Structure

CellPress
OPEN ACCESS

Review

The future of integrated structural biology

Harald Schwalbe,[1,2,*] Pauline Audergon,[2] Natalie Haley,[2] Claudia Alen Amaro,[2] Jon Agirre,[3] Marc Baldus,[4] Lucia Banci,[5] Wolfgang Baumeister,[6] Martin Blackledge,[7] Jose Maria Carazo,[8] Kristina Djinovic Carugo,[9] Patrick Celie,[10] Isabella Felli,[5] Darren J. Hart,[7] Thomas Hauß,[11] Lari Lehtiö,[12] Kresten Lindorff-Larsen,[13] José Márquez,[9] André Matagne,[14] Roberta Pierattelli,[19] Antonio Rosato,[5] Frank Sobott,[15] Sridhar Sreeramulu,[1] Jan Steyaert,[16] Joel L. Sussman,[17] Lukas Trantirek,[18] Manfred S. Weiss,[11] and Matthias Wilmanns[20]

[1]Center for Biomolecular Magnetic Resonance (BMRZ), Institute for Organic Chemistry, Max-von-Laue-Str. 7, 60438 Frankfurt/M., Germany
[2]Instruct-ERIC, Oxford House, Parkway Court, John Smith Drive, Oxford OX4 2JY, UK
[3]York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 3BG, UK
[4]NMR Spectroscopy, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, Utrecht 3584 CH, the Netherlands
[5]Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine—CIRMMP, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Italy
[6]Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany
[7]Institut de Biologie Structurale, Université Grenoble Alpes-CEA-CNRS UMR5075, 71 Avenue des Martyrs, 38000 Grenoble, France
[8]Biocomputing Unit, National Centre for Biotechnology (CNB CSIC), Campus Universidad Autónoma de Madrid, Darwin 3, Cantoblanco, 28049 Madrid, Spain
[9]European Molecular Biology Laboratory (EMBL) Grenoble, Grenoble, France
[10]Division of Biochemistry, The Netherlands Cancer Institute, Amsterdam, the Netherlands
[11]Macromolecular Crystallography, Helmholtz-Zentrum, Albert-Einstein-Str. 15, 12489 Berlin, Germany
[12]Faculty of Biochemistry and Molecular Medicine and Biocenter Oulu, University of Oulu, Oulu, Finland
[13]Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Copenhagen, Denmark
[14]Laboratory of Enzymology and Protein Folding, Centre for Protein Engineering, InBioS Research Unit, University of Liège, Building B6C, Quartier Agora, Allée du 6 Août, 13, 4000 Liège (Sart-Tilman), Belgium
[15]Astbury Centre for Structural Molecular Biology and School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK
[16]VIB-VUB Center for Structural Biology, VIB, Pleinlaan 2, Brussels, Belgium
[17]Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel
[18]Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 753/5, 62500 Brno, Czech Republic
[19]Department of Chemistry "Ugo Schiff", University of Florence and Magnetic Resonance Center, University of Florence, 50019 Sesto Fiorentino, Italy
[20]European Molecular Biology Laboratory (EMBL) Hamburg, Hamburg, Germany
*Correspondence: schwalbe@nmr.uni-frankfurt.de
https://doi.org/10.1016/j.str.2024.08.014

SUMMARY

Instruct-ERIC, "the European Research Infrastructure Consortium for Structural biology research," is a pan-European distributed research infrastructure making high-end technologies and methods in structural biology available to users. Here, we describe the current state-of-the-art of integrated structural biology and discuss potential future scientific developments as an impulse for the scientific community, many of which are located in Europe and are associated with Instruct. We reflect on where to focus scientific and technological initiatives within the distributed Instruct research infrastructure. This review does not intend to make recommendations on funding requirements or initiatives directly, neither at the national nor the European level. However, it addresses future challenges and opportunities for the field, and foresees the need for a stronger coordination within the European and international research field of integrated structural biology to be able to respond timely to thematic topics that are often prioritized by calls for funding addressing societal needs.

INTRODUCTION

Instruct-ERIC (Instruct) is a European research infrastructure for integrated structural biology. Its mission is to provide access to cutting-edge technologies in this field of research. Integrated structural biology provides important fundamental insight into the structure and dynamics of biological macromolecules (biomacromolecules). This structural insight is an essential basis for a mechanistic understanding of all transformations that occur in organisms from all living kingdoms.

When Instruct was founded 15 years ago, the move from structural biology to *integrated* structural biology was perceived as the essential next step to advance our fundamental understanding of cellular biology. Integrated structural biology combines insight from multiple complementary techniques. Such integrative approaches allow gathering information at all available sizes and time scales. Pushing those limits for increased temporal and spatial resolution remains of prime interest for structural biology, as it allows us to get deeper structural insights into how biomacromolecules exert function.

Integrated structural biology uses a multitude of technologies. No single laboratory can maintain expertise and hardware at the forefront of all possible techniques. Thus, access to all technologies is essential and Instruct provides access to hardware,

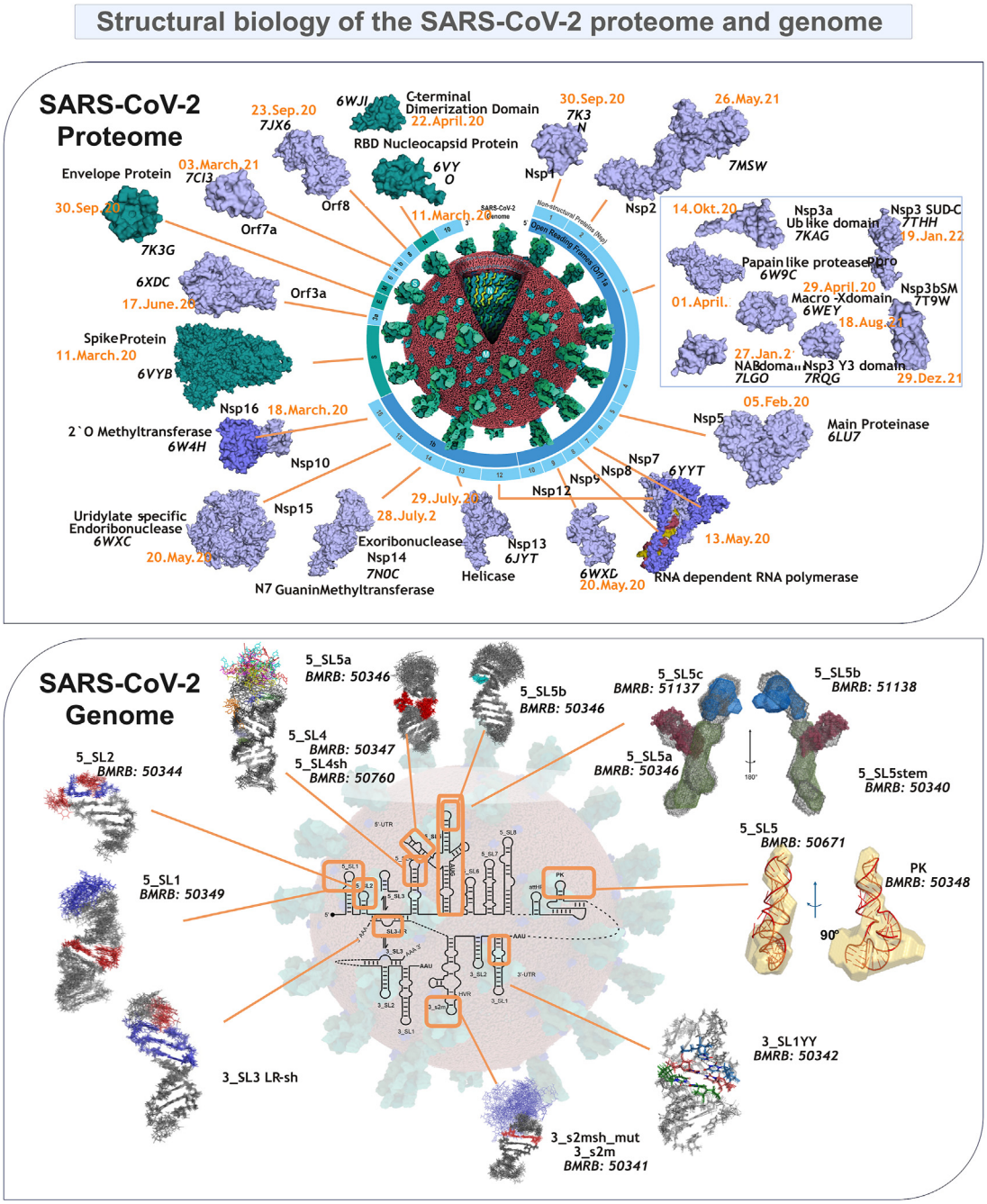

Figure 1. Integrated structural biology toward understanding COVID-19

SARS-CoV-2 protein structures solved mainly by X-ray crystallography (Graphic adapted from https://cdn.rcsb.org/pdb101/learn/resources/flyers/covid-genome/sarscov2-genome-prots.pdf and J. H. Lubin et al. [2021][1]) and cryo-electron microscopy (top). SARS-CoV-2 genome (RNA elements) structures solved by NMR spectroscopy (bottom).

analytic tools and human expertise through its network of distributed research infrastructures.

In the following, we showcase the advantages of coordinated research initiative and access capability on the example of the reaction of the structural biology community to the emergence of COVID-19. Soon after its outbreak, different fields of integrated structural biology demonstrated their power and maturity (Figure 1) to solve the structures of the viral proteins and RNAs to provide a fundamental understanding of viral biology. Viral protein structures were solved by X-ray crystallography and cryoelectron microscopy (cryo-EM), and viral RNA structures are currently being determined by nuclear magnetic resonance (NMR) spectroscopy. The impact of mutations in variants of concern on the viral proteins, especially of the viral receptor

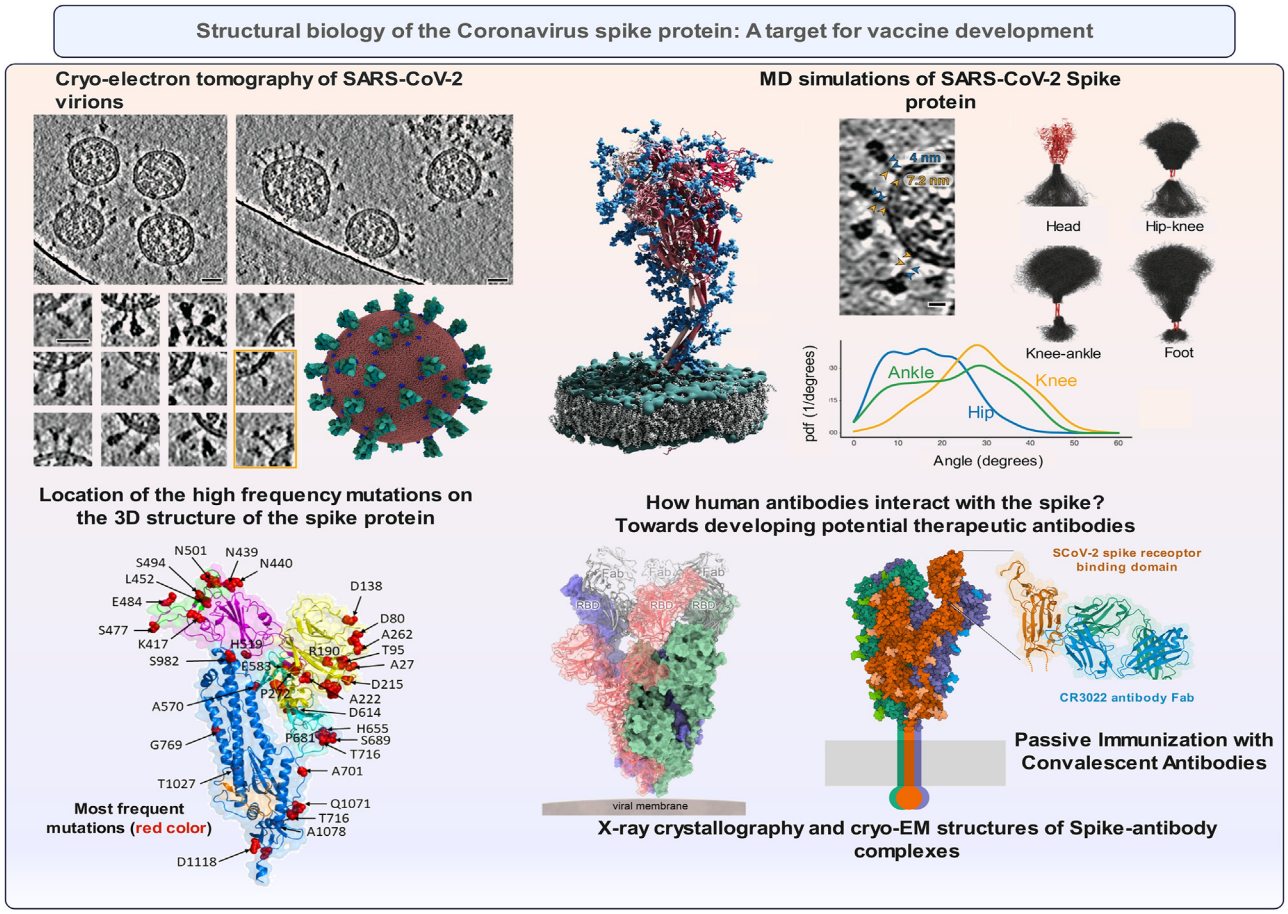

Figure 2. Integrated structural biology and molecular dynamics of the essential and complex SPIKE protein from SARS-CoV-2 and the impact of mutations in variants of concern

(Top left and right) Cryo-electron tomography of SARS-CoV-2 virions and molecular dynamics simulations of SARS-CoV-2 Spike (S) protein (graphic adapted from Beata Turoňová et al.[2]). (Bottom left) Based on clustering and alignment tools the most frequent mutations observed (red color) in the S protein (graphic adapted from Negi, S.S. et al.[3]). (Bottom right) Cryo-EM structure of the S protein in complex with Fabs (graphic adapted from Walsh, M. A. et al.[4] and https://pdb101.rcsb.org/learn/resources-to-fight-the-covid-19-pandemic/passive-immunization-with-convalescent-antibodies.)

spike, was monitored on the fly by structural biology techniques (Figure 2), and dynamic aspects fundamental for its function were described by molecular dynamics simulations.

In this review, scientists with a record of accomplishment in integrated structural biology discuss their perspectives on the most important developments in this field. Further, it collates different perspectives and discusses existing technologies, individually and in their integration, new areas for integrated structural biology as well as technologies that either need to be expanded or newly integrated.

DEFINITION OF INTEGRATED STRUCTURAL BIOLOGY

In its broadest sense, structural biology determines the arrangements of biological systems over spatial scales spanning ten orders of magnitude. Traditionally, "structural biology" refers to atomic-scale studies of biomacromolecules. Further, the use of integrative approaches can steer toward understanding superstructures of large biomolecular assemblies from atomic to cellular resolution.

Pioneering discoveries, all honored by Nobel prizes, by John H. Northrop and Wendell M. Stanley for preparation of enzymes and proteins, Max F. Perutz and John C. Kendrew for the first protein structure determination and Christian B. Anfinsen for protein folding studies transformed our understanding of the very essence of protein chemistry and protein function. They set off a field of "structural biology," attracting chemists and physicists to decipher the structural intricacies of macromolecules derived from biology. With the improved ability to express and purify recombinant and isotopically enriched proteins and improved methodologies to determine their near atomic-resolution 3D structures by means of X-ray crystallography and NMR spectroscopy, these early works on structural biology provided mechanistic insights into modes of action of proteins.

Structure determination is often a time- and sample-demanding enterprise, requiring either isolation from natural sources or samples produced using recombinant technologies, both in their non-modified forms and with post-transcriptional or post-translational modifications. Therefore, the current atlas of structural biology space is far from complete and lacks

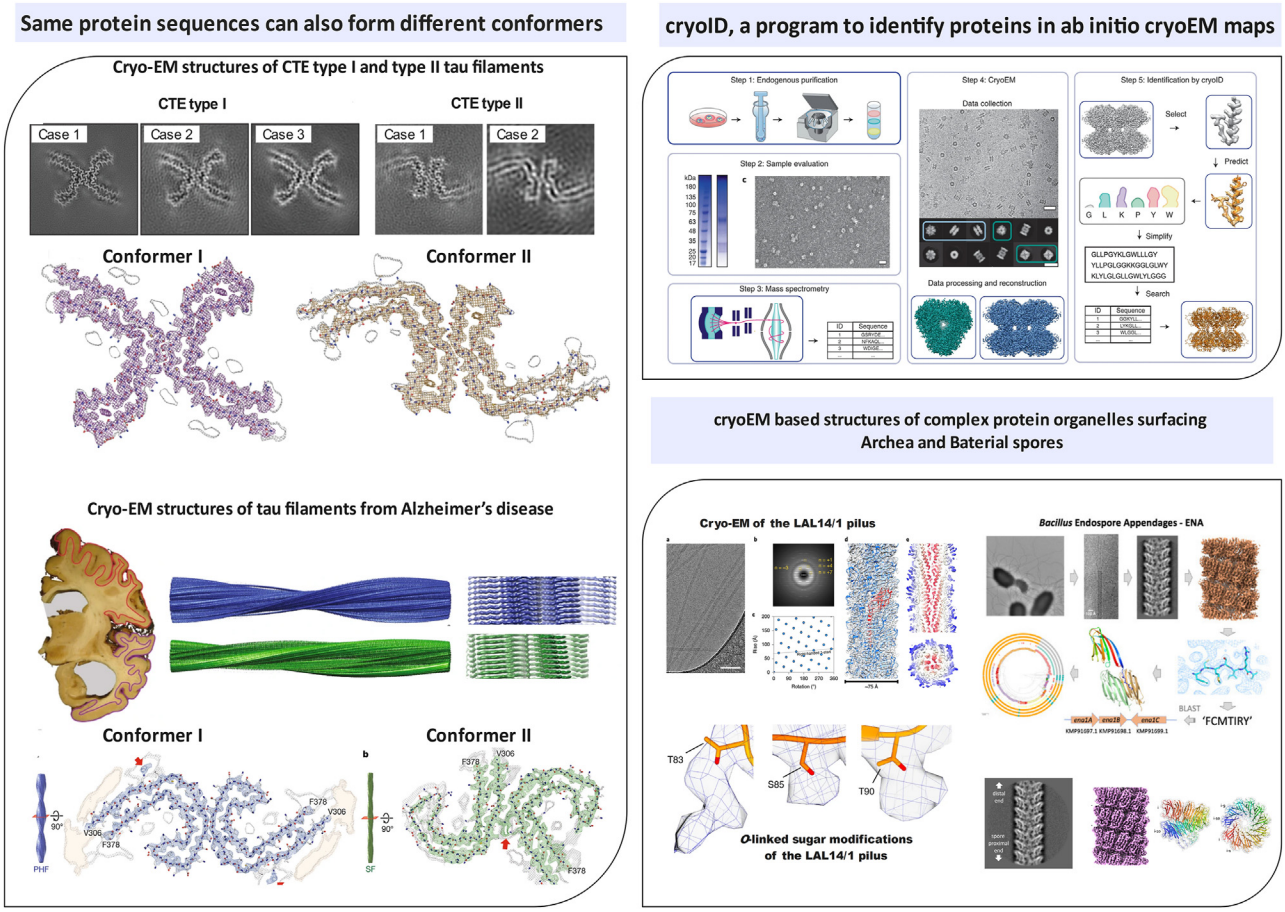

Figure 3. Near-atomic resolution applications of cryo-EM technology

(Left) Cryo-EM structures of the chronic traumatic encephalopathy (CTE) type I tau filaments and their high resolution cryo-EM maps overlayed with corresponding atomic models (graphic adapted from Falcon B. et al.[5]). (Top right) Workflow to identify proteins in cryo-EM maps from heterogeneous mixtures from endogenous sources using endogenous structural proteomics approach (graphic adapted from Ho, Chi-Min et al.[6]). (Bottom right) O-linked glycosylation glycosylation as observed by cryo-EM (graphic adapted from Wang F. et al.[7]) and molecular identity of Endospore Appendages (Enas) of *Bacillus cereus* using cryo-EM (graphic adapted from Pradhan B. et al.[8]).

contributions from fields like glycosylated proteins and (modified) RNA. The lack of RNA structures exists despite a recent surge in interest in RNA biology as well as in studies of post-transcriptional and post-translational modifications.

At the same time, structural biology takes breath-taking turns and breakthroughs every year. A powerful example comes from studies of amyloid filaments extracted from the brains of individuals with different neurodegenerative diseases, revealing that these diseases may be correlated to unique conformational "strains" of the same protein (Figure 3).[5,9] In another example, the near-atomic-resolution cryo-EM maps from unidentified protein complexes enriched directly from the malaria-causing parasite *Plasmodium falciparum* were reconstructed. Similarly, Wang et al. and Pradhan and colleagues used cryo-EM for the molecular identification of organelles on the surface of Archaea or bacterial spores, respectively.[7,8] Cryo-EM comes in two flavors: single particle analysis (SPA) and cryoelectron tomography (cryo-ET). SPA requires multiple copies of repetitive structures and has become a mature method continuously depositing structures in the protein databank (PDB) at high resolution. *In situ* structural biology is one of the most exciting recent developments in integrated structural biology. Cryo-ET is a method that holds great promise for studying supramacromolecular structures *in situ*, i.e., in their unperturbed natural habitats.

Over the ensuing decades, structural biology matured into a field of research capable of unraveling the molecular basis of biological processes, increasingly enabling landmark discoveries in physiology and medicine and is still today informing the way new drugs are discovered and improved.

CHALLENGES FOR INTEGRATED STRUCTURAL BIOLOGY

Over the past decade, numerous developments have taken place that challenge the position of integrated structural biology as a scientific field and the associated question of to where to expand capacities for future research. Here, we identify the following challenges.

(1) The success of artificial intelligence (AI)-based protein structure prediction requires a renewed positioning of integrated structural biology. How do we react to this

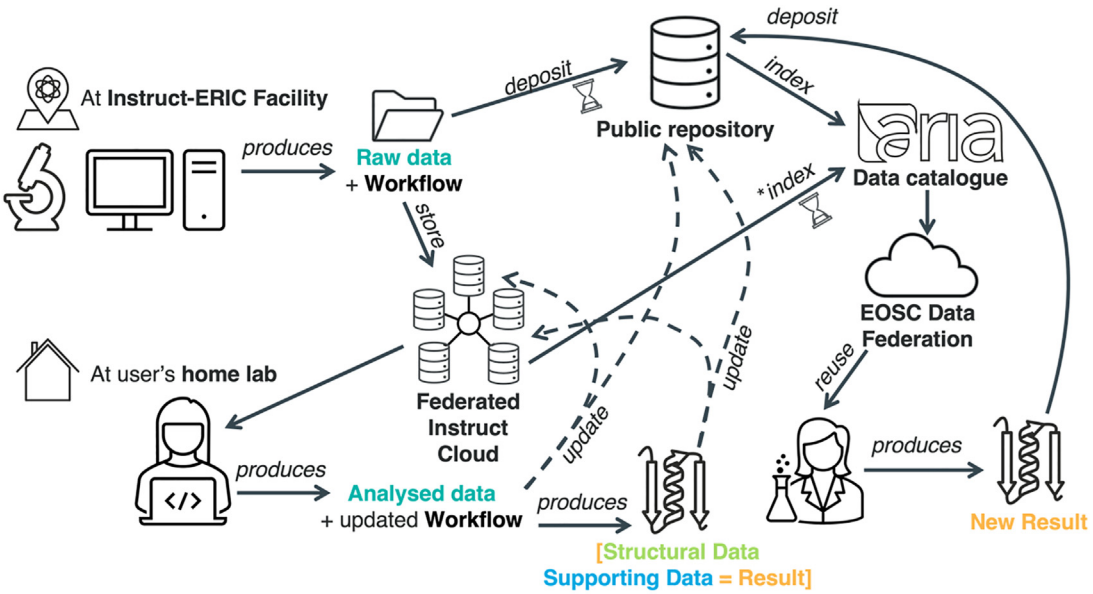

Figure 4. Overview of data management workflow in Instruct-ERIC

development? Are all experimental technologies still needed given the success of AI in structure prediction?

(2) The developments of cryo-ET raise the possibility that single-particle cryo-EM, X-ray, and NMR structure determination become routine techniques that are no longer at the forefront of scientific developments.

(3) Given the aforementioned points, the question of distributed versus centralized facilities becomes of renewed importance.

Interestingly, the answers collated in this review, according to the diverse perspectives from different structural methods and technologies are highly coherent: the more rapid access to high quality AI-derived structural models will not compete with experimental structure determination but will instead fuel and expedite hypothesis-driven research and more sophisticated follow-up experimental studies. These follow-up studies will enrich AI-derived models with missing key information, e.g., information on hydrogen bonding, roles of structural water molecules, and inherent dynamics including ensemble representations of biomacromolecules and their complexes to describe biological function.

Additionally, capturing the dynamic behavior of biomolecules over time, adding a temporal dimension to the traditional 3D structural information, is highly important. This includes understanding how biomolecular structures change and interact in response to different environmental conditions or during biological processes (e.g., replication, splicing, transcription, and translation) and has led to the coining of the term "4D structural biology." This encompasses techniques and methodologies aimed at capturing and analyzing these dynamic aspects of biomolecular structures. To achieve this, synchronized induction of structural transitions and signal-to-noise reduction to detect the accumulated differences during structural transition are key challenges.

Integrated structural biology studies, which integrate findings from different structural biology techniques based on fundamentally distinct physical-chemical principles that have brought about multiple technical challenges, including the following.

(4) Data integration: Different techniques generate different data types with varying resolutions, noise levels, and limitations, making combining them into a cohesive model difficult.

(5) Computational analysis: Developing robust computational methods that can handle complex datasets and provide accurate predictions of molecular structures and dynamics is an ongoing challenge.

(6) Data sharing and standardization: There is a growing need for standardized formats and protocols for sharing appropriately curated data and their analyses.[10] Ensuring that different research groups can consistently share and interpret data is critical to advancing the field (Figure 4). Further, the availability of productive raw experimental data for later data analysis by other groups will fuel AI-driven approaches.

Correlating structural biology data with biological-functional data, a critical step in understanding the function of biological macromolecules and complexes in living systems, poses further challenges to the integrated structural biology concept.

(7) Correlating structural dynamics with biological function: Correlating structural dynamics with biological function requires time-resolved structural biology techniques and integration with other biophysical and biochemical experiments.

(8) Complexity of biological systems towards 4D structural biology: An accelerating trend in structural biology is to determine 4D structures of complexes. Structure determination using high-throughput (HT) serial crystallography

and cryo-EM time-resolved approaches are extending the scope of X-ray crystal structure analysis to the determination of sets of related structures or time-series recorded during dynamic processes on the millisecond timescale. Cryo-EM allows determination of large structures unhindered by crystal contacts and in near physiological conditions.

(9) Decoding principles of *de novo* folding of biomacromolecules: Addressing this issue will require continued innovation in sample preparation, imaging conditions, and hardware applied, including microfluidics, data processing, and analysis methods, particularly in in-cell NMR spectroscopy and correlative (fluorescence) microscopy.

(10) Development of novel tools to characterize complex biological processes in the cellular context. These processes range from time-dependent and cellular state-specific metabolic fluxes (by time-resolved NMR) to signaling cascades (cryo-EM, solid-state NMR [ssNMR], fluorescence microscopy/spectroscopy, and their combinations).

CURRENT STATE-OF-THE-ART OF INTEGRATED STRUCTURAL BIOLOGY TECHNOLOGIES

We here review the current state-of-the-art of integrated structural biology technologies, having in mind that these technologies are in place in the distributed centers of Instruct.

Sample preparation

Many structural biology projects start with the preparation of the biological sample. Highly sophisticated heterologous and cell-free expression systems have been developed to produce prokaryotic and eukaryotic proteins, including post-translational modifications. Similarly, natural and modified RNAs have become increasingly available. Thus, sample preparation has become a key asset for Instruct as part of access for broad user communities.

Automated sample preparation

The introduction of robotic units with automated liquid handling allows the preparation and manipulation of large numbers of samples in a reasonable time, with high accuracy and precision. They enable HT screening of variables to optimize protein expression (including cloning, domain boundary definition, mutagenesis, transformation, cell line generation, and colony picking). Challenging and diverse targets, including intrinsically disordered proteins (IDPs), domains from giant complexes, human disease-related and viral pathogens, can be made available for structural studies using this approach.

Protein expression systems in bacteria (*Escherichia coli* and *Bacillus subtilis*) or yeast (*Pichia pastoris*) are in place. Cells can be transferred for purification and refolding of inclusion bodies in a 96-well format in a second robotic workstation. Capillary electrophoresis allows fast and easy automated separation of nucleic acids and proteins; microplate readers are used for enzymatic assays; biolayer interferometry (BLI) allows the analysis of biomolecular interactions in 96- and 384-well microplates, protein quantification and automated biomolecular interaction assays. For example, BLI is used to measure the avidity[11–13] and affinity[14,15] of polyclonal and monoclonal antibodies, respectively, with specificity for any purified antigen. In two completely unrelated studies, BLI was used to follow the kinetics of interaction of engineered zinc finger proteins with single-stranded and structured RNAs,[16] and to characterize the simple 1:1 noncovalent interaction between the bone morphogenetic protein 2 and its cognate antagonist Noggin.[17]

Nanobodies, membrane proteins, eukaryotic IDPs

Nanobodies form a particular class of proteins often associated with membrane protein research. These small (~15 kDa) proteins derived from camelids can usually be expressed as recombinant proteins in *E. coli*, after subcloning and phage- or yeast-display screening. Nanobodies not only serve as stabilizing agents for proteins but can also aid structure determination (cryo-EM and crystallization) or act as inhibitors or activators to modulate protein activity. Nanobodies, including variants like bi-valent and bi-specific nanobodies, will therefore not only remain a precious resource for structural biology but may also have applications in disease treatment.[18]

The production and characterization of membrane proteins, which are notoriously difficult to express, has also advanced in the past years. In particular, the improvements in both eukaryotic- and cell-free expression systems, as well as progress in methods for solubilization, screening, and lipid cubic phase (LCP) crystallization have enhanced structural characterization. Because of their biological role in signaling and hence their status as druggable molecules, membrane proteins are recognized as the most important class of drug targets. Structure-function analyses will remain of high interest to the scientific community and industry.

An upcoming trend in the field of structural biology is the characterization of proteins containing intrinsically disordered regions (IDRs)[19] and proteins containing post-translational modifications[20] like glycosylation[21] and phosphorylation. In particular, the role of phosphorylation sites located within IDRs in protein function[22] has become a topic of interest. In this respect, the expression of these proteins in eukaryotic cells is essential to obtain native post-translational features.

Design of new proteins

The continuous stimulus to the field of protein structure prediction provided by the CASP (critical assessment of structure prediction) initiative has fostered the development of AI-based protein structure-prediction programs has paved the path in developing programs like AlphaFold2,[23] RoseTTAFold,[24] and Colabfold[25] that now allow the determination of structures from amino acid sequences.

The recent advances in AI also provide opportunities for structure-based construct design for protein production. Although these computationally derived structures may reduce the necessity of producing an experimentally determined protein structure *per se*, these models still require experimental validation to assess structure-function relationships.

RNA preparation

For RNA structural biology, sample preparation methodologies include solid-phase chemical synthesis and *in vitro* transcription methodologies. While chemical synthesis methodologies are limited in size, they allow the introduction of modified nucleotides that confer chemical stability to RNA for in-cell studies,

particularly for time-resolved studies introducing photolabile protecting groups.[26,27] Within the drug development process, chemically modified mRNA plays a key role in advancing mRNA technology. Incorporation of modified nucleotides can result in significant improvement in the half-life of the mRNA, thereby improving the translation efficiency and immunological profile. Enzymatic methodologies alleviate the synthetic size limitation of chemical synthesis methods, but for future RNA structural biology, incorporation of modified RNA nucleotides with high fidelity will become essential, as also evidenced by the impact of modified nucleotides on mRNA-based vaccine development. For example, the replacement of uridine with *N*1-methyl-pseudouridine can circumvent innate immune responses but result in more efficient translation. Further, incorporation of 16 modified nucleotides and six ^{19}F-labeled nucleotides was achieved using chemo-enzymatic methods, thus, enabling the characterization of RNA fold, ligand binding and kinetics.[28]

Mass spectrometry

Mass spectrometry (MS) has evolved from a method for mass detection of biomacromolecules toward probing the structure and dynamics of proteins and the complexes they form with other (bio)molecules. There are now several specialized MS methods, each with unique sample preparation, data acquisition, and data processing protocols.[29] These structural MS methods can be divided into three groups.

Native and top-down approaches

Native and top-down approaches preserve non-covalent interactions, which define biomolecular conformations and interactions, in the gas phase of the mass spectrometer. They deliver information on mass, identity, integrity, and stoichiometry of binding partners, subunit exchange kinetics (native MS); global size and shape of analytes, folding states and overall complex topology (ion mobility); and subunit composition, relative fold stability and interaction strength, as well as exposed and highly dynamic residues (top-down fragmentation).

Labeling approaches

Labeling approaches include the following: (i) hydrogen-deuterium exchange (HDX), where exposed and locally unstructured amide backbone hydrogens are reversibly exchanged against deuterium; (ii) covalent labeling that may utilize specific chemical reactivity or nonspecific hydroxyl radicals produced *in situ* (e.g., with a UV laser or synchrotron radiation, on msec–μsec timescales) to permanently modify residues; and (iii) chemical or photo-crosslinking where specific residues within a defined intra- or intermolecular distance (“contacts”) are covalently linked with each other. Results from labeling experiments often provide valuable restraints for computational modeling. Especially crosslinking data are used for this purpose. However, experiments involving HDX and covalent labeling typically compare between two states, and direct structural interpretation is still in its infancy.

Proteomics approaches

Proteomics approaches use peptide-level readouts on whole (sub-)proteomes to sample, e.g., the folding and interaction state of individual proteins (limited proteolysis), their thermal stability (cellular thermal shift assays or thermal proteome profiling), or other proteins in their proximity (turbo-ID). These are essentially biochemical experiments with an MS readout using bottom-up quantitative proteomics methods *in vitro* or *in vivo* upon application of a given stimulus.

A current challenge in method development is to increase spatial resolution, e.g., to elevate HDX-MS from the typical peptide-level resolution to single amino acid-level resolution. This involves using novel experimental methods such as electron-driven fragmentation that minimizes H/D scrambling, along with software strategies that model intrinsic exchange rates. The next challenge involves targeting samples of increasing complexity. This includes classic *in vitro* sampling from dilute, aqueous solutions to, environments like detergents or lipids for membrane proteins, or conditions that induce liquid-liquid phase separation (LLPS), allowing these elusive states to be investigated. Moreover, there is a need to develop MS approaches for *ex vivo* (sub-cellular extracts) and *in vivo* structural biology.

X-Ray crystallography

X-ray structure determination

X-ray crystallography might well be the technique most impacted by the structure prediction AI tool AlphaFold. While determination of the phases by anomalous scattering methods used to be the major application of X-ray crystallography, now it is possible to have a reliable model with AlphaFold. For most amino acid sequences, a three-dimensional structure can be computed; however, the computed structure may or may not answer the biological question posed. One may ask, what is left for X-ray crystallography then? Now, the first answer comes from the fact that AlphaFold can predict the fold of an amino acid sequence to near experimental accuracy; nevertheless, an experimental structure of a protein is still information richer than a computed model. This holds particularly true for fields such as structure-based drug discovery, wherein high experimental accuracy is the key to success. Further, X-ray crystallography and AF are there to provide atomic resolution models for the interpretation of cryo-EM maps. However, identities and oxidation states of metals can only be reliably established using X-ray crystallography.

Advances in X-ray crystallography technologies have led to large-scale fragment screening experiments, an important step in the early stages of drug discovery projects. Further, serial crystallography experiments, first pioneered at XFELs and now more and more relayed to synchrotrons, help conduct time-resolved studies and thereby adding the time dimension to the traditionally static crystal structures.

For NMR studies, the existence of a wealth of high-resolution structural data from crystallographic studies provides models to interpret the conformational dynamics detected in liquid-state NMR studies. This is at the level of precision that can be obtained by X-ray structure, where integrated X-ray/NMR can provide detailed insight into subtle, but functionally highly relevant structural detail. Figure 5 provides an example of the structure determination of GPCRs and follow-up studies by NMR to understand the molecular origin of receptor signaling (Figure 5).

Finally, the AI success was enabled by the vast number of structures determined by X-ray crystallography. The efficiency of X-ray structure production is impressively documented by the COVID Moonshot initiative,[31] which has significantly contributed to the development of antiviral medication (Figure 6).

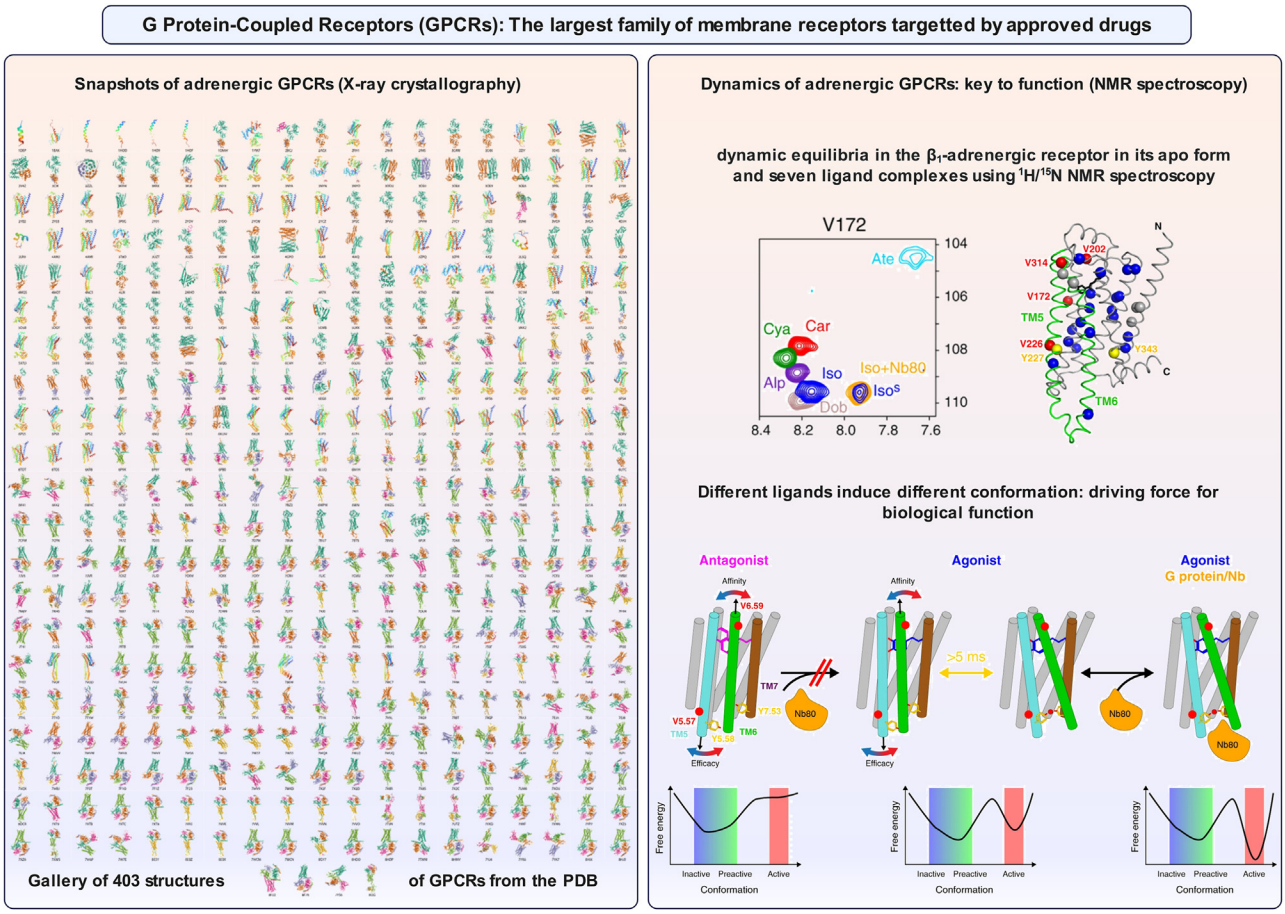

Figure 5. Showcase example of integrated structural biology research on GPCR receptor signaling states
(Left) Snapshots of 403 structures of GPCRs from the PDB. (Right) GPCRs (β_1-adrenergic receptor) conformational dynamics and ligand interactions as studied by NMR spectroscopy (graphic adapted from Grahl A. et al.[30]).

Structural biology, particularly X-ray crystallography, has pioneered scientific data management and the principles of FAIR (findable, accessible, interoperable, and reusable) data. Several tools have been developed to keep track of vital metadata containing information on the sample, crystal conditions, and diffraction data. Examples include IceBear[35] and CRIMS[36] of Instruct and ISPyB[37] tracking experiments at synchrotrons.

X-ray tomography

Cryo-ET allows the precise localization of molecular structures *in situ* at (quasi)-atomic resolution. However, the technique is limited by the low penetration of electrons, meaning that sample preparation is tortuous and comprises the analysis of ''slabs'' rather than entire cells. This limitation can be alleviated by using X-rays, where the penetration is much higher, enabling visualizing relatively thick objects. Advances in X-ray tomography, combined with 4th generation synchrotron sources, have led to resolutions of 10 nm. With an order of magnitude difference in resolution in comparison to electron tomography, X-ray tomography can fill the gap between light and electron microscopy, offering substantial improvements in sample size and significant reductions in experiment time scales.

Cryo-EM

Cryo-EM comes in two flavors: SPA and cryo-ET. SPA requires multiple copies of repetitive structures and has become a mature method consistently contributing structures to the protein databank (PDB). With the ''resolution revolution'' resulting from advances in technology (direct electron detectors) and methodology (image processing), it routinely delivers resolutions in the 2–3 Å range. It is particularly powerful for studying large macromolecular structures but less so with smaller (<50 kDa) macromolecules. A remaining problem is the unpredictable behavior of macromolecules at the water-air interface, which often results in (partial) unfolding or preferred orientations affecting the quality of 3D reconstructions.

In situ structural biology is one of the most exciting recent developments in integrated structural biology. Cryo-ET is a method that holds great promise for studying supramacromolecular structures *in situ*, i.e., in their unperturbed natural habitats. It allows the localization and analysis of molecular structures in their native environment and in few cases at quasi-atomic resolution. Recent highlights that demonstrate the power of this approach are the classification of in cell ribosome structures in *Mycoplasma pneumoniae* from the Mahamid lab[38] or the recent

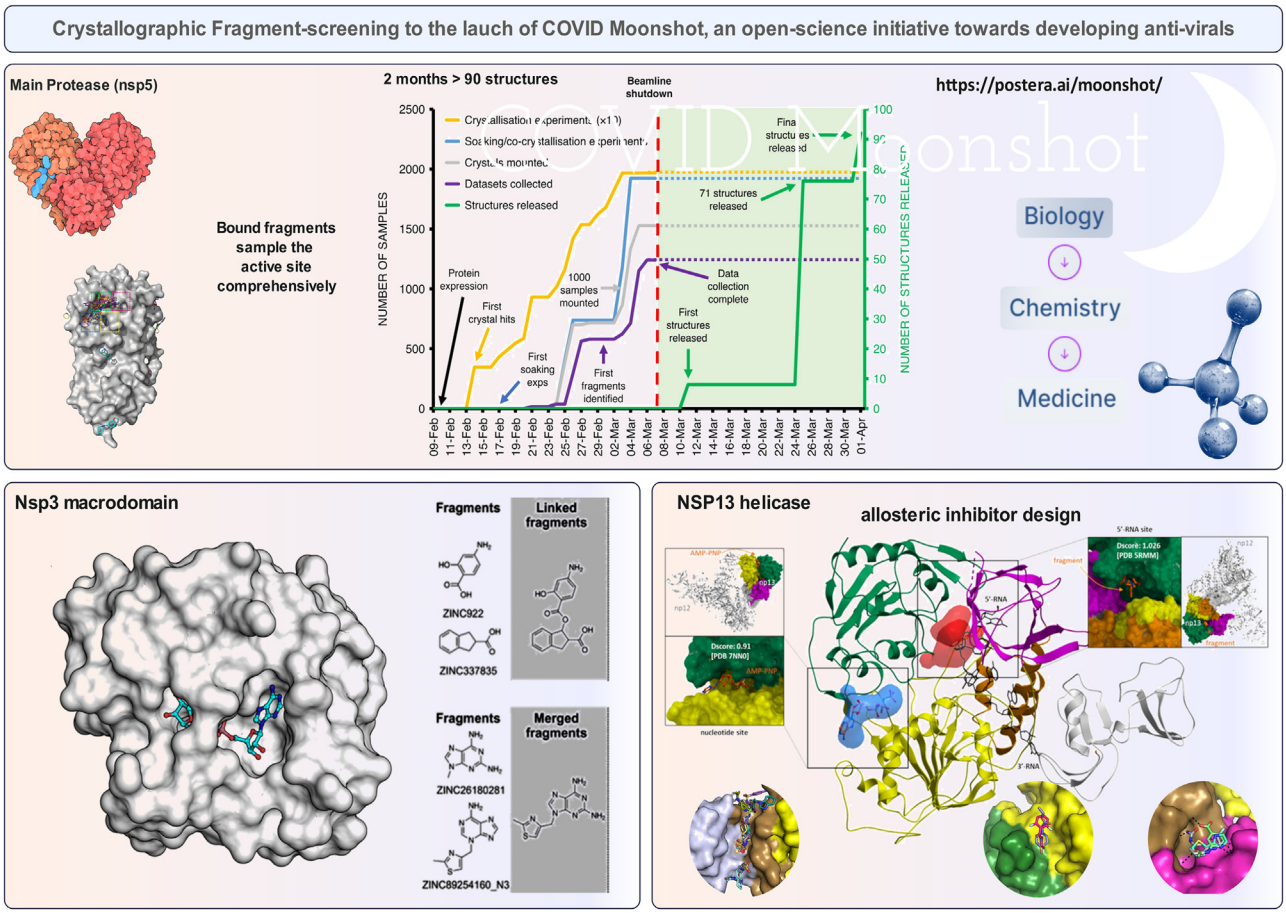

Figure 6. Fragment-based drug discovery using X-ray fragment screening applied to selected SARS-CoV-2 proteins

(Top panel) Crystallographic-based fragment screening against the main protease of SARS-CoV-2 and medicinal chemistry follow-up via the COVID Moonshot project (graphic adapted from Douangamath A. et al.[32]). (Bottom left and right) Crystallographic-based fragment screening against the macrodomain of the nsp3 and nsp13 proteins of SARS-CoV-2 (graphic adapted from Schuller M. et al.[33] and Newman, J. A[34]).

manuscript from the Beck lab where a drug bound to a ribosome was visualized at 2.5 Å in human cells.[39] Additionally, further technical and software developments are required to locate smaller assemblies in cells.

Future developments in cryo-EM

Technical and software developments to improve the workflows in cryo-ET include the following.

(1) Current process in speeding up automation of focused-ion-beam (FIB); indeed, thinning by FIB technology should become more routine.
(2) Tomographic data acquisitions will become much faster.
(3) Improved instrumentation (automated cell freezing, plasma-FIB, laser phase plates, higher voltages, Cc correctors, liquid helium temperatures).
(4) Liftout-techniques for cryo-ET of organisms, including *Drosophila* embryos and *C. elegans.*
(5) AI-guided enhancements to cryo-ET processing. Automated annotation of complexes in cryo-ET could be achieved using AI.
(6) Machine learning is also increasingly used for denoising of tomograms and particle picking. AlphaFold provides templates for an exhaustive search of tomograms. Since the options for subtomogram averaging improving contrast and resolution are limited to large and abundant structures, the goal must be to get to subnanometer resolutions without averaging. This requires further advances in technology and methodology and realizing the full potential of cryo-ET.

Strong synergies exist between developments in cryo-ET and super-resolution light microscopy. For example, super-resolution microscopy at cryogenic temperature allows the location of fluorescently marked complexes to be determined at high precision and supports the annotation of cryo-tomograms by correlative approaches.

NMR spectroscopy

NMR and electron paramagnetic resonance (EPR) spectroscopy are closely related magnetic resonance spectroscopy methods as far as their quantum chemical foundations are concerned.

In areas where traditional X-ray-based protein structure determination has remained difficult, the approach will start with an AlphaFold-computed structure model. NMR is unique in the

ability to design experiments to rapidly validate structural models (such as chemical shifts and residual dipolar couplings). Additionally synergistic developments are planned to make NMR more readily applicable across a broad range of contexts.

Technological developments for NMR

The technologies underlying NMR spectroscopy are constantly evolving. They support NMR applications, both in the liquid (ls NMR) and in the solid (ss NMR) state and include ultrahigh-field NMR (1.2 GHz and beyond), dynamic nuclear polarization (DNP), and ultra high-speed magic angle sample spinning (MAS).

Worldwide, the first magnet at 1.2 GHz has been installed at the Italian Instruct Centre in Florence. New magnet design technology was required to combine the traditional superconductors that could generate the highly homogenous fields required using high-temperature superconductors, including rare-earth barium copper oxide (REBCO) materials.

Technological breakthroughs at the European company Bruker are opening avenues to further increase magnetic field strength to 1.5 GHz systems, expected to become operational by the end of the decade. When combined with ultrafast MAS, the combined technological advantages of high field and MAS spinning at a high rate will revolutionize proton-detected solid-state NMR spectroscopy at ultrahigh field. Further, coupling of nuclear spin excitation with methodologies for increasing the transfer of polarization, e.g., from electrons to protons either inside the NMR active volume (DNP) or outside the NMR magnet (hyperpolarization, PHIP, and optical pumping) have boosted signal-to-noise by 2–3 orders of magnitude and reduced measurement times from years to days.

Application envelope of NMR spectroscopy

NMR stands out in its extremely broad application window, from analysis of large biological systems in all phases of matter (liquid, solid, and phase-separated) to molecules of low molecular weight and their quantification and flux in biofluids (NMR-based metabolomics) and in food, medicines to materials including those serving sustainable energy production, e.g., battery research.[40]

NMR and exchanging systems

In aqueous solutions and physiological temperatures, biomolecules, including proteins and nucleic acids (NA), do not adopt a single conformation but rather exist in rapid equilibrium with alternative low-populated high-energy (excited) states, which are invisible for conventional biophysical techniques, such as X-ray crystallography, cryo-EM, and MS, challenging to study by computational approaches, e.g., molecular dynamics, due to their limitations in sampling of the biomolecule conformational space, and out of reach of available AI tools trained on ground-state structural data. The emerging evidence indicates that sampling among different conformations is vital for biological function[41,42] and that stabilizations of alternate conformations can lead to diseases[41,43] extensive differences in key structural elements between the conformational states, each displaying distinct biological activity, then give rise to multiple intrinsic regulatory mechanisms.

The unique capabilities of NMR spectroscopy in exchanging systems enable the detection, characterization, and visualization of these sparsely populated conformational states.[44] These states, albeit transient, have facilitated the demonstration of direct couplings between activity and intrinsic dynamics in several proteins, including reductases,[45] kinases,[46,47] chaperones,[44,48,49] and hydrolases.[50] Recently, Shukla et al.[51] illustrated that reduced enzymatic activities, inhibitor affinities, and residence times for mutants of histone deacetylase are all reflected by the rate constants between intrinsically sampled conformations, indicating that dynamic sampling of conformations dictates both enzymatic activity and inhibitor potency for this enzyme. NMR spectroscopy has also revealed the importance of intrinsic dynamics for RNA functionality[42,52,53] and suggested that dynamic transitions between Watson-Crick base pairs and alternative low-populated Hoogsteen conformations, influenced by local sequence context, contribute to DNA recognition and repair.[54,55]

Undoubtedly, NMR relaxation-based approaches are crucial for comprehending biomolecular activities under *in vitro* conditions. Adapting NMR relaxation techniques to in-cell NMR settings will be indispensable for gaining insights into biomolecular activities *in vivo*. The main challenge in applying NMR relaxation approaches to measurements of biomolecules in cells will be overcoming problems with NMR's low sensitivity.

IDPs by NMR

Paradoxically, advances in AI structure prediction have been accompanied by a much broader appreciation of the fact that large sections of all known proteomes are too dynamic to be described in terms of a single set of structural coordinates. IDPs as well as regions of disorder in proteins (IDRs), whose existence and biological role remained controversial only two decades ago, represent an important family of proteins whose highly dynamic properties unquestionably define their biological function.[56–60] IDPs are implicated in important pathologies, for example, controlling essential regulatory and signaling pathways involved in cancer, playing instrumental roles in numerous host-pathogen interactions and forming the molecular basis of neurodegenerative disease.

Novel NMR approaches will continuously be developed to probe the conformational dynamics of IDPs and to study protein-protein and protein-ligand interactions. So-called chemical exchange NMR maps the trajectory from the conformational equilibrium sampled by the IDP in its free state, via possible transitory intermediate states, to the bound state ensemble. Not all IDPs fold upon binding (Figure 7), as shown by the interaction between nuclear pore proteins—which fill the nuclear pore in a dense matrix of fluid IDPs—and nuclear transporters. Here, ultra-weak (>mM) interactions with the transporter are compensated by the extremely high concentration of the linear motifs within the pore. Similarly, although at the other end of the affinity scale, the ultra-tight (pM) interaction between two highly charged nuclear proteins, histone H1 and prothymosin-α, allows for the two IDPs to remain highly dynamic in the complex due to the strongly electrostatic nature of their interaction.[61]

NMR is the prime technology to unravel the mechanisms of LLPS, an emerging area of research. Indeed, LLPS is increasingly recognized as a crucial process in cellular biology, where it plays a role in the formation of membrane-less organelles and condensates, including nucleolus and stress granules.[62]

In-cell NMR

NMR techniques can be applied to living cells. Recent progress has enabled the studies of small and large biomolecules

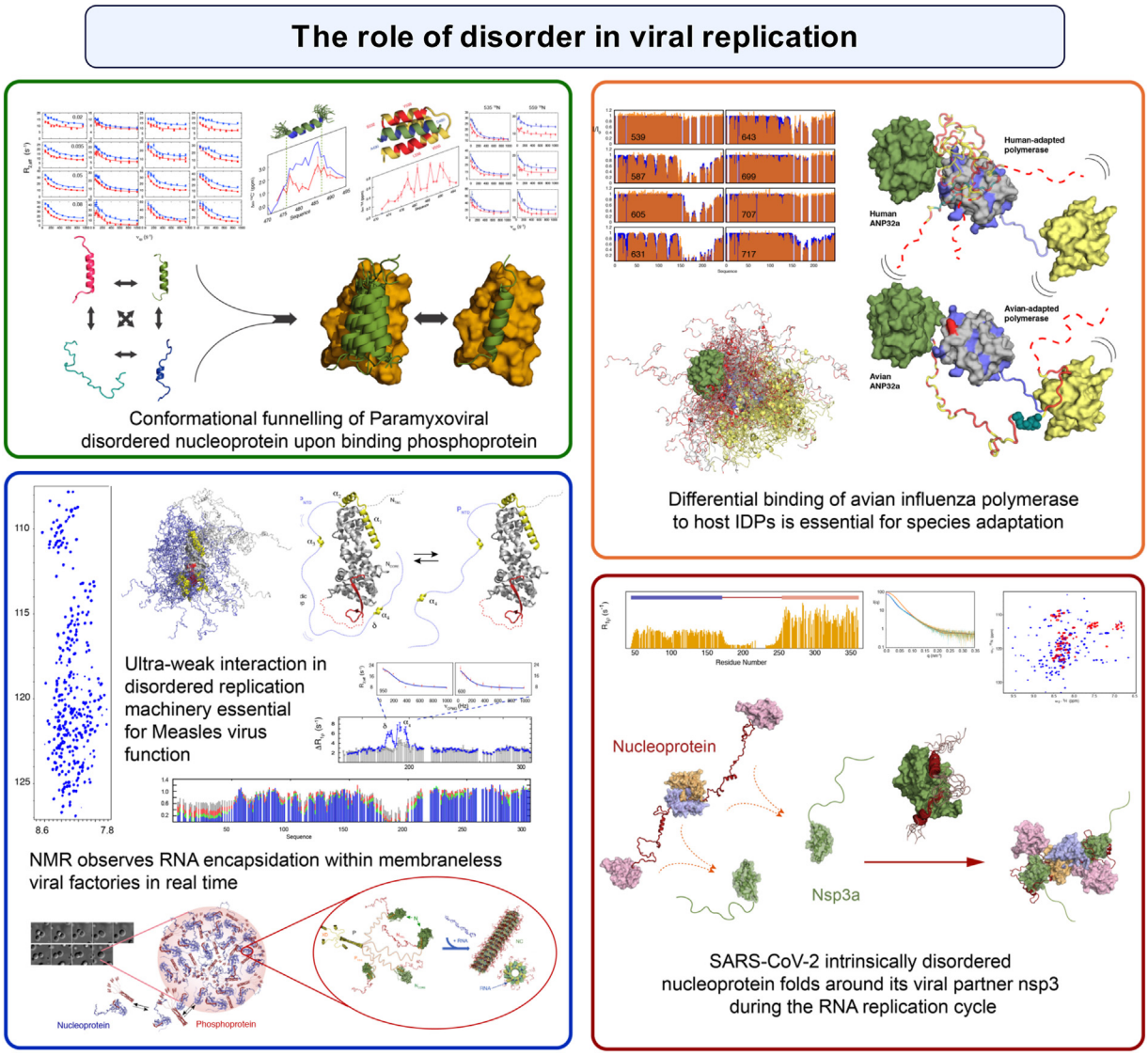

Figure 7. NMR characterization of the role of intrinsically disordered proteins in viral replication and liquid-liquid phase separation

(Top left) NMR exchange reveals the trajectory from the disordered free state ensemble of Sendai virus nucleoprotein to the phosphoprotein-bound state via conformational funneling. (Top right) Paramagnetic NMR reveals the differential binding of avian influenza polymerase to host IDPs that is essential for host adaptation. (Bottom right) Intrinsically disordered SARS-CoV-2 nucleoprotein folds around its viral partner during the RNA replication cycle. (Bottom right) Measles virus disordered phosphoprotein chaperones the nucleoprotein via an ultra-weak interaction prior to encapsidation of the viral RNA that occurs within membraneless viral factories.

in living cells—including bacteria, yeast, oocytes, and cultured insect/human cells—bridging classical structural and cellular biology methodologies.[63] In-cell NMR, therefore, has the potential to provide precious structural and mechanistic insights on pharmacologically relevant target proteins and nucleic acids in a highly physiological environment (Figure 8).

The main challenges that in-cell NMR approaches will face in the next five years are related to the need to (i) increase the intrinsically low sensitivity of NMR, (ii) move toward more physiologically relevant models of cells and tissues, and (iii) overcome the limitations of increased spin relaxation when studying molecules tumbling at slow rates due to their interaction with the cellular environment.

NMR of RNA

Different from the structure determination of proteins, NMR plays a unique role in the structure determination of RNAs. Since the year 2000, the importance of RNA beyond its role as an information carrier has been increasingly recognized and with the use of mRNA-based vaccines, it is now also in the focus as an emerging therapeutic agent and a new class of biologics. The unique position of NMR in determining RNA structure stems from the inherent properties of RNA to feature highly dynamic sites. Bulges and loops play similar roles in RNA as IDRs of proteins. These dynamic elements can be sites for hinge motions and are often the target elements for protein-RNA interactions. They can also be target sites for small molecule binding to RNA as non-A-form structural elements can feature specific

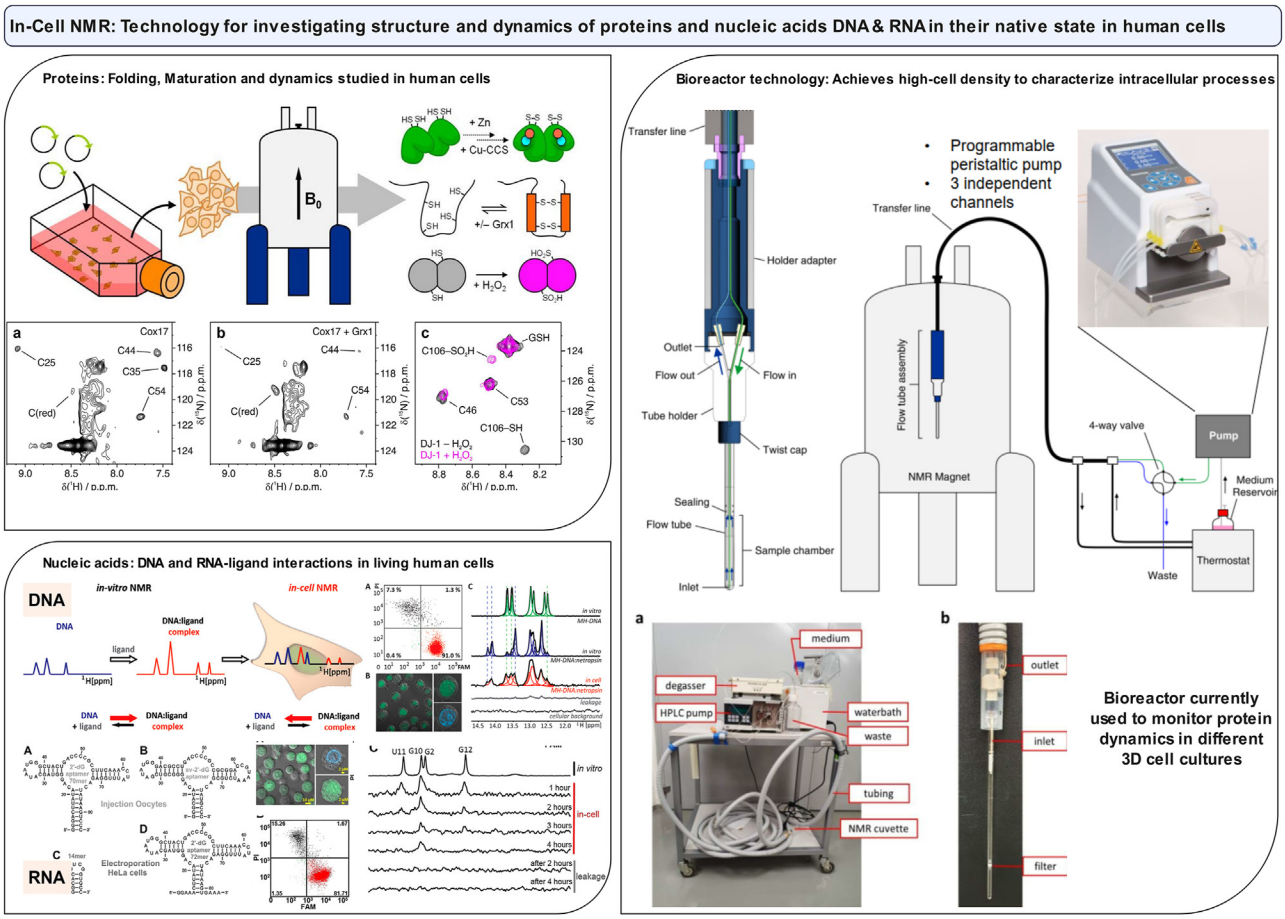

Figure 8. General setup for in-cell solution-state NMR studies on proteins and nucleic acids

(Top left) In-cell NMR in human cells. The redox state of intracellular [^{15}N]cysteine-labeled Cox17 monitored by in-cell NMR (graphic adapted from Luchinat E. et al.[64]; "This is an unofficial adaptation of an article that appeared in an ACS publication. ACS has not endorsed the content of this adaptation or the context of its use."). (Bottom left) In-cell NMR spectroscopy of functional 2′-deoxyguanosine riboswitch aptamer in eukaryotic cells (graphic adapted from Broft P. et al.[65]). (Right) Bioreactor technology coupled to NMR spectroscopy to study protein-ligand interaction[66] or protein dynamics in cells. (This graphic is adapted from Barbieri, L., Luchinat, E. Monitoring Protein-Ligand Interactions in Human Cells by Real-Time Quantitative In-Cell NMR using a High Cell Density Bioreactor. J. Vis. Exp. (169), e62323, https://doi.org/10.3791/62323 [2021].)

binding sites. Further, RNA chains have been difficult to trace in recent cryo-TM studies and methods to predict RNA structures are only now developed, in particular FARFAR.[67] One specific aspect for future structural studies of RNA is to understand the role of RNA modification on the structure, dynamics and function of RNA. To be able to conduct these studies, new sample preparation methodologies are constantly developed (see chapter 3).

NMR and drug discovery

Fragment-based drug discovery has been introduced by Fesik and colleagues in his seminal contribution entitled SAR by NMR, demonstrating how the covalent fusion of two fragments binding adjacent binding pockets of protein can lead to substantial improvement in binding affinity.[68] Despite these early developed concepts, their application in academia has been limited in scope as the size of fragment libraries have been limited, and follow-up medicinal chemistry campaigns have been scattered. This situation has changed, in particular due to the outbreak of COVID-19, where fragment-screening efforts have been intensified both by X-ray and NMR. In fact, the identical so-called DSI-poised library has been the starting point for massive structural fragment-screening campaigns (Figure 9). This library and other fragment libraries developed in other European RIs (EU-OPENSCREEN) are available for academic users through Instruct. One advantage of NMR over X-ray for fragment screening applications is the broader range of target classes that can be studied including both IDPs[71] as well as RNA.[69]

Molecular dynamics and AI approaches in integrated structural biology

Computational structural biology has come a long way in its mission to assist, inform and, in straightforward cases, even avoid experimental structure determination. Building on bioinformatics and more recently harnessing AI, methods are being developed for the prediction of protein-protein interfaces and the impact point mutations might have on them; the prediction of protein-ligand interactions and affinity between them; the functional analysis of post-translational modifications[22]; the characterization of existing proteins and the design of new ones. Beyond proteins, the prediction of RNA 3D structure will likely be an ongoing challenge for years to come. The now

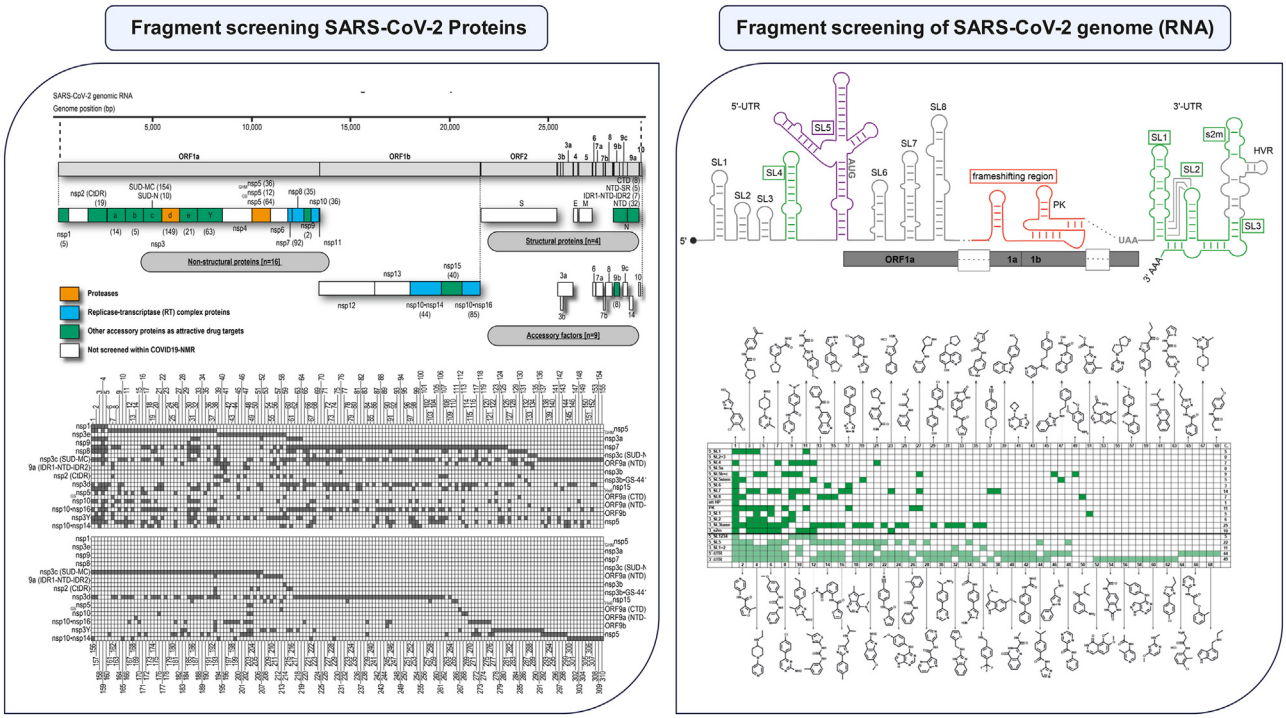

Figure 9. Fragment-based drug discovery applied to SARS-CoV-2 proteins and genome (RNA) by NMR spectroscopy within the Covid19-nmr project

Several fragment hits were identified against the screened targets. (Graphics adapted from Sreeramulu S et al.[69] and Berg H et al[70]).

ubiquitous use of deep learning—a form of AI—in the prediction of protein 3D folds, such as implemented in AlphaFold2,[72] RoseTTAfold,[24] and their myriad variants, may present scientists with workable structural hypotheses in many cases, although certainly not all.

Albeit impressive in their own right, these predictors have been trained on protein data from a database that contains mainly crystallographic structures determined to a typical resolution of 2 Å, and therefore, are strongly biased toward that; the fact that structures are structured means that the predictors never saw unstructured data as part of their training; thus their output will be reduced to speculation when folding IDPs or disordered domains. Moreover, the absence of important moieties that do get captured in experimental structures—co- and post-translational modifications, co-factors, and important ligands—is not easily addressed, as some of them are important for proteins to fold properly; currently, the best approach is to supplement these models' post-prediction with experimentally supported moieties.[73,74] At a larger scale, homo- and hetero-multimeric arrangements are still a problem for fold predictors, requiring experimental evidence to guide the generation of the most probable assemblies.

Molecular mechanics has recently been instrumental in understanding the dynamics of the fully glycosylated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike N-glycans, which accounts for a large portion of the structure (as solved by cryo-EM), is essential in the interplay between the receptor binding domain (RBD) and the rest of the spike's structure, therefore playing a key role in its function[75] (see Figure 2). At a more complex level, machine learning-driven multiscale modeling can help understand the interplay between local lipid mixtures in membranes and the multimerization of signaling a protein like RAS, revealing a mechanism for regulating cell signaling cascades.[76]

In the following, we will discuss the integration of experimental data with molecular dynamics simulations and AI for NMR spectroscopy and X-ray crystallography.

Integration of NMR with Molecular dynamics and AI

There are many interconnected growth areas in the field of NMR spectroscopy of biological (macro)molecules and they are intertwined. The first involves expanding the use of AI/deep learning techniques for a range of NMR data processing and analysis applications, while the second entails increasing the accessibility of high-quality data collected globally.

Regarding the development AI/DL methods for biomolecular NMR, the range of potential applications is wide, going from spectra processing, e.g., based on NUS data,[77] to peak picking and assignment in multi-dimensional NMR.[78] Further, the combination of self-consistent relaxation datasets reporting on motions up to tens of nanoseconds from diverse IDPs with extensive molecular simulation has already provided essential new insight into the nature of the dynamics of IDPs on these timescales and will surely provide appropriate benchmarks for future improvements in our ability to simulate IDP dynamics *in silico*. Similarly, interfacing NMR and MD simulation will be vital to improve RNA structure determination and simulation. While there are various proof-of-principles of the aforementioned concepts, their actual use in everyday laboratory practice is still in its infancy, and other themes in biomolecular NMR are still unaddressed.

Activities to improve NMR data availability would leverage existing collaborations with the BMRB database. The development of appropriate data and metadata standards to describe the experiments and their results would be required as part of efforts to increase the availability of NMR data. These standards would apply to both in-cell NMR and fragment screening campaigns as well as NMR structure determination, an area that is more developed but not yet fully satisfactory.

Integration of macromolecular structure determination with AI

Deep automation of sample handling, data acquisition, and data analysis in macromolecular crystallography (MX) has increased the efficiency and value of structural biology research infrastructures. New HT approaches like large-scale ligand and fragment screening are providing easier access to chemical tools for research and closer integration of fundamental and translational research. These new capacities require improved large-scale data processing and hit identification pipelines exploiting AI-based approaches. AlphaFold2 has largely solved the problem of predicting the initial fold for most proteins and this success has demonstrated that making large volumes of high-quality structural data available under open access is key to fueling future breakthroughs in AI-based research, opening it to new user communities. It will also produce large-scale datasets to fuel the future development of AI. This applies to all technologies and to all domains of structural biology and is particularly exciting in the field of drug design, where, for the first-time, large datasets produced by academic groups will be available in the public domain, providing the basis for future developments in AI-based drug design. We expect that there will also be a tremendous increase in the need to train on advanced usage of relevant software approaches and need to implement services allowing combining and integration of experimental and computational data to obtain optimized structural models. During the writing of this review, AlphaFold3 has been published and commented upon.[79,80] Strikingly, previously perceived gaps have been filled and conceptually new prediction methodology has been integrated. Especially the possibility to predict the structure of macromolecular complexes is a game changer, with numerous impact on understanding functional important protein-protein as well as protein-nucleic acid complexes, but also protein-ligand interactions.

Interfacing molecular dynamics with experimental data to describe macromolecular dynamics

Macromolecular dynamics is another area that can accelerate our understanding of the free energy landscapes of proteins and RNA, including different conformational states. Exciting possibilities afforded by cryo-EM include the generation of conformational ensembles from class-averaged maps or—with even bigger potential—truly exploiting the single-molecule nature of cryo-EM. This will require substantial developments in methods for computational analysis as well as getting the best possible experimental data.

As structural biology moves to larger assemblies and biologically more complex samples, it will be increasingly hard to generate "chemically pure" samples. For example, a large fraction of complexes on a cryo-EM grid may lack one or more subunits, and therefore we need better computational tools to deal with experimental data that are averaged over distinct species or assemblies.

Another area for important developments is the integration of experiments and computation to study time-resolved phenomena. Experimentally, there have been important developments in obtaining time-resolved data using NMR, crystallography, and cryo-EM. We still need computational methods to simulate these processes and refine the models against experiments.

One type of system that encompasses many of the issues discussed previously is biomolecular condensates or other systems that display similar behavior. These systems are often dynamic and chemically heterogeneous. Examples span from something like the nuclear pore complex to stress granules. Their dynamics make them challenging to study by crystallography and cryo-EM, and their large nature and slowish dynamics make NMR difficult. Ideally, we would even like to be able to study these in a native environment. To tackle these kinds of systems we need to combine all the methods we have and develop new ones.

FUTURE DIRECTIONS

Integrated structural biology will evolve in the same way as molecular biology did—it will be done routinely and will not be considered a "separate discipline." To reach this deep integration toward must-do research within any cell biology project will require addressing the following frontiers.

Frontier 1: Optimized sample preparation, deep automation, and broader availability of cryo-EM instrumentation

Some facilities, including X-ray and NMR facilities, will generate break-through approaches in the next few years, enabling high-density data acquisition. Know-how and expertise in automation will be transferred to the other applications, bringing them to the level of other facilities. Additionally, the widespread availability of much less expensive, high-resolution cryo-EM instruments will make them as easy to use as a desktop light microscope and more importantly, it will allow optimized sample preparation.

Impact

New complex scientific questions can be answered; while increase efficiency and higher output will broaden access to new communities, further necessitating pan-European access mechanisms.

Frontier 2: *In situ* cellular structural biology

Substantial parts of integrated structural biology will evolve into cellular structural biology, which will be performed not just on cultured cells in Petri dishes but on living multicellular organisms. This will open new horizons, with much of what was considered "true" before the *in vivo* revolution being discarded. Tomography, both cryo-EM and X-ray (at synchrotrons), will become the principal experimental approaches. These technologies will permit studying biomacromolecules in their native environment without first purifying the individual components and will be achievable at atomic resolution!

Cryo-EM, cryo-ET, X-ray tomography, and in-cell NMR will further develop, and the combined integrated use of a broad technology envelope will provide key breakthrough insight.

Future Direction: The Five Forces Analysis

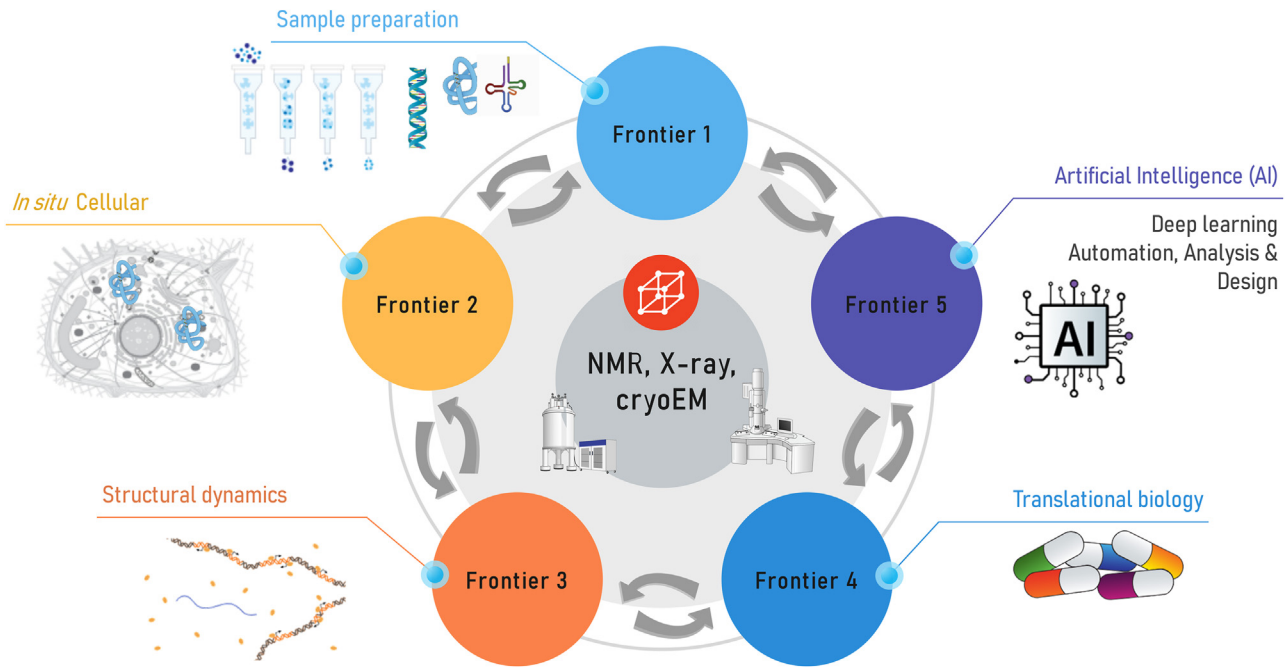

Figure 10. Integrated structural biology—A logical diagram illustrating the interplay between various technologies

Concepts like digital twins of cells to integrate also computational analysis data annotation, data integration will develop.

Impact

New scientific questions can be addressed by studying complex systems in a more biologically relevant context. Bridge size scales through collaboration within the imaging community represents a logical next step.

Frontier 3: Correlating structural dynamics with biological function

Integrated structural biology will further evolve into 4D integrated structural biology by holistically studying the dynamics of biomacromolecules and their complexes. It will become increasingly apparent that a static high-resolution structure tells us only a glimpse of the molecule's behavior. These studies will benefit from advanced NMR instruments, super-resolution microscopes, and potentially new, yet-to-be-developed instruments. Furthermore, ongoing advancements in computing power will accelerate molecular dynamics simulations, enabling deeper insights in a fraction of the current time. Such research will be conducted by an increasingly broad user community, with the recent technologies enabling exploration of the conformational space in complex systems as never before. Combination of experimental findings with computational and new biophysical approaches promises to deepen our understanding of dynamics.

Impact

Focusing on a tighter link between structural dynamics and biological function will lead to new understanding of proteins and RNA as intrinsically dynamic systems, influencing our understanding of their function and improving modeling and prediction capabilities.

Frontier 4: Bridging fundamental and applied research through translational biology

New frontiers include developing high-density applications for ligand screening, advancing instruments for higher sensitivity, adapting HT approaches from X-ray/NMR to cryo-EM, and facilitating the study of challenging systems in biological relevant environments.

Impact

Faster and easier access to translational oriented infrastructures for users in academia and industry. Stimulate cooperation between academy and industry. Strengthen links with emerging communities such as chemical biology and screening communities.

Frontier 5: Harnessing the data revolution, AI in structural biology

Integrated structural biology will become a big data-driven discipline, with hundreds of millions of molecular structures accurately determined or predicted. Access to research infrastructures will be key to reach this vision. The integration of AI-driven structure prediction with experimental methods in integrated structural biology will significantly expand its impact across various fields, including evolution, disease diagnosis and treatment, developmental biology, and our overall understanding of life processes. To enable non-experts to conduct

integrated structural biology projects on a routine basis, easy access to research infrastructures with the help of dedicated staff will be essential.

Impact

Documentation of improvements and the necessity of experimental validation of AI predictions would further strengthen the importance of integrated structural biology.

Converging technological advancements in all areas (Figure 10) of integrated structural biology pave the way for exciting science, particularly considering challenges such as increasing life expectancy, global pandemics, and climate change. As these issues transcend national boundaries, global solutions are imperative, highlighting the importance of transnational research collaboration. Within Europe, the impact of such collaboration on European citizens is a major determinant for European Research Infrastructure Consortia. The implementation of open science principles is democratizing research and facilitating rapid access to technologies and resources, as well as the rapid distribution of key insights. It is also worth noting that there are challenges confronting sharing data and services across international borders. As data and resources become more distributed and increasingly are accessed in the cloud, export sanctions (for example against investigators from Russia) and regulations such as the General Data Protection Regulation (GDPR) pose challenges to open science. Maintaining open science policies is, however, crucial for the continued welfare of our global society.

ACKNOWLEDGMENTS

Instruct hub: this work was supported by European Union's Horizon 2020 and Horizon Europe research and innovation programmes under grant agreement EOSC-Life No 824087, iNEXT-Discovery No 871037, Fragment-Screen No 10194131, BY-COVID No 101046203, and EOSC4Cancer 101058427. Work at BMRZ was supported by the state of Hesse. H.S. also acknowledges European Union's Horizon 2020 research and innovation program iNEXT-discovery under grant agreement No 871037 and by European Union's Horizon 2020 research and innovation program Fragment-Screen under grant agreement No 101094131. IBS (D.J.H. and M.B.) acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). The support of the Italian Ministry for University and Research (MUR, FOE funding) to the CERM/CIRMMP center (L.B., R.P., A.R., and I.F.) of Instruct-ERIC is gratefully acknowledged. This work is part of the project "Potentiating the Italian Capacity for Structural Biology Services in Instruct-ERIC (ITACA.SB, Project no. IR0000009) within the call MUR 3264/2021 PNRR M4/C2/L3.1.1, funded by the European Union – Next Generation EU. J.A. is a Royal Society University Research Fellow (awards UF160039 and URF\R\221006). Work at CEITEC (L.T.) was supported by the National Institute for Cancer Research (Project No. LX22NPO5102), funded by the European Union – Next Generation EU and by the EU project Fragment-Screen, grant agreement ID: 101094131. JMC acknowledges the financial support from the Ministry of Science, Innovation and Universities (BDNS n. 716450) to the Instruct Image Processing Center (I2PC) as part of the Spanish participation in Instruct – ERIC, the European Strategic Infrastructure Project (ESFRI) in the area of Structural Biology. L.L., Biocenter Oulu Structural Biology core facility at the University of Oulu, is a member of Biocenter Finland, Instruct-ERIC Centre, Finland, and FINStruct and the work was supported by the Research Council of Finland (grant no. 345512).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Lubin, J.H., Zardecki, C., Dolan, E.M., Lu, C., Shen, Z., Dutta, S., Westbrook, J.D., Hudson, B.P., Goodsell, D.S., Williams, J.K., et al. (2022). Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. Proteins *90*, 1054–1080. https://doi.org/10.1002/prot.26250.

2. Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science *370*, 203–208. https://doi.org/10.1126/science.abd5223.

3. Negi, S.S., Schein, C.H., and Braun, W. (2022). Regional and temporal coordinated mutation patterns in SARS-CoV-2 spike protein revealed by a clustering and network analysis. Sci. Rep. *12*, 1128. https://doi.org/10.1038/s41598-022-04950-4.

4. Walsh, M.A., Grimes, J.M., and Stuart, D.I. (2021). Diamond Light Source: contributions to SARS-CoV-2 biology and therapeutics. Biochem. Biophys. Res. Commun. *538*, 40–46. https://doi.org/10.1016/j.bbrc.2020.11.041.

5. Falcon, B., Zivanov, J., Zhang, W., Murzin, A.G., Garringer, H.J., Vidal, R., Crowther, R.A., Newell, K.L., Ghetti, B., Goedert, M., et al. (2019). Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. Nature *568*, 420–423. https://doi.org/10.1038/s41586-019-1026-5.

6. Ho, C.M., Li, X., Lai, M., Terwilliger, T.C., Beck, J.R., Wohlschlegel, J., Goldberg, D.E., Fitzpatrick, A.W.P., and Zhou, Z.H. (2020). Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. Nat. Methods *17*, 79–85. https://doi.org/10.1038/s41592-019-0637-y.

7. Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., Su, Z., de Oliveira, G.A.P., Osinski, T., Sherman, N., DiMaio, F., Wall, J.S., Prangishvili, D., et al. (2019). An extensively glycosylated archaeal pilus survives extreme conditions. Nat. Microbiol. *4*, 1401–1410. https://doi.org/10.1038/s41564-019-0458-x.

8. Pradhan, B., Liedtke, J., Sleutel, M., Lindbäck, T., Zegeye, E.D., ÓSullivan, K., Llarena, A.K., Brynildsrud, O., Aspholm, M., and Remaut, H. (2021). Endospore Appendages: a novel pilus superfamily from the endospores of pathogenic Bacilli. EMBO J. *40*, e106887. https://doi.org/10.15252/embj.2020106887.

9. Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., Crowther, R.A., Ghetti, B., Goedert, M., and Scheres, S.H.W. (2017). Cryo-EM structures of tau filaments from Alzheimer's disease. Nature *547*, 185–190. https://doi.org/10.1038/nature23002.

10. Ioannidis, J.P.A., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., and Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. Lancet *383*, 166–175. https://doi.org/10.1016/S0140-6736(13)62227-8.

11. Pannus, P., Neven, K.Y., De Craeye, S., Heyndrickx, L., Vande Kerckhove, S., Georges, D., Michiels, J., Francotte, A., Van Den Bulcke, M., Zrein, M., et al. (2022). Poor Antibody Response to BioNTech/Pfizer Coronavirus Disease 2019 Vaccination in Severe Acute Respiratory Syndrome Coronavirus 2-Naive Residents of Nursing Homes. Clin. Infect. Dis. *75*, E695–E704. https://doi.org/10.1093/cid/ciab998.

12. Tomasi, L., Thiriard, A., Heyndrickx, L., Georges, D., Van Den Wijngaert, S., Olislagers, V., Sharma, S., Matagne, A., Ackerman, M.E., Ariën, K.K., et al. (2022). Younger Children Develop Higher Effector Antibody Responses to SARS-CoV-2 Infection. Open Forum Infect. Dis. *9*, ofac554. https://doi.org/10.1093/ofid/ofac554.

13. Kemlin, D., Gemander, N., Depickère, S., Olislagers, V., Georges, D., Waegemans, A., Pannus, P., Lemy, A., Goossens, M.E., Desombere, I., et al. (2023). Humoral and cellular immune correlates of protection against COVID-19 in kidney transplant recipients. Am. J. Transplant. *23*, 649–658. https://doi.org/10.1016/j.ajt.2023.02.015.

14. Cawez, F., Duray, E., Hu, Y., Vandenameele, J., Romão, E., Vincke, C., Dumoulin, M., Galleni, M., Muyldermans, S., and Vandevenne, M. (2018). Combinatorial Design of a Nanobody that Specifically Targets Structured RNAs. J. Mol. Biol. *430*, 1652–1670. https://doi.org/10.1016/j.jmb.2018.03.032.

15. Cawez, F., Mercuri, P.S., Morales-Yãnez, F.J., Maalouf, R., Vandevenne, M., Kerff, F., Guérin, V., Mainil, J., Thiry, D., Saulmont, M., et al. (2023). Development of Nanobodies as Theranostic Agents against CMY-2-Like

Class C b-Lactamases. Antimicrob. Agents Chemother. *67*, e0149922. https://doi.org/10.1128/aac.01499-22.

16. De Franco, S., Vandenameele, J., Brans, A., Verlaine, O., Bendak, K., Damblon, C., Matagne, A., Segal, D.J., Galleni, M., Mackay, J.P., and Vandevenne, M. (2019). Exploring the suitability of RanBP2-type Zinc Fingers for RNA-binding protein design. Sci. Rep. *9*, 2484. https://doi.org/10.1038/s41598-019-38655-y.

17. Robert, C., Kerff, F., Bouillenne, F., Gavage, M., Vandevenne, M., Filée, P., and Matagne, A. (2023). Structural analysis of the interaction between human cytokine BMP-2 and the antagonist Noggin reveals molecular details of cell chondrogenesis inhibition. J. Biol. Chem. *299*, 102892. https://doi.org/10.1016/j.jbc.2023.102892.

18. Wang, J., Kang, G., Yuan, H., Cao, X., Huang, H., and de Marco, A. (2021). Research Progress and Applications of Multivalent, Multispecific and Modified Nanobodies for Disease Treatment. Front. Immunol. *12*, 838082. https://doi.org/10.3389/fimmu.2021.838082.

19. Morris, O.M., Torpey, J.H., and Isaacson, R.L. (2021). Intrinsically disordered proteins: Modes of binding with emphasis on disordered domains. Open Biol. *11*, 210222. https://doi.org/10.1098/rsob.210222.

20. Kamacioglu, A., Tuncbag, N., and Ozlu, N. (2021). Structural analysis of mammalian protein phosphorylation at a proteome level. Structure *29*, 1219–1229.e3. https://doi.org/10.1016/j.str.2021.06.008.

21. Jonker, H.R.A., Saxena, K., Shcherbakova, A., Tiemann, B., Bakker, H., and Schwalbe, H. (2020). NMR Spectroscopic Characterization of the C-Mannose Conformation in a Thrombospondin Repeat Using a Selective Labeling Approach. Angew. Chem. Int. Ed. Engl. *59*, 20659–20665. https://doi.org/10.1002/anie.202009489.

22. Bludau, I., Willems, S., Zeng, W.F., Strauss, M.T., Hansen, F.M., Tanzer, M.C., Karayel, O., Schulman, B.A., and Mann, M. (2022). The structural context of posttranslational modifications at a proteome-wide scale. PLoS Biol. *20*, e3001636. https://doi.org/10.1371/journal.pbio.3001636.

23. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

24. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. https://doi.org/10.1126/science.abj8754.

25. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

26. Blümler, A., Schwalbe, H., and Heckel, A. (2022). Solid-Phase-Supported Chemoenzymatic Synthesis of a Light-Activatable tRNA Derivative. Angew. Chemie - Int. Ed. *61*, 10. https://doi.org/10.1002/anie.202111613.

27. Keyhani, S., Goldau, T., Blümler, A., Heckel, A., and Schwalbe, H. (2018). Chemo-Enzymatic Synthesis of Position-Specifically Modified RNA for Biophysical Studies including Light Control and NMR Spectroscopy. Angew. Chemie - Int. Ed. *57*, 12017–12021. https://doi.org/10.1002/anie.201807125.

28. Sudakov, A., Knezic, B., Hengesbach, M., Fürtig, B., Stirnal, E., and Schwalbe, H. (2023). Site-Specific Labeling of RNAs with Modified and 19F-Labeled Nucleotides by Chemo-Enzymatic Synthesis. Chem. Eur J. *29*, e202203368. https://doi.org/10.1002/chem.202203368.

29. Britt, H.M., Cragnolini, T., and Thalassinos, K. (2022). Integration of Mass Spectrometry Data for Structural Biology. Chem. Rev. *122*, 7952–7986. https://doi.org/10.1021/acs.chemrev.1c00356.

30. Grahl, A., Abiko, L.A., Isogai, S., Sharpe, T., and Grzesiek, S. (2020). A high-resolution description of β1-adrenergic receptor functional dynamics and allosteric coupling from backbone NMR. Nat. Commun. *11*, 2216. https://doi.org/10.1038/s41467-020-15864-y.

31. Boby, M.L., Fearon, D., Ferla, M., Filep, M., Koekemoer, L., Robinson, M.C., COVID Moonshot Consortium‡, Chodera, J.D., Lee, A.A., London, N., et al. (2023). Open science discovery of potent noncovalent SARS-CoV-2 main protease inhibitors. Science *382*, eabo7201. https://doi.org/10.1126/science.abo7201.

32. Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C.D., Resnick, E., Strain-Damerell, C., Aimon, A., Ábrányi-Balogh, P., et al. (2020). Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. Nat. Commun. *11*, 5047. https://doi.org/10.1038/s41467-020-18709-w.

33. Schuller, M., Correy, G.J., Gahbauer, S., Fearon, D., Wu, T., Díaz, R.E., Young, I.D., Carvalho Martins, L., Smith, D.H., Schulze-Gahmen, U., et al. (2021). SARS-CoV-2 identified through crystallographic screening and computational docking. Sci. Adv. *7*, eabf8711. https://doi.org/10.1126/sciadv.abf8711.

34. Newman, J.A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão-Neto, J., Dunnett, L., Gorrie-stone, T., Skyner, R., Fearon, D., et al. (2021). Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. Nat. Commun. *12*, 4848. https://doi.org/10.1038/s41467-021-25166-6.

35. Daniel, E., Maksimainen, M.M., Smith, N., Ratas, V., Biterova, E., Murthy, S.N., Rahman, M.T., Kiema, T.R., Sridhar, S., Cordara, G., et al. (2021). IceBear: An intuitive and versatile web application for research-data tracking from crystallization experiment to PDB deposition. Acta Crystallogr. D Struct. Biol. *77*, 151–163. https://doi.org/10.1107/S2059798320015223.

36. Cornaciu, I., Bourgeas, R., Hoffmann, G., Dupeux, F., Humm, A.S., Mariaule, V., Pica, A., Clavel, D., Seroul, G., Murphy, P., and Márquez, J.A. (2021). The automated crystallography pipelines at the embl htx facility in grenoble. J. Vis. Exp. *172*, e62491. https://doi.org/10.3791/62491.

37. Delagenière, S., Brenchereau, P., Launer, L., Ashton, A.W., Leal, R., Veyrier, S., Gabadinho, J., Gordon, E.J., Jones, S.D., Levik, K.E., et al. (2011). ISPyB: An information management system for synchrotron macromolecular crystallography. Bioinformatics *27*, 3186–3192. https://doi.org/10.1093/bioinformatics/btr535.

38. Xue, L., Lenz, S., Zimmermann-Kogadeeva, M., Tegunov, D., Cramer, P., Bork, P., Rappsilber, J., and Mahamid, J. (2022). Publisher Correction: Visualizing translation dynamics at atomic detail inside a bacterial cell. Nature *611*, E13. https://doi.org/10.1038/s41586-022-05455-w.

39. Xing, H., Taniguchi, R., Khusainov, I., Kreysing, J.P., Welsch, S., Turoňová, B., and Beck, M. (2023). Translation dynamics in human cells visualized at high resolution reveal cancer drug action. Science *381*, 70–75. https://doi.org/10.1126/science.adh1411.

40. Grey, C.P., and Tarascon, J.M. (2016). Sustainability and in situ monitoring in battery development. Nat. Mater. *16*, 45–56. https://doi.org/10.1038/nmat4777.

41. Sekhar, A., and Kay, L.E. (2019). An NMR View of Protein Dynamics in Health and Disease. Annu. Rev. Biophys. *48*, 297–319. https://doi.org/10.1146/annurev-biophys-052118-115647.

42. Ganser, L.R., Kelly, M.L., Herschlag, D., and Al-Hashimi, H.M. (2019). The roles of structural dynamics in the cellular functions of RNAs. Nat. Rev. Mol. Cell Biol. *20*, 474–489. https://doi.org/10.1038/s41580-019-0136-0.

43. Halvorsen, M., Martin, J.S., Broadaway, S., and Laederach, A. (2010). Disease-associated mutations that alter the RNA structural ensemble. PLoS Genet. *6*, e1001074. https://doi.org/10.1371/journal.pgen.1001074.

44. Alderson, T.R., and Kay, L.E. (2021). NMR spectroscopy captures the essential role of dynamics in regulating biomolecular function. Cell *184*, 577–595. https://doi.org/10.1016/j.cell.2020.12.034.

45. Boehr, D.D., McElheny, D., Dyson, H.J., and Wright, P.E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. Science *313*, 1638–1642. https://doi.org/10.1126/science.1130258.

46. Kerns, S.J., Agafonov, R.V., Cho, Y.J., Pontiggia, F., Otten, R., Pachov, D.V., Kutter, S., Phung, L.A., Murphy, P.N., Thai, V., et al. (2015). The energy landscape of adenylate kinase during catalysis. Nat. Struct. Mol. Biol. *22*, 124–131. https://doi.org/10.1038/nsmb.2941.

47. Xie, T., Saleh, T., Rossi, P., and Kalodimos, C.G. (2020). Conformational states dynamically populated by a kinase determine its function. Science *370*, eabc2754. https://doi.org/10.1126/science.abc2754.

48. Saio, T., Guan, X., Rossi, P., Economou, A., and Kalodimos, C.G. (2014). Structural basis for protein antiaggregation activity of the trigger factor

chaperone. Science *344*, 1250494. https://doi.org/10.1126/science.1250494.

49. Karamanos, T.K., and Clore, G.M. (2022). Large Chaperone Complexes Through the Lens of Nuclear Magnetic Resonance Spectroscopy. Annu. Rev. Biophys. *51*, 223–246. https://doi.org/10.1146/annurev-biophys-090921-120150.

50. Werbeck, N.D., Shukla, V.K., Kunze, M.B.A., Yalinca, H., Pritchard, R.B., Siemons, L., Mondal, S., Greenwood, S.O.R., Kirkpatrick, J., Marson, C.M., and Hansen, D.F. (2020). A distal regulatory region of a class I human histone deacetylase. Nat. Commun. *11*, 3841. https://doi.org/10.1038/s41467-020-17610-w.

51. Shukla, V.K., Siemons, L., and Hansen, D.F. (2023). Intrinsic structural dynamics dictate enzymatic activity and inhibition. Proc. Natl. Acad. Sci. USA *120*, e2310910120. https://doi.org/10.1073/pnas.2310910120.

52. Zhao, B., Guffy, S.L., Williams, B., and Zhang, Q. (2017). An excited state underlies gene regulation of a transcriptional riboswitch. Nat. Chem. Biol. *13*, 968–974. https://doi.org/10.1038/nchembio.2427.

53. Baronti, L., Guzzetti, I., Ebrahimi, P., Friebe Sandoz, S., Steiner, E., Schlagnitweit, J., Fromm, B., Silva, L., Fontana, C., Chen, A.A., and Petzold, K. (2020). Base-pair conformational switch modulates miR-34a targeting of Sirt1 mRNA. Nature *583*, 139–144. https://doi.org/10.1038/s41586-020-2336-3.

54. Nikolova, E.N., Kim, E., Wise, A.A., O'Brien, P.J., Andricioaei, I., and Al-Hashimi, H.M. (2011). Transient Hoogsteen base pairs in canonical duplex DNA. Nature *470*, 498–502. https://doi.org/10.1038/nature09775.

55. Xu, Y., McSally, J., Andricioaei, I., and Al-Hashimi, H.M. (2018). Modulation of Hoogsteen dynamics on DNA recognition. Nat. Commun. *9*, 1473. https://doi.org/10.1038/s41467-018-03516-1.

56. Tompa, P., Davey, N.E., Gibson, T.J., and Babu, M.M. (2014). A Million peptide motifs for the molecular biologist. Mol. Cell *55*, 161–169. https://doi.org/10.1016/j.molcel.2014.05.032.

57. Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. Biochim. Biophys. Acta *1804*, 1231–1264. https://doi.org/10.1016/j.bbapap.2010.01.017.

58. Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. *6*, 197–208. https://doi.org/10.1038/nrm1589.

59. Tompa, P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci. *27*, 527–533. https://doi.org/10.1016/S0968-0004(02)02169-2.

60. Uversky, V.N. (2002). Natively unfolded proteins: A point where biology waits for physics. Protein Sci. *11*, 739–756. https://doi.org/10.1110/ps.4210102.

61. Borgia, A., Borgia, M.B., Bugge, K., Kissling, V.M., Heidarsson, P.O., Fernandes, C.B., Sottini, A., Soranno, A., Buholzer, K.J., Nettels, D., et al. (2018). Extreme disorder in an ultrahigh-affinity protein complex. Nature *555*, 61–66. https://doi.org/10.1038/nature25762.

62. Fuxreiter, M., and Vendruscolo, M. (2021). Generic nature of the condensed states of proteins. Nat. Cell Biol. *23*, 587–594. https://doi.org/10.1038/s41556-021-00697-8.

63. Luchinat, E., Cremonini, M., and Banci, L. (2022). Radio Signals from Live Cells: The Coming of Age of In-Cell Solution NMR. Chem. Rev. *122*, 9267–9306. https://doi.org/10.1021/acs.chemrev.1c00790.

64. Luchinat, E., and Banci, L. (2018). In-Cell NMR in Human Cells: Direct Protein Expression Allows Structural Studies of Protein Folding and Maturation. Acc. Chem. Res. *51*, 1550–1557. https://doi.org/10.1021/acs.accounts.8b00147.

65. Broft, P., Dzatko, S., Krafcikova, M., Wacker, A., Hänsel-Hertsch, R., Dötsch, V., Trantirek, L., and Schwalbe, H. (2021). In-Cell NMR Spectroscopy of Functional Riboswitch Aptamers in Eukaryotic Cells. Angew. Chem. Int. Ed. Engl. *60*, 865–872. https://doi.org/10.1002/anie.202007184.

66. Barbieri, L., and Luchinat, E. (2021). Monitoring protein-ligand interactions in human cells by real-time quantitative in-cell nmr using a high cell density bioreactor. J. Vis. Exp. *169*, e62323. https://doi.org/10.3791/62323.

67. Kladwang, W., VanLang, C.C., Cordero, P., and Das, R. (2011). A two-dimensional mutate-and-map strategy for non-coding RNA structure. Nat. Chem. *3*, 954–962. https://doi.org/10.1038/nchem.1176.

68. Shuker, S.B., Hajduk, P.J., Meadows, R.P., and Fesik, S.W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. Science *274*, 1531–1534. https://doi.org/10.1126/science.274.5292.1531.

69. Sreeramulu, S., Richter, C., Berg, H., Wirtz Martin, M.A., Ceylan, B., Matzel, T., Adam, J., Altincekic, N., Azzaoui, K., Bains, J.K., et al. (2021). Exploring the Druggability of Conserved RNA Regulatory Elements in the SARS-CoV-2 Genome. Angew. Chem. Int. Ed. Engl. *60*, 19191–19200. https://doi.org/10.1002/anie.202103693.

70. Berg, H., Wirtz Martin, M.A., Altincekic, N., Alshamleh, I., Kaur Bains, J., Blechar, J., Ceylan, B., de Jesus, V., Dhamotharan, K., Fuks, C., et al. (2022). Comprehensive Fragment Screening of the SARS-CoV-2 Proteome Explores Novel Chemical Space for Drug Development. Angew. Chem. Int. Ed. Engl. *61*, e202205858. https://doi.org/10.1002/anie.202205858.

71. Altincekic, N., Korn, S.M., Qureshi, N.S., Dujardin, M., Ninot-Pedrosa, M., Abele, R., Abi Saad, M.J., Alfano, C., Almeida, F.C.L., Alshamleh, I., et al. (2021). Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. Front. Mol. Biosci. *8*, 653148. https://doi.org/10.3389/fmolb.2021.653148.

72. Jumper, J., and Hassabis, D. (2022). Protein structure predictions to atomic accuracy with AlphaFold. Nat. Methods *19*, 11–12. https://doi.org/10.1038/s41592-021-01362-6.

73. Bagdonas, H., Fogarty, C.A., Fadda, E., and Agirre, J. (2021). The case for post-predictional modifications in the AlphaFold Protein Structure Database. Nat. Struct. Mol. Biol. *28*, 869–870. https://doi.org/10.1038/s41594-021-00680-9.

74. Hekkelman, M.L., de Vries, I., Joosten, R.P., and Perrakis, A. (2023). AlphaFill: enriching AlphaFold models with ligands and cofactors. Nat. Methods *20*, 205–213. https://doi.org/10.1038/s41592-022-01685-y.

75. Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., Mclellan, J.S., et al. (2020). Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. ACS Cent. Sci. *6*, 1722–1734. https://doi.org/10.1021/acscentsci.0c01056.

76. Ingolfsson, H.I., Neale, C., Carpenter, T.S., Shrestha, R., Lopez, C.A., Tran, T.H., Oppelstrup, T., Bhatia, H., Stanton, L.G., Zhang, X., et al. (2022). Machine learning–driven multiscale modeling reveals lipid-dependent dynamics of RAS signaling proteins. Proc. Natl. Acad. Sci. USA *119*, e2113297119. https://doi.org/10.1073/pnas.2113297119.

77. Qu, X., Huang, Y., Lu, H., Qiu, T., Guo, D., Agback, T., Orekhov, V., and Chen, Z. (2020). Accelerated Nuclear Magnetic Resonance Spectroscopy with Deep Learning. Angew. Chem. Int. Ed. Engl. *59*, 10297–10300. https://doi.org/10.1002/anie.201908162.

78. Klukowski, P., Riek, R., and Güntert, P. (2022). Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. Nat. Commun. *13*, 33875–33879. https://doi.org/10.1038/s41467-022-33879-5.

79. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

80. Roy, R., and Al-Hashimi, H.M. (2024). AlphaFold3 takes a step toward decoding molecular behavior and biological computation. Nat. Struct. Mol. Biol. *31*, 997–1000. https://doi.org/10.1038/s41594-024-01350-2.