Bioorg. Med. Chem. 112 (2024) 117873

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Let's get biophysical – *How to get your favorite protein's digits*☆

Marcus J.C. Long [b,*], Yimon Aye [a,*]

[a] *Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne 1015, Switzerland*
[b] *University of Lausanne (UNIL), Epalinges 1066, Switzerland*

ARTICLE INFO

Keywords:
Biophysical tools
Protein structure and function
Interdisciplinary research

ABSTRACT

In these days of information overload and high-throughput analysis, it is easy to lose focus on the study of individual proteins. It is our conjecture that such investigations are still crucially important and offer uniquely penetrative insights. We thus present a discussion of biophysical methods to allow readers to get to know their protein of interest better. Although this perspective is not written with the expert in mind, we hope that for interdisciplinary scientists, or researchers who do not routinely perform biophysical analyses, the content will be helpful and inspiring.

Protein biochemistry remains the meat and potatoes of biological and medicinal research, and indeed, any disciplines that harness purified proteins to understand interactions at the molecular level. With increased power of resolution and detection, similar experiments are now also routinely used on partially-purified protein mixtures isolated from cells or tissues. Regardless of the origin of the samples to be analyzed, or the goals of the experiment, biophysical methods have their own inherent limits, and pros and cons that need to be considered both during planning of experiments and subsequent interpretation of results.

In this mini-review, we briefly discuss techniques we and others use to understand structure, binding, and activity/function of purified proteins (either recombinantly expressed or native entities isolated from cells). We focus on key scenarios that these methods can address, and where they may be limited in delivering a holistic picture. We hope such discussions will aid researchers—particularly interdisciplinary scientists and chemical biologists alike, not specialized in structural biology/biophysics research, but nonetheless wishing to exploit these tools—in deciding which methods could best match their specific research questions. Indeed, biophysical techniques are particularly useful to get up close and personal with proteins in which one has specific interest. Such an interest in these proteins could come from high throughput screens, literature searches, or through less formal means such as interactions during scientific symposia. Nonetheless, it is likely that once we start to hone in on studying a specific protein, some of our publications and key conclusions will hinge on correct biophysical characterization! Thus, this opinion piece, also set to a menu for a 3-course dinner to spur the broader readership, is written with the spirit that it could help make that happen for researchers if and where applicable.

1. Prelude – *protein isolation*

We will focus here on how to characterize specific proteins and not deal too much with purification. However, we note that, before the characterization of a protein starts, someone has to prepare it. Protein purification, although often considered routine, is indeed a complex skill that requires experience and care. We have purified recombinant proteins from *E. coli*,[1,2] but also used expression in mammalian systems, from which we have extracted sufficient protein for negative stain EM studies and activity assays.[3] The majority of these studies relied on His_6 tagged proteins and their affinity to nickel or cobalt beads. Although we have also used small domains, such as glutathione-S-transferase[4] and *E. coli* dihydrofolate reductase in some instances.[5] In general for relatively small (<~100 kDa), non post-translationally modified proteins, *E. coli* tends to be an ideal organism for preparation; several different *E. coli* strains are available to aid preparation of specific proteins, including strains aiding of folding of proteins at low temperatures,[6] and strains with codons apposite for mammalian expression.[7] For larger proteins, those requiring special folding procedures, and those that are post-translationally modified, mammalian or insect cells are often needed. For instance, in a recent paper we found that only expression in mammalian cells was able to afford enzymatically active Cyp-33e1 oxidoreductase.[8]

In terms of purity, it is also worth noting that different methods require different levels of purity. Indeed, enzymatic activity assays are

☆ Contribution to the Special Issue dedicated to Professor Chuan He, Winner of the 2023 Tetrahedron Prize for Creativity in Bioorganic & Medicinal Chemistry.

* Corresponding authors.
E-mail addresses: marcusjohncurtis.long@unil.ch (M.J.C. Long), yimon.aye@epfl.ch (Y. Aye).

https://doi.org/10.1016/j.bmc.2024.117873
Received 9 June 2024; Received in revised form 31 July 2024; Accepted 12 August 2024
Available online 13 August 2024
0968-0896/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

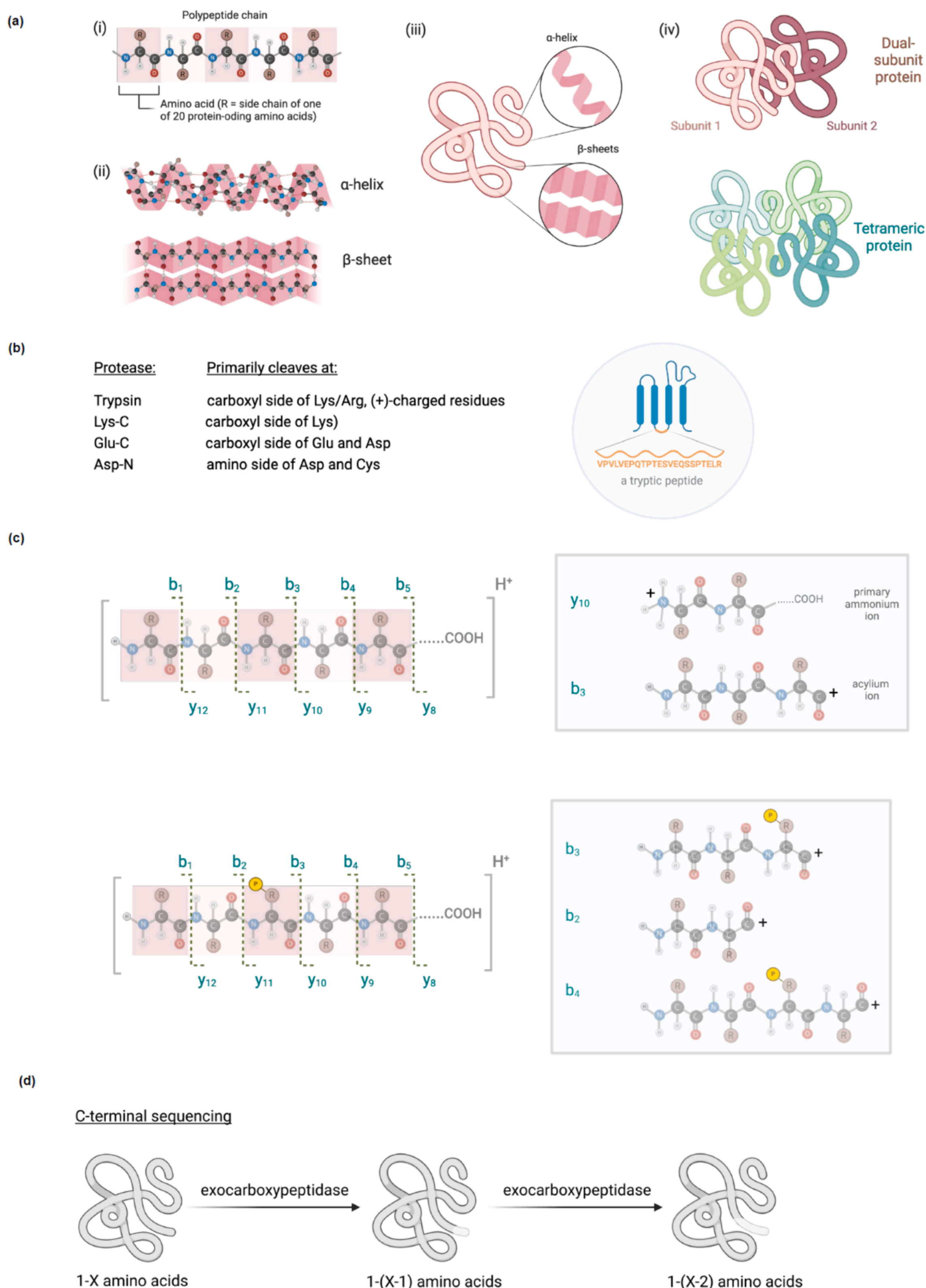

Figure 1. Protein structure and mass spectrometry. (a) Depictions of protein primary (i), secondary (ii), tertiary (iii), and quaternary (iv) structures. (b) Different proteases used in digest MS experiments. Inset (cricle) on right shows a representative tryptic peptide, wherein the last residue must be either arginine (as in this example) or lysine. (c) y and b ions resulting from collision-induced dissociation in tandem MS identification, examplified using a resulting 12-mer peptide post digestion. Top row: inset on right shows a representative fragmented ammonium and acylium ions, respectively, arbitrarily shown for y_{10} and b_3 ions. Lower row: modification site identification, exemplified using phosphorylation (yellow sphere) at the 3rd residue of a 12-mer peptide post digestion: consecutive fragments flanking a modified residue (here, b_2, b_3, and b_4 ions, see inset on right) are needed to confidently assign position of a modification. (d) C-terminal sequencing by exocarboxypeptidase and intact MS. Biorender software is used for illustrations.

typically very tolerant to contaminant proteins, providing those contaminants do not contribute to the measured activity. Other methods, such as crystallography can require highly homogeneous protein, which can be limiting in some cases. Purity of protein can depend on numerous factors including washing steps during purification, the nature of resin used,[4] and other factors beyond the scope of this publication.

Of course, many proteins (>70 % in eukaryotes) consist of several domains.[9] Computational methods to detect individual domains are known. Each domain, or other truncations, can often be purified and studied indepedently, greatly simplifying preparation, and analysis. It very much depends on the nature of the protein and the research question whether this approach is applicable or not; but should purification of the full protein otherwise be impossible, domain purification can constitute a good starting point. Alternatively, swapping similar proteins with different domains, either proteins from different species, or different isoforms, can also bring new insights.[10]

2. Protein structure – *first orders*

As our article focuses on methods to assess protein structure or structure-dependent function, for *amuse bouche*, we present an overview on protein structure. Protein structure is a complex, multifactorial problem that cannot be described by a single parameter. This complexity inherent to protein structure can be reflected by breaking it down into four categories (Fig. 1a):

1) primary structure, the specific sequence of amino acids that make up the protein;
2) secondary structure, local structure including β-sheets and helices (the most common of which is α-, but others such as π[11] and 3_{10}[12] exist);
3) tertiary structure, the overall fold of a single polypeptide chain, incorporating all secondary structural elements; and
4) quaternary structure, the convergence of multiple tertiary structures (e.g., oligomerization) as well as cofactor incorporation.

Unsurprisingly, different biophysical methods are applicable to measure specific structural aspects; some go a stage further and focus on specific regions within a protein. We will discuss this aspect in each section.

Although this breakdown of structructural units, and by extension domain theory,[13] gives an idea that local interactions are increadibly important for protein structure and function, this is not entirely the case. Protein structure and function are increadibly complex, and indeed mutation of one region of a protein can have dramatic implications on remote regions,[14] through unexpected mechanisms.[15].

3. Mass spectrometry – *For starters, something exciting and impactful to raise your spirits?*

Mass spectrometry (MS) is a physical method used to measure mass of particles. It does so through volatilizing samples, then ionizing them, and subsequently detecting them. The actual output of MS is the mass divided by the charge (m/z), although this is typically converted to mass. Because the nuts and bolts of this process are particularly complex, depending on volatility, ionization properties and stability of the ions, all of which are highly sensitive to molecular structure and particular experimental conditions, MS is inherently unable to give absolute levels of specific components in mixtures, or even compare individual peaks between samples. Several methods have been used to circumvent these issues that we will briefly discuss below. Nevertheless, because it can accurately measure particle mass, MS is one of the most useful methods to identify unknown proteins involved in specific process or pathways one is interested in. As outlined below, MS is also applicable to other questions worthwhile posing early on in proceedings.

3.1. Digest mass spectrometry – Biting off what you can chew

Applicability:	**(semi)-purified protein, complex mixtures**
Main structural aspect investigated:	**primary structure**
Others general applications:	**protein modification/truncation (with caveats)**
Is it quantitative?	**relative amounts are readily quantifiable; absolute amounts not particularly reliable.**

3.1.1. Protein identification – Guess who?

Digest MS involves treating protein samples with a protease, separating the fragments using liquid chromatography, and then analyzing individual fragments by tandem MS. (Note: tandem MS is a variant of MS where each ion is analyzed for total mass, then fragmented into constituent components). The fragmentation patterns give precise sequence information that allows identification of each fragment within a specific protein; hence, this method is ideal for assigning primary sequence. For this reason, it is crucial that fragmentation patterns are contiguous, and that the digest fragments are consistent with the protease used for digestion (Fig. 1b). Fragmentation derived from standard collision induced dissociation (CID) occurs at the amide C—N bond, giving rise to y or b ions, where the charge resides on the primary ammonium, or an acylium ion, respectively (Fig. 1c: top panel). As MS depends on volatility and ionization, it should be remembered when planning a digestion experiment that short (<4 amino acid) and very long peptides (often, >~24 amino acids) are not readily applicable to this method. Bearing this in mind, proteases apposite for covering specific stretches of the primary sequence of one's target protein can be chosen.

For proteomics studies, it is often desired to compare enriched proteomes following different treatment conditions. Unfortunately, the issues of MS quantitation historically rendered such comparisons prone to error. Several methods have been developed to improve *relative* quantitation. These include the use of stable amino acid isotopomers, which allow quantitative comparison between samples because isotopomers have identical physical properties (volatility and ionization, for instance), but are significantly different in molecular mass to be resolved in MS. Post-lysis and post-digest isotope labeling, e.g., tandem-mass tagging (TMT), can also be used. The TMT approach has some limitations, principally associated with labeling occurring at peptide stage. However, TMT can nowadays be carried out in 11 or more multiplex, whereas, for instance, in-cell/in-vivo mass tagging through stable isotope labeling of amino acids in cell culture (SILAC), can only be used in triplicate. For readers interested to delve deeper into MS proteomics capabilities and for chemical biological or biochemical investigations, we refer them to our recent mini-review.[16] With the improvement of modern MS methods, it is now quite common to use label free quantification (LFQ) MS, which compares ion signatures per run in native digest samples. Although this has no internal control, with multiple replicates, and careful processing using either data-dependent or data-independent acquisition (DDA/DIA), and taking advantage of the fact that enriched samples often contain large amounts of noise, numerous researchers been able to identify specific associations and relevant functional targets using this method. For instance, we have deployed many of the above techniques in concert with digest proteomics to uncover specific electrophile-sensor proteins.

3.1.2. Protein modifications – change and chop

As posttranslational modifications (PTMs) change a protein's molecular weight, MS can be used to identify PTMs. This method usually requires a relatively large amount of sample protein, as the aim is to cover as much of a specific protein's sequence as possible (only a few, ideally unique, peptides are necessary to identify a protein). Using tandem MS, specific PTMs can be assigned to specific fragments, and then to specific residues. As above, to assign specific modified residues, a

contiguous fragmentation pattern that allows loss of the specific PTM at the specific residue should be observed (Fig. 1c: bottom panel). Alternatively, specific PTM-associated proteins/peptides can be enriched from samples/crude extracts. Enrichment can be achieved using affinity resins; for instance, anti-R-ε-GG antibodies for ubiquitin modifications. Subsequently the proteins or peptides harboring specific PTMs can be analyzed by MS. In our investigations into electrophile PTMs within specific proteins, we have frequently deployed MS to identify modified residues. This has been performed in *in vitro* purified samples, as well as proteins derived from cells in which the proteins are either overexpressed or expressed at endogenous levels,[17] often with an epitope tag. Just as all other MS methods, we point out that detection of modified peptides can only be as good as those peptides are volatile, ionizable, and stable. Moreover, many modifications involve stable bond formation, such as isopeptides, or monophosphates, and are relatively easy to detect by MS. However, many of the PTMs we and others work with are labile. These include cysteine modifications by Michael-acceptor-derived electrophiles and related electrophilic drugs. These often need specific mild volatilization and ionization conditions. Indeed, not observing an expected modification does not mean that no modification has occurred. Conversely, observing a modification gives little indication of stoichiometry and biological relevance. Thus, in terms of residue modification identification, the PTM-mapping MS data should always be independently investigated by orthogonal approaches including functional mutagenesis in relevant systems.

3.1.3. *Protein truncations – things are getting interesting, not worth cutting it short?*

MS can also be used to assign cleavage position of specific proteins. This often uses a technique called digest sequencing. As with PTM-mapping, this experiment uses purified protein, with the aim of achieving very high coverage, to allow assignment of the total sequence of the protein present. In this instance, having a control, full-length protein, as well as good coverage, are particularly important. Nonetheless, it can be difficult to be totally confident of cleavage site. Indeed, it is often worth attempting several proteases to try to ensure that correct terminal coverage has been achieved.

3.2. *Intact mass spectrometry – taking the whole weight*

Intact MS measures molecular weight of whole proteins, potentially with PTM or bound ligand. It can be performed on native, or denatured proteins. Fragmentation of ions by tandem MS is possible, allowing primary sequence to be identified. There are two principal methods of ionization currently used in intact MS: electrospray ionization (ESI); and matrix assisted laser desorption ionization (MALDI). ESI is compatible with HPLC, and so it is used almost exclusively in digest proteomics and other automated procedures. ESI is also believed to favor fragmentation relative to MALDI, aiding sequence assignment. MALDI requires the use of a matrix and hence is not so readily applicable to automated procedures. However, MALDI is often used to measure total protein mass, and validate *in vitro* protein synthesis/modification procedures as it is less prone to ion fragmentation.

Intact MS has some other interesting applications beyond simple mass measurements. MALDI has been adapted to MS imaging, where the intensities of ions for a specific analyte are measured across a suitably prepared biological sample. This method has relatively high resolution (~10 μM), which is on the order of the size of a nucleus.[18] It can measure both small molecules and proteins. Other interesting avenues are the measurement of conformational changes,[19] native complexes and dynamical associations. Intact MS is also commonly used in C-terminal MS sequencing, where exocarboxypeptidases are used to winnow the C-terminus one amino acid at a time to allow sequencing (Fig. 1d).

4. Other methods to determine primary structure – *an interesting side dish*

4.1. *Edman degradation – chemical N-terminal sequencing*

Applicability:	**purified protein**
Main structural aspect investigated:	***N*-terminus**
Others general applications:	**few**
Is it quantitative?	**no**

N-terminal processing of proteins is relatively common, due to *N*-terminal methionine cleavage, and signal peptide cleavage. Although *N*-termini can be sequenced by digest MS sequencing, there are always potential worries about how complete coverage is, potentially leading to aberrant assignment of termini. An alternative to this approach is Edman degradation, a chemical method to directly determine the *N*-terminus of proteins. Although it requires a relatively large amount of sample, it is particularly precise and can give reads of up to ca. 20 amino acids, which is typically of sufficient length to assign positions of truncation, for example. Edman degradation is certainly a good option if a large quantity of the purified protein is accessible, and a new *N*-terminus is expected within the cleavage fragment. It is worth noting however that Edman degradation is blocked by *N*-terminal modification, such as formylation/acetylation, although this is not common for cleaved peptides.

5. Methods to determine three-dimensional structure – *returning to the fold for the main course*

Having negotiated "the first course", and identified a protein, its modifications, and truncations, it is now important to start asking some more specific questions. Indeed, although primary structure information is hugely important, in the end, protein folding, i.e., secondary structure and beyond, are key to open a window into the soul of one's favorite protein. There are several methods applicable to answer these questions.

5.1. *Crystallography – peering into infinity*

Applicability:	**purified protein**
Main structural aspect investigated:	**all structural elements and ligand associations**
Other notes:	**little limit in size or other parameters**
Is it quantitative?	**N/A**

Crystallography is a venerable structural method that has its roots in the genesis of structural biology and structure-guided enzymology. Indeed, since the structure of lysozyme was solved by David Chilton Phillips in 1965, revealing several important aspects of how enzymes function, protein structure has been considered to give crucial insights into the inner workings of proteins. This has extended to ligand interactions, and inhibitor mode of action. With almost 60 years of work behind it, there is now a huge inventory of structures available in the Protein Data Bank (PDB). This is a very useful resource for rationalizing interactions, and predicting mutants. For instance, when we identify an electrophile-sensitive cysteine, we typically investigate its surface availability, and surrounding residues by perusing the crystal structure. Of course, not all proteins have been crystallized, and even less so have been crystallized with specific ligands bound. New additions to the structural armory, particularly α-fold (2) that can predict protein structure from primary sequence have helped broaden the remit of our structural understanding.[20] This trend has extended to drug design and discovery,[21] although this is not the focus of our perspective. Of course, care should be exercised when interpreting computed structures and common pitfalls underpinning modern structural prediction tools have been nicely reviewed elsewhere.[22] In our case, we often use Swissmodel to perform homology modeling,[23] for instance of zebrafish proteins, for which there are rarely structures available.[24]

In the context of empirical crystallographic data, conditions under

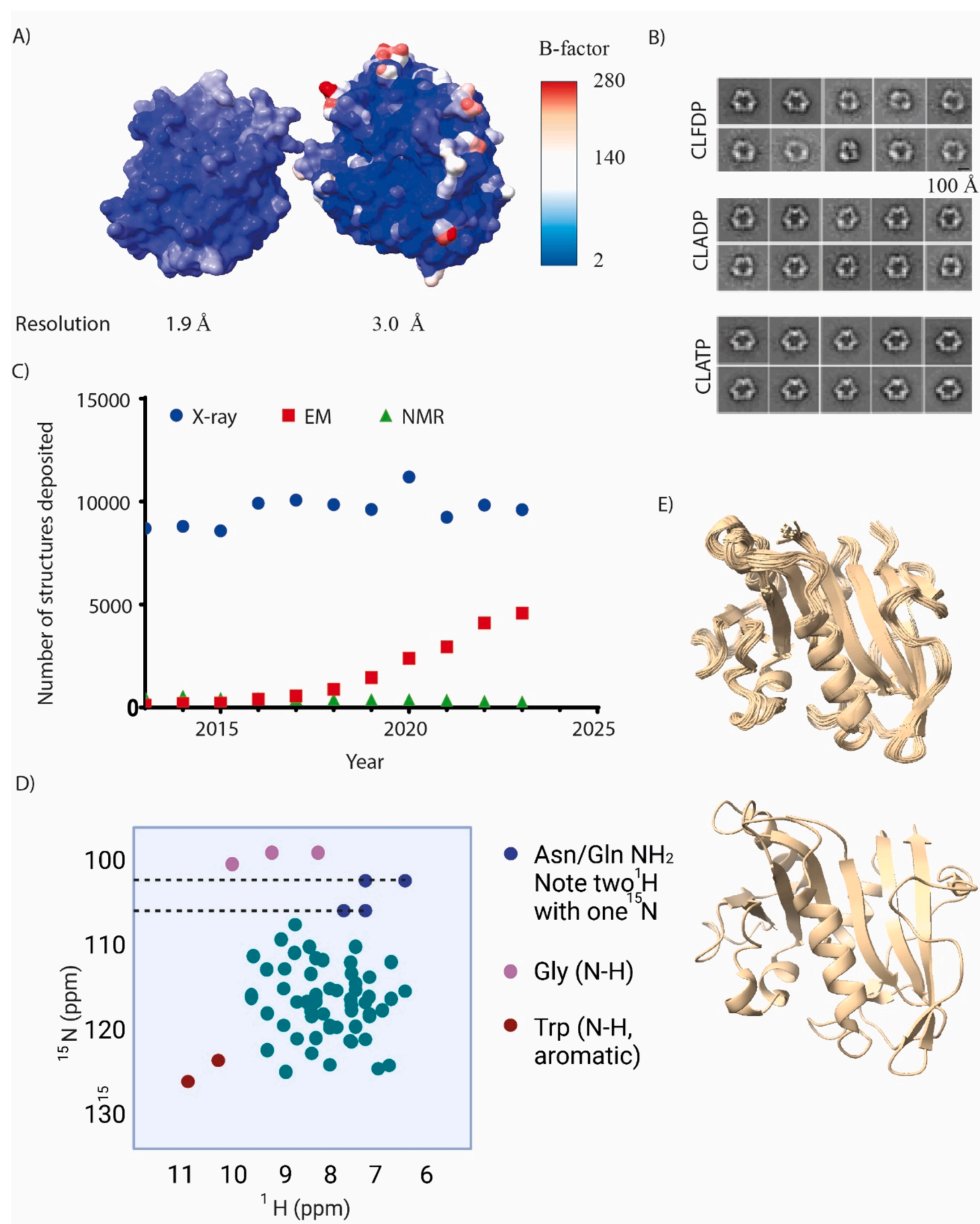

Figure 2. Different structural techniques discussed in the main text. (a) Two different structures of human GST-α4.4 dimers at high (PDB: 3IK7) and low (PDB: 1GUM) resolution. For each structure, a monomer is colored by B-factor, blue lowest to red highest, as per the scale on the right. (b) Representative pictures of different conformations of RNR-α hexamers detected by treatment with different drugs. Figures reproduced from Fu et al. Nat Chem Biol 2018 14 943. Image copyright belongs to the authors of this perspective. (c) Number of structures deposited in the PDB per year based on method. (d) HSQC NMR spectrum, with some rough guide-lines about characteristic peaks. See manuscript text for details. (e) Human dihydrofolate reductase shown as both an NMR (PDB: IYHO) (top panel) and a crystal structure (PDB: 3NZD) (bottom panel) in ribbon form.
Source: PDB Statistics (rcsb.org)).

which specific crystal structures were obtained (pH, ionic strength, reducing agents, etc.), the presence or absence of ligands, and also the overall resolution, as well as the B-factors across the crystal structure, are worth closer attention. The B-factor—often referred to as the temperature factor—reflects precision within the structure. This should be relatively low for well-defined atoms, but can become large for atoms whose precise structure is not defined. A similar logic applies to overall resolution: as resolution increases, the B-factor should diminish.[25] We illustrate this point by putting side by side crystal structures of glutathione-S-transferase with 3.00 Å and 1.97 Å resolution, respectively (Fig. 2a). It is clear that the B-factors for some atoms are much larger in the former. Emerging frontiers of this method include observing chemical/enzymatic processes within crystal structures,[26] although the rigid packing in the crystals may limit the scope of such efforts.[27]

Of course, crystallography is perhaps the most sensitive method to

protein purity and compatibility with the method itself. Indeed, for most crystallographic efforts, large amounts of pure protein are required. Transmembrane proteins can be particularly difficult to analyze by these methods, and a host of methods aimed precisely at crystallizing membrane proteins[28] have been reported.[29]

5.2. Electron microscopy (EM) – is the atmosphere getting electric?

Applicability:	**purified protein, or potentially mixtures**
Main structural aspect investigated:	**all structural elements and ligand associations**
Other notes:	**little limit in size or other parameters**
Is it quantitative?	**specific particles in a sample can be classified and quantified**

EM is becoming one of the most powerful structural methods. This method is applicable to answer questions pertaining to gross structural arrangements, principally by negative stain EM. But it is also able to give high resolution information (up to ∼1.5 Å) in the form of cryo-electron microscopy.

5.2.1. Negative stain EM – it's OK to focus on the negative

Negative stain EM is a technique that is robust and can give gross structural information, particularly of large proteins, complexes and aggregates. It has relatively low resolution, ∼10–20 Å, and hence is not particularly useful for small proteins <∼80 kDa, although some reports claim that smaller proteins than this can be visualized.[30] Negative stain EM is applicable to homogeneous, as well as relatively non-homogeneous samples, and indeed, for the latter case, negative stain EM is preferred to cryo-EM. In this way negative stain EM can give information on conformational heterogeneity, for instance, during enzymatic activation and associated changes in conformational dynamics. We have used this method to classify specific conformationally-distinct hexamers of the enzyme ribonucleotide reductase subunit-α (RNR-α, monomer weight, ∼90 kDa) induced by different approved nucleotide therapeutics that target RNR-α (Fig. 2b). We further backed up the negative-stain-EM-derived conformationally distinct states of the resulting hexamers, by demonstrating that different hexamers shaped by different drugs were differentially susceptible to protease digestion.

5.2.2. Cryo-EM – putting things on ice to get a better perspective

Cryo-EM is a high-resolution method that can give the same level of structural detail and insight as X-ray crystallography. Such outputs are possible because at very low temperatures damage caused by electron beams (that is high at room temperature) is limited. Cryo-EM typically requires relatively high structural homogeneity, although strategies to deal with heterogeneity are available.[31] This method is applicable to both soluble and membrane proteins.[32] Indeed, in the last 10 years, cryo-EM has been the largest source of membrane protein structures.[33] The number of total EM-structures deposited to the PDB has increased year on year in this time, although X-ray structures still constitute the largest single soure of structures in the PDB[34] (Fig. 2**c**).

5.3. NMR – take advantage of the rough and tumble

Applicability:	**purified proteins & either small-molecule or biomolecule-based ligands**
Main structural aspect investigated:	**all structural elements and ligand associations**
Other notes:	**large proteins are not amenable to this method; high solubility can be needed**
Is it quantitative?	**capable of providing a large amount of quantitative information**

Solution NMR is a structural method that investigates specific spin-active nuclei, for instance, hydrogen or nitrogen, in a sample. As the method is particularly sensitive to nuclear environment, in an NMR spectrum, non-equivalent nuclei within a sample show up as specific peaks, each of which have a defined value, referred to as a chemical shift. In standard protein NMR measurements, ^{1}H and ^{15}N are used in a two-dimensional experiment, to separate out peaks that would otherwise be occluded in one-dimensional experiment. In this experiment broadly known as heteronuclear single quantum coherence (HSQC), N—H bonds show up as specific peaks in a two-dimensional grid. Thus, the technique focuses specifically on peptide N—H bonds, as well as sidechain N—H's, such as tryptophan aromatic N—H and asparagine/glutamine NH_2 (Fig. 2d); typical spectra run from 6-11 ppm for ^{1}H, and 105–140 ppm for ^{15}N. Several of these residues, as well as specific peptide N—H bonds have relatively characteristic chemical shifts that can aid rapid analysis of spectral quality (e.g. glycine typically has relatively low ^{15}N chemical shifts (∼110 ppm); glutamine/asparginine NH_2 usually have low ^{1}H (∼7.5 ppm) and ^{15}N (∼110 ppm) chemical shifts and have two protons on same nitrogen; tryptophan aromatic NH, usually has high ^{1}H (10.5 ppm) and ^{15}N (130 ppm) chemical shifts). It should be noted that the natural isotope of nitrogen contains an even number of protons and neutrons within its nucleus, and is silent in NMR. It is thus necessary to feed bacteria for protein preparation with heavy nitrogen (^{15}N, typically in the form of heavy ammonia), in minimal media in order to prepare proteins for NMR.

One of the standard limitations of NMR was classically the size of molecule that can be measured. This limitation traces its roots back to the fact that NMR relies on molecular tumbling, that leads to the line breadth of samples increasing as the size of the molecule increases. This effect is prohibitive for molecules of several kDa's in size. With the advent of highly powerful spectrometers and transverse relaxation-optimized spectroscopy (TROSY) pulse sequences, that seek to limit the effect of line broadening due to size, it is now quite feasible to measure NMR of proteins in excess of 50 kDa, although, very large proteins are still problematic.

NMR can give a large amount of both qualitative and quantitative structural information. In general, NMR spectra can inform on protein folding. Spectra manifesting well-dispersed peaks are often indicative of a complex, heterogeneous environment provided by a protein structure. Folding states of different mutants can be compared by investigating changes in peaks upon mutation – in general, residues spatially close to the mutated site will change, as NMR is particularly sensitive to chemical environment. However, residues distal from the mutated site should be unchanged. This can be defined by parameters such as minimal chemical shift perturbations, i.e., the difference in chemical shift between each residue. A similar argument applies to ligand binding, where chemical shift perturbations will occur only when an interaction occurs. There are indeed a huge range of programs designed specifically for NMR analysis, peak assignment, and the like.[35]

Determination of whether a protein is folded or not by NMR is more or less independent of the assignment of peaks within the NMR spectrum of the protein.[36] However, to assign protein structure by NMR requires more detailed information both in terms of peak assignment and spatial distribution of the peaks. Oftentimes, further dimensions are needed to separate out peaks better (usually ^{13}C is used). Although NMR usually provides information on through-bond interactions (i.e., unaffected by spatial distribution), it can also provide information on proximity by measuring through-space effects, using the nuclear Overhauser effect (NOE).[37] Thus, NMR has all the necessary properties to solve protein structures. Aside from requiring a large amount of computational analysis, such pursuits often require complex protein preparation procedures, which can be technically difficult. Of course, numerous proteins have already had their NMR structures solved, including peak assignments, which simplifies matters considerably. One aspect that is particularly striking when viewing NMR structures of proteins versus X-ray crystallographic data, especially, is that the former show many conformations of the protein (Fig. 2**e**). This is because in solution the protein structure can "breathe", whereas in the restricted environment of a crystal mostly the protein is fixed in (a small number of) conformations.

NMR indeed is a uniquely versatile method that lends itself to numerous specialized procedures that we cannot completely cover here. Aside from using heavy isotopes to allow NMR visualization, the incorporation of atoms not present endogenously, particularly fluorine (^{19}F) open means to simplify spectra, and potentially improve responsivity. There are also numerous NMR experimental approaches that are ideally suited to answer specific biological questions. Protein ligand interactions and screening can be studied by several techniques, including saturation-transfer difference (STD) NMR,[38] a technique based on NOE. Moreover, proteins can show a gamut of different motions, associated with simple bond rotations, and larger conformational changes. NMR is equipped to study many, if not most of these dynamical changes.[39]

5.4. *Circular dichroism – the right time to split?*

Applicability:	**purified protein**
Main structural aspect investigated:	**secondary structure, stability, and ligand association**
Other notes:	**no real limit in size, but high analyte concentration can affect low wavelength absorbance, limiting data acquisition**
Is it quantitative?	**can give percentage of secondary structure, but this is a rough guide**

Residue-specific information is typically critical for understanding, for instance, how mutations affect protein structure, or potentially how ligand binding occurs. However, in many instances, a more global view of protein structure is sufficient, or perhaps even preferred. In this case, CD can be particularly useful. This technique informs on protein secondary structure. Characteristic CD spectra for α-helices and β-sheets are established, and are significantly different from each other and unfolded polypeptide chains. Based on these behaviors, there are several programs that can assign structural composition based on CD spectra, although these likely should be interpreted carefully. One simple test is to use analogy to published crystal (or more ideally NMR) structures to see if the secondary structural characteristics are sensible.

We have regularly used CD spectra to compare gross structural similarities across recombinantly expressed mutants. We have also used CD to show that there are gross structural changes in specific proteins when they are treated with reactive electrophilic ligands. Other uses of CD include measuring thermal stability of proteins. This technique can also be applied to assessing mutant proteins, which ideally should be similarly stable to the wt-protein. Ligand binding can also be measured. In general, ligand binding should stabilize bound proteins, typically with a change in temperature related to the Gibbs free energy of binding-the higher the affinity, the larger the thermal stabilization.[36] Assays that use fluorescent-dye binding to assess protein stability can give similar information, and are amenable to high-throughput experimentation.[40] However, CD remains a preferred method if relatively few mutants or conditions are to be investigated.

5.5. *Small angle X-ray scattering (SAXS)*

Applicability:	**purified protein (ideally free of aggregates, and contaminants)**
Main structural aspect investigated:	**protein aggregation, size, shape, and ligand binding (relatively low resolution)**
Other notes:	**no real limit in size**
Is it quantitative?	**can provide kinetics or information on structural changes**

SAXS is a solution method that can inform on relatively large structural perturbations/transitions/polydispersity in macromolecules. In many ways it is thus complementary to crystallography and other similar methods that offer more profound structural information. In several instances it has been used in conjunction with those methods. SAXS can provide several parameters that correlate with important physical parameters in solution. These include the radius of gyration, *R*g, which relates to the overall size of the molecule in solution, molecular weight and maximum dimension (*D*max). These parameters can be compared to those calculated for specific proteins, to understand specific aspects of a protein such as folding and flexibility. Improvements in data collection, and analysis have allowed for increases in rapidity of SAXS experiments, and uses available.[41,42] Nonetheless, SAXS requires a strong X-ray beam source and is usually performed at a specialist facility using synchrotron sources.

6. Methods investigating function – *just a taster for dessert, but who knows where it go?*

In the end, protein structure, modification, and the like are mainly methods to gain deeper comprehension of protein function. We thus end our tasting menu with a dessert trolley presenting some methods that we commonly deploy to detect enzyme activity and protein functional associations. This section, like the above, is by no means exhaustive, but it should inform on choice of assay.

6.1. *Enzyme kinetics – hooking up again, or making a quick exit?*

Applicability:	**purified protein, lysates, or potentially cells**
Main structural aspect investigated:	**all aspects of protein structure as these are all linked to activity**
Other notes:	**no real limit in size**
Is it quantitative?	**highly quantitative and sensitive; can be used to quantify absolute amounts of protein**

For purified enzymes, or enzymes in complex mixtures, particularly those overexpressed, enzyme kinetics constitutes a diagnostic, quantitative, and accurate means to assess protein expression and folding. These assays can be readily extended to inhibition, and activation. Indeed, to assess concentration of enzymes, active-site titrations, are considered the most accurate. Numerous methods exist for measuring enzymatic activity. In general, a continuous assay that can be measured by absorbance, fluorescence, etc., in real time gives the simplest protocol. Several fluorescent assays can now be deployed in live cells, or even live model organisms. This offers good flexibility and offers new possibilities after initial discovery phases. However, for proteins that are not particularly active, or whose chemistry is not amenable to such assays, numerous radioactive, western blot, and bead-based assays exist. In the case of RNR-α despite there being several reported activity assays for this enzyme, we have typically opted to measure conversion of radiolabelled [5-^{3}H]-CDP as it is highly sensitive, and with sufficient practice, progress curves and time-dependent inhibition can be efficiently measured.

6.2. *Analytical chromatography – (hydrodynamic) size can matter*

Applicability:	**purified protein, specific mixtures, or lysates**
Main structural aspect investigated:	**protein quaternary associations**
Other notes:	**no real limit in size**
Is it quantitative?	**highly quantitative and sensitive, size resolution depends on column properties**

This is a method to separate proteins by size, with the largest proteins eluting first. Several different columns with different optimal size resolution parameters are available for assay optimization. This method is often deployed to separate different oligomeric states of proteins (for instance for X-ray crystallography, or when a specific oligomer of a protein is active[43]), to assess association, or to fractionate protein mixtures. Of course, in order to be suitable for analysis by analytical chromatography complexes must be relatively stable (certainly on the order of ~1 h, the time needed for a typical run). We have deployed this method to study the effect of RNR-α binding drugs on protein oligomeric

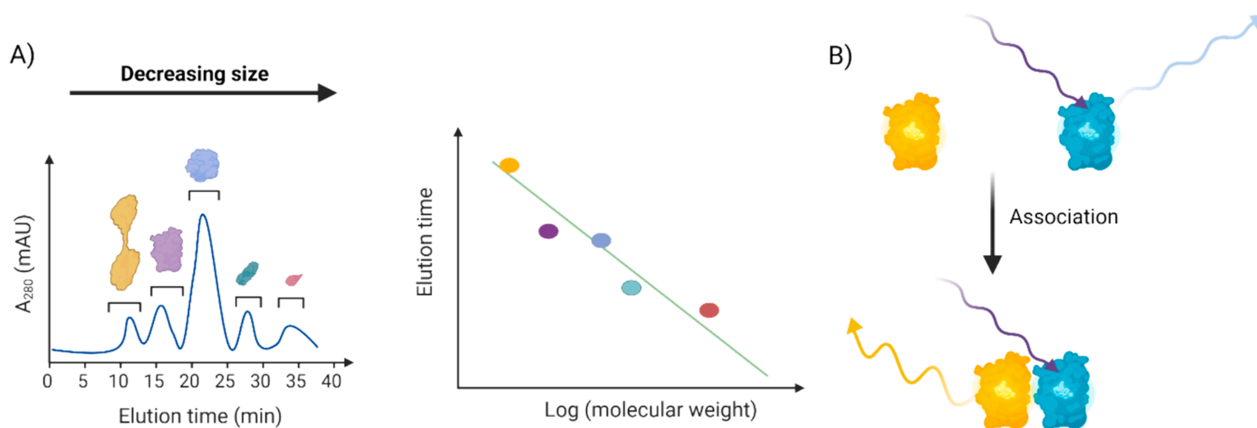

Figure 3. Schematic illustrations for some of the general biophysical techniques measuring proteins' functional properties. (a) Size exclusion chromatography separates proteins by size from largest to smallest (left panel: elution spectrum). Right panel: the logarithm of the molecular weight should be negatively correlated with elution time. (b) FRET between a CFP (blue) and YFP (yellow) reporting protein protein association of two designated proteins of interest (omitted in figure for clarity) genetically fused to CFP and YFP, providing also the resulting fusion constructs do not alter native (untagged) proteins' function/activity.

state both with purified proteins and proteins isolated from cells (Fig. 3a). In both instances, we showed that these inhibitors—either by direct treatment of purified protein or following drug treatment of cells—all caused RNR-α to shift to a higher molecular weight, hexameric form.

6.3. *Fret-based association assays – changing colors*

Applicability:	**purified protein, specific mixtures, lysates or cells**
Main structural aspect investigated:	**protein quaternary associations**
Other notes:	**no real limit in size**
Is it quantitative?	**highly quantitative and distance sensitive**

Fluorescence resonance energy transfer, FRET (and the closely related bioluminescence resonance energy transfer, BRET) occurs when two fluorophores with matching wavelengths (one a donor, with higher wavelength of excitation, the other an accepter with lower wavelength of excitation) are brought into proximity, for instance by an association[44] or a large change in conformation.[45] When the two fluorophores are in proximity, excitation of the acceptor fluorophore leads to excitation of the donor fluorophore through energy transfer. Thus, the emission of the acceptor decreases, while the emission of the acceptor increases. Because FRET/BRET are very sensitive to distance between the fluorophores, these methods can give accurate indication of distances between differently fluorescently-labeled proteins. We used this method to create a FRET reporter assay for RNR-α hexamerization (Fig. 3b) that does not rely on the relatively arduous radioactive activity/inhibition assay. In this case, we used non-site-specifically-fluorophore labeled recombinant RNR-α monomers that were mixed together. FRET is possible with fluorescent proteins, typically CFP and YFP are used. We have used several FRET-based reporters, developed by the Zhang laboratory, to report on activity of kinase AKT in live zebrafish larvae beyond cultured cells, for instance. Co-expression of differentially fluorescently labeled receptors that undergo clustering during signaling has also been used.[46]

7. Conclusion – *got a taste for biophysics?*

Biophysical methods present a varied and rich fleet of methods that are complementary to other biological and chemical-biological investigations, and can provide unique insights in of themselves. Their deployment often requires specific experimental equipment or techniques, although these are often available at core facilities or through collaboration. We hope that this piece has whetted your appetites for biophysical experiments.

CRediT authorship contribution statement

Marcus J.C. Long: Writing – review & editing, Writing – original draft. **Yimon Aye:** Writing – review & editing, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

Postdoctoral Fellowship from the Novartis Foundation for Medical-Biomedical Research (M.J.C.L.). EPFL (Y.A.).

References

1. Fu Y, et al. Nuclear RNR-α antagonizes cell proliferation by directly inhibiting ZRANB3. *Nat Chem Biol.* 2018;14:943–954. https://doi.org/10.1038/s41589-018-0113-5.
2. Wisitpitthaya S, et al. Cladribine and fludarabine nucleotides induce distinct hexamers defining a common mode of reversible RNR inhibition. *ACS Chem Biol.* 2016;11:2021–2032. https://doi.org/10.1021/acschembio.6b00303.
3. Aye Y, et al. Clofarabine targets the large subunit (α) of human ribonucleotide reductase in live cells by assembly into persistent hexamers. *Chem Biol.* 2012;19: 799–805. https://doi.org/10.1016/j.chembiol.2012.05.015.
4. Pan Y, et al. Glutathione (GSH)-decorated magnetic nanoparticles for binding glutathione-S-transferase (GST) fusion protein and manipulating live cells. *Chem Sci.* 2011;2:945–948. https://doi.org/10.1039/c1sc00030f.
5. Long MJC, Pan Y, Lin H-C, Hedstrom L, Xu B. Cell compatible trimethoprim-decorated iron oxide nanoparticles bind dihydrofolate reductase for magnetically modulating focal adhesion of mammalian cells. *J Am Chem Soc.* 2011;133: 10006–10009. https://doi.org/10.1021/ja202767g.
6. Morão LG, Manzine LR, Clementino LOD, Wrenger C, Nascimento AS. A scalable screening of E. coli strains for recombinant protein expression. *PLoS One* 2022;**17**: e0271403. doi:10.1371/journal.pone.0271403.
7. Aye Y, Stubbe J. Clofarabine 5′-di and -triphosphates inhibit human ribonucleotide reductase by altering the quaternary structure of its large subunit. *Proc Natl Acad Sci.* 2011;108:9815–9820. https://doi.org/10.1073/pnas.1013274108.
8. Liu J, et al. Organ-specific electrophile responsivity mapping in live C. elegans. *Cell.* 2024 (in press).
9. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol.* 2007;8:319–330. https://doi.org/10.1038/nrm2144.

10. Fu Y, et al. Uncoupling of allosteric and oligomeric regulation in a functional hybrid enzyme constructed from Escherichia coli and human ribonucleotide reductase. *Biochemistry*. 2013;52:7050–7059. https://doi.org/10.1021/bi400781z.
11. Cooley RB, Arp DJ, Karplus PA. Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. *J Mol Biol*. 2010;404:232–246. https://doi.org/10.1016/j.jmb.2010.09.034.
12. Li SC, Deber CM. A measure of helical propensity for amino acids in membrane environments. *Nat Struct Biol*. 1994;1:368–373. https://doi.org/10.1038/nsb0694-368.
13. Wang Y, Zhang H, Zhong H, Xue Z. Protein domain identification methods and online resources. *Comput Struct Biotechnol J*. 2021;19:1145–1153. https://doi.org/10.1016/j.csbj.2021.01.041.
14. Hedstrom L, Szilagyi L, Rutter WJ. Converting trypsin to chymotrypsin: the role of surface loops. *Science*. 1992;255:1249–1253. https://doi.org/10.1126/science.1546324.
15. Kozome D, Sljoka A, Laurino P. Remote loop evolution reveals a complex biological function for chitinase enzymes beyond the active site. *Nat Commun*. 2024;15:3227. https://doi.org/10.1038/s41467-024-47588-8.
16. Long MJC, Liu J, Aye Y. Finding a vocation for validation: taking proteomics beyond association and location. *RSC Chem Biol*. 2023;4:110–120. https://doi.org/10.1039/D2CB00214K.
17. Zhao Y, et al. Function-guided proximity mapping unveils electrophilic-metabolite sensing by proteins not present in their canonical locales. *Proc Natl Acad Sci*. 2022;119, e2120687119. https://doi.org/10.1073/pnas.2120687119.
18. Good CJ, et al. High spatial resolution MALDI imaging mass spectrometry of fresh-frozen bone. *Anal Chem*. 2022;94:3165–3172. https://doi.org/10.1021/acs.analchem.1c04604.
19. Dixit SM, Polasky DA, Ruotolo BT. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. *Curr Opin Chem Biol*. 2018;42:93–100. https://doi.org/10.1016/j.cbpa.2017.11.010.
20. Bertoline LMF, Lima AN, Krieger JE, Teixeira SK. Before and after AlphaFold2: an overview of protein structure prediction. *Front Bioinform*. 2023;3, 1120370. https://doi.org/10.3389/fbinf.2023.1120370.
21. Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature*. 2023;616:673–685. https://doi.org/10.1038/s41586-023-05905-z.
22. Wuyun Q, et al. Recent progress of protein tertiary structure prediction. *Molecules*. 2024;29:832.
23. Binbay FA, Rathod DC, George AAP, Imhof D. Quality assessment of selected protein structures derived from homology modeling and AlphaFold. *Pharmaceuticals (Basel)*. 2023;16. https://doi.org/10.3390/ph16121662.
24. Surya SL, et al. Cardiovascular small heat shock protein HSPB7 is a kinetically privileged reactive electrophilic species (RES) sensor. *ACS Chem Biol*. 2018;13:1824–1831. https://doi.org/10.1021/acschembio.7b00925.
25. Carugo O. How large B-factors can be in protein crystal structures. *BMC Bioinf*. 2018;19:61. https://doi.org/10.1186/s12859-018-2083-8.
26. Schotte F, Soman J, Olson JS, Wulff M, Anfinrud PA. Picosecond time-resolved X-ray crystallography: probing protein function in real time. *J Struct Biol*. 2004;147:235–246. https://doi.org/10.1016/j.jsb.2004.06.009.
27. Butryn A, et al. An on-demand, drop-on-drop method for studying enzyme catalysis by serial crystallography. *Nat Commun*. 2021;12:4461. https://doi.org/10.1038/s41467-021-24757-7.
28. MacKinnon R. Potassium channels and the atomic basis of selective ion conduction. *Nobel Lecture*. 2003:214–235.
29. Tao X, Zhao C, MacKinnon R. Membrane protein isolation and structure determination in cell-derived membrane vesicles. *Proc Natl Acad Sci*. 2023;120, e2302325120. https://doi.org/10.1073/pnas.2302325120.
30. Ohi M, Li Y, Cheng Y, Walz T. Negative staining and image classification—powerful tools in modern electron microscopy. *Biol Procedures Online*. 2004;6:23–34. https://doi.org/10.1251/bpo70.
31. Milne JLS, et al. Cryo-electron microscopy – a primer for the non-microscopist. *FEBS J*. 2013;280:28–45. https://doi.org/10.1111/febs.12078.
32. Kühlbrandt W. Forty years in cryoEM of membrane proteins. *Microscopy*. 2022;71:i30–i50. https://doi.org/10.1093/jmicro/dfab041.
33. Harrison PJ, Vecerkova T, Clare DK, Quigley A. A review of the approaches used to solve sub-100 kDa membrane proteins by cryo-electron microscopy. *J Struct Biol*. 2023;215, 107959. https://doi.org/10.1016/j.jsb.2023.107959.
34. *PDB Statistics_Growth of Structures from X-ray Crystallography Experiments Released per Year (accessed: 28/7/2024)*, <https://www.rcsb.org/stats/growth/growth-xray>.
35. Lee W, Rahimi M, Lee Y, Chiu A. POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. *Bioinformatics*. 2021;37:3041–3042. https://doi.org/10.1093/bioinformatics/btab180.
36. Shi Y, et al. Boc(3)Arg-linked ligands induce degradation by localizing target proteins to the 20S proteasome. *ACS Chem Biol*. 2016;11:3328–3337. https://doi.org/10.1021/acschembio.6b00656.
37. (1999).
38. Aretz J, Rademacher C. Ranking hits from saturation transfer difference nuclear magnetic resonance-based fragment screening. *Front Chem*. 2019;7. https://doi.org/10.3389/fchem.2019.00215.
39. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. *Nature*. 2007;450:964–972. https://doi.org/10.1038/nature06522.
40. Elgert C, Rühle A, Sandner P, Behrends S. Thermal shift assay: Strengths and weaknesses of the method to investigate the ligand-induced thermostabilization of soluble guanylyl cyclase. *J Pharm Biomed Anal*. 2020;181, 113065. https://doi.org/10.1016/j.jpba.2019.113065.
41. Josts I, et al. Structural kinetics of MsbA investigated by stopped-flow time-resolved small-angle X-ray scattering. *Structure*. 2020;28. https://doi.org/10.1016/j.str.2019.12.001, 348–354.e343.
42. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett*. 2015;589:2570–2577. https://doi.org/10.1016/j.febslet.2015.08.027.
43. Cho JK, et al. Geranylated flavonoids displaying SARS-CoV papain-like protease inhibition from the fruits of Paulownia tomentosa. *Bioorg Med Chem*. 2013;21:3051–3057. https://doi.org/10.1016/j.bmc.2013.03.027.
44. Martín-Sánchez F, et al. Measuring NLR oligomerization III: detection of NLRP3 and NLRC4 complex by bioluminescence resonance energy transfer. *Methods Mol Biol*. 2023;2696:93–103. https://doi.org/10.1007/978-1-0716-3350-2_6.
45. Kowalski-Jahn M, Schihada H, Schulte G. Conformational GPCR BRET sensors based on bioorthogonal labeling of noncanonical amino acids. *Methods Mol Biol*. 2023;2676:201–213. https://doi.org/10.1007/978-1-0716-3251-2_15.
46. Block V, et al. CVID-associated B cell activating factor receptor variants change receptor oligomerization, ligand binding, and signaling responses. *J Clin Immunol*. 2023;43:391–405. https://doi.org/10.1007/s10875-022-01378-3.