4-17-2020

Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution

Mark R. Chance
Case Western Reserve University, mark.chance@case.edu

Erik R. Farquhar
Case Western Reserve University, erik.farquhar@case.edu

Sichun Yang
Case Western Reserve University, sichun.yang@case.edu

David T. Lodowski
Case Western Reserve University, david.lodowski@case.edu

Janna G. Kiselar
Case Western Reserve University, janna.kiselar@case.edu

Author(s) ORCID Identifier:

Mark R. Chance

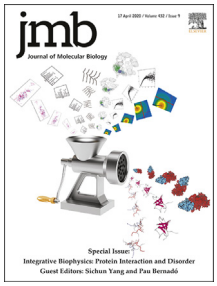
Erik R. Farquhar

Sichun Yang

David T. Lodowski

Recommended Citation

Mark R. Chance, Erik R. Farquhar, Sichun Yang, David T. Lodowski, Janna Kiselar. Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution. Journal of Molecular Biology, Volume 432, Issue 9, 2020, Pages 2973-2984, https://doi.org/10.1016/j.jmb.2020.02.011.



Protein Footprinting: Auxiliary Engine to Power the Structural Biology Revolution

Mark R. Chance[1,2,3], Erik R. Farquhar[2], Sichun Yang[1,3], David T. Lodowski[1,3] and Janna Kiselar[1,3]

*1 - **Case Center for Proteomics and Bioinformatics,** USA*
*2 - **Case Center for Synchrotron Biosciences,** USA*
*3 - **Department of Nutrition,** Case Western Reserve University, School of Medicine, 10900 Euclid Ave., Cleveland, OH, 44106, USA*

***Correspondence to Mark R. Chance:** Case Western Reserve University, 10900 Euclid Ave., BRB930, Cleveland, OH, 44106, USA, Fax: 1.216.368.6846. mark.chance@case.edu.*
https://doi.org/10.1016/j.jmb.2020.02.011
Edited by Pau Bernado

Abstract

Structural biology is entering an exciting time where many new high-resolution structures of large complexes and membrane proteins are determined regularly. These advances have been driven by over fifteen years of technology advancements, first in macromolecular crystallography, and recently in Cryo-electron microscopy. These structures are allowing detailed questions about functional mechanisms of the structures, and the biology enabled by these structures, to be addressed for the first time. At the same time, mass spectrometry technologies for protein structure analysis, "footprinting" studies, have improved their sensitivity and resolution dramatically and can provide detailed sub-peptide and residue level information for validating structures and interactions or understanding the dynamics of structures in the context of ligand binding or assembly. In this perspective, we review the use of protein footprinting to extend our understanding of macromolecular systems, particularly for systems challenging for analysis by other techniques, such as intrinsically disordered proteins, amyloidogenic proteins, and other proteins/complexes so far recalcitrant to existing methods. We also illustrate how the availability of high-resolution structural information can be a foundation for a suite of hybrid approaches to divine structure-function relationships beyond what individual techniques can deliver.

Advances in structural biology and biophysics applications of macromolecular crystallography, cryo-EM, and NMR have revolutionized our access to protein and nucleic acid structural information. These structural advances, coupled to advances in genomic sequencing and molecular biology, have provided rapid identification of drug targets and have revolutionized rational drug development for both small molecules and biologics [1]. A major theme emerging in structural biology is that of combining different methods through appropriate data integration. In Fig. 1 we illustrate this cycle of analysis, which starts with high-resolution structural information (or models) from macromolecular crystallography, Cryo-EM, or other methods and through the application of a range of orthogonal approaches asks questions like, What are the structures of multicomponent macromolecular complexes in varying molecular states? How do they structurally interconvert between these functional states? What structural features drive the kinetics and thermodynamics of assembly?

The structural genomics revolution provided template structures for most soluble domains from crystallographic and NMR data [2,3], but these "Lego blocks" representing discrete tertiary structural elements need additional data to "assemble" them into their physiological context as components of complex macromolecular machines. Techniques like small-angle x-ray scattering (SAXS) and electron microscopy (EM) are powerful approaches that provide global shape envelopes (Fig. 1); these can

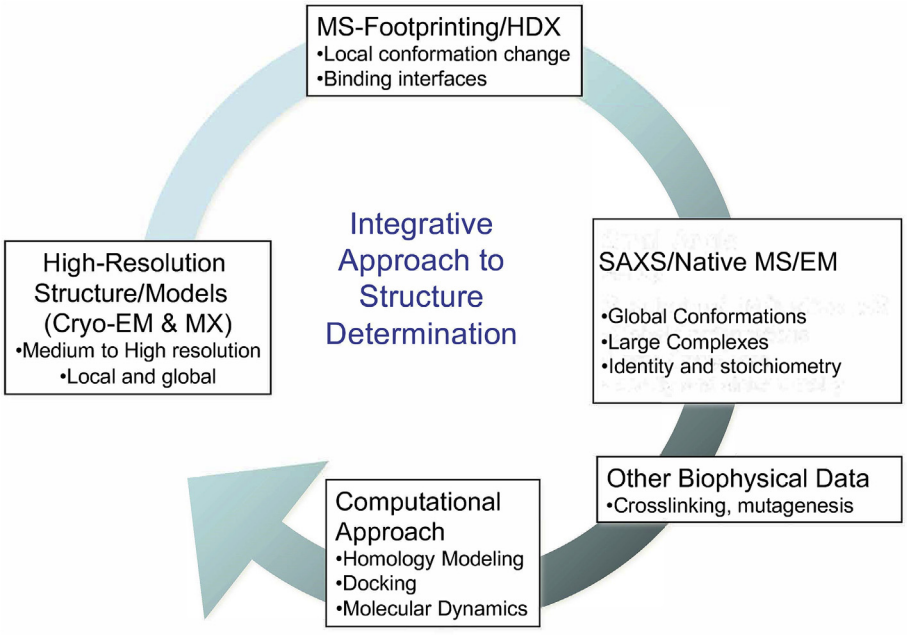


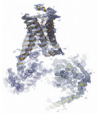
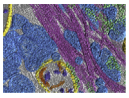
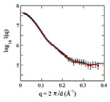
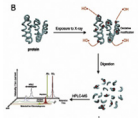
Fig. 1. Integrated workflow for structure assessment. Reprinted with permission [1].

help to understand the different ways of assembling these "Lego blocks" [4,5]. Along with these shape measures, native mass spectrometry (MS) [6,7] has been quite valuable in elucidating the composition of these large assemblies. In this MS approach (also called "top-down" because an intact protein species has been introduced to the instrument), macromolecular complexes are ionized and introduced to the gas phase, then subjected to cycles of analysis and fragmentation to understand subunit stoichiometries, topology, and binding to ligands.

This top-down approach is complementary to classical "bottom-up" mass spectrometry, where proteins are digested by proteases into a set of constituent peptides (ranging from 6 to 15 residues each), and the peptides are separated by chromatography and individually analyzed. In bottom-up footprinting experiments, the protein complexes are chemically labeled, then digested to the peptide level and introduced to the instrument while preserving as much of the labeling as possible, and the labeling patterns of the peptides are analyzed to infer important structural and dynamics information. The most popular bottom-up approaches are crosslinking, hydrogen-deuterium exchange (HDX), and irreversible covalent labeling. Crosslinking typically uses bi-functional chemical reagents with linkers of defined length (from so-called "zero length" to longer) that react with side chain groups (e.g., Lys) linking side chains that can achieve close approach. This data can be deconvolved to determine the proteins that are adjacent in the case of a large complex (inter-protein crosslinks), or it can be used to understand domain-domain distances for single proteins, as well (intra-protein crosslinks) [8,9]. Hydrogen-deuterium exchange (HDX) and various other footprinting (FP) approaches use mono-functional reagents that specifically label the backbone (in the case of HDX) [10] or side chains (in the case of FP) [1,11–13]. Changes in the reactivity of these reagents with respect to the specific sites on the protein can reveal the peptide (or sub-peptide) level details of secondary (HDX) or tertiary and quaternary (HDX and FP) structural changes.

In order to better understand the strengths and weaknesses of these evolving approaches, and determine the right integrative approach for a particular problem, we have prepared a summary of the salient features of major structural biology techniques familiar to most modern research laboratories. Table 1 lays out the major techniques, provides a graphic illustrating the method or results, lists the number of structures associated with that technique currently in the Protein Data Bank (https://www.rcsb.org/stats/summary), states limitations of molecular size or amounts of material associated with typical studies, and adds additional commentary on significance. The vast number of deposited structures, exceeding 160,000 (as of 2/2020), are dominated x-ray crystallography, due to the high-throughput nature of modern synchrotron data collection and structure solution. Other methods also make very important contributions, with NMR exceeding 12,000 structures and single particle analysis by Cryo-EM increasing quite rapidly with now over 4000 structures deposited. Structures

Table 1. Comparison of major structural biology techniques in terms of resolution, limitations, and other relevant figures of merit. Tomography figure courtesy of Chiu lab.

Technique (PDB deposits)		Size (Sample state)	Resolution Limits	Amounts	Notes
NMR (12835)		<100 kDa (solution)	~3–4 Å	μmoles/milligrams	Requires isotopically labeled recombinant protein, disordered regions can be observed but may not be assigned
X-ray Crystallography (141165)		Limited by crystal quality	<1–3 Å	μmoles/milligrams	Mutant constructs necessary for many membrane proteins, disordered regions invisible. Gold standard for structural water
Cryo-EM: Single particle (4092)		>100 kDa (vitrified ice)	Mostly >3 Å	nanomoles/μgrams	Resolution and size limits improving, best samples have symmetry, disordered regions invisible
Cryo-EM: Tomography		Cells or tissues	30-40 Å	thin sections/individual cells	Resolution improving; captures large-scale spatial organization in cells
SAXS		>10 kDa (solution)	>20 Å	nanomoles/μgrams	Native material can usually be used, (similar to FP samples)
Footprinting: HRF-MS [and HDX-MX]		HRF-MS: No limit [<100KDa for HDX] (solution)	Peptide to single-residue (single base for NA)	picomoles/nanograms	Native material can usually be used (both), absolute surface area can be estimated (HRF) disordered regions visible (HRF). Studies in cells/tissue possible (HRF)

based on integrative or hybrid methods, including low-resolution EM, cross-linking, footprinting, and SAXS, are also emerging, and several dozen structures have been deposited in a separate database, called PDB-Dev (https://pdb-dev.wwpdb.org/), designed specifically for these integrative approaches.

NMR is a valuable solution-based technique, most effective on proteins <100 kDa. Milligrams of material and micromolar concentrations of a sample are typical; some experiments require expensive isotope labeling and/or expensive expression systems for some proteins. Data collection times can be hours or days. X-ray crystallography, conducted at synchrotrons, is very high throughput, as hundreds of crystals can be analyzed in a single day. It has a very high-potential resolution and is currently essential for reliably identifying sites of water or metal ion occupancy. Protein stability issues and limitations of crystallization are challenges for membrane proteins and larger complexes [14,15]. Cryo-EM has provided a breakthrough for solving medium to high-resolution structures of large complexes and for membrane proteins, and sample amounts are considerably reduced, from the milligrams required for crystallography and NMR to micrograms for the Cryo-EM vitrification sample preparation process. An additional complexity is that molecules at the air/water interface can be denatured as a result of the vitrification and data collection times may also be hours/days. Another current limitation of Cryo-EM is that >90% of the structures do not exceed 3 Å resolution, and samples that lack symmetry or are under 100 kDa suffer relatively rapid radiation damage limiting data collection and or resolution [16,17]. Thus, NMR and Cryo-EM have completely orthogonal optimums for protein size. Similar to Cryo-EM, Cryo-electron tomography also has the potential to elucidate the molecular envelopes of large macromolecular complexes while operating *in situ,* albeit currently at resolutions of 20–30 Å [18,19]. Small-angle x-ray scattering can be applied in solution, even for relatively large complexes and provides shape information at ~20 Å resolution [20]. With the strengths and weaknesses of each of these techniques in mind, structural mass spectrometry can be quite useful to fill in gaps in both approach and information. Some advantages of MS include:

only nanograms of material are needed (three orders of magnitude less than cryo-EM and six orders less than crystallography/NMR); samples generally do not need to be engineered/re-engineered or labeled (e.g., can be native); samples are assessed in solution or other convenient matrices; samples can be of almost any size or complexity (with the caveat that HDX is less effective for systems >100 kDa due to back exchange); and MS experiments can provide high-resolution assessments of secondary, tertiary, and quaternary structure [1].

Fig. 1 illustrates the workflow we have adopted to answer key questions in structure assessment. The workflow's foundation includes high-resolution structural information from crystallography, NMR or cryo-EM, and relevant homology modeling, and then envisioning experiments that assess the temporal and spatial features of the assembly and dynamics of the molecular systems under study, particularly those that represent a perturbation of the ''canonical'' structure. It is often the case that crystallographic or Cryo-EM information is available only for one form of a complex of interest, e.g., with a ligand while the ''apo'' structure without a ligand may not be known. In these cases, techniques sensitive to local structural information (HDX and FP) can be used to infer the ligand-binding site and sites of potential allosteric change while global measures, such as native-MS, EM and SAXS can determine ligand dependent shape changes for the overall protein/complex. Repeated test and validation cycles where structural models are evaluated by mutagenesis or other orthogonal biophysics experiments are essential for establishing rigor of results.

A number of interesting and challenging problems in structural biology can now be solved by this enabling suite of approaches; some of these are outlined in Table 2. In this perspective, we illustrate novel integrations of SAXS, FP, and molecular docking to overcome these challenges to provide structure. In terms of potential applications, structures of macromolecular complexes, for example, antibody/antigen complexes and checkpoint interactions in immune cells, are of high priority for analysis, but some systems, due to disorder or size, are not amenable to standard approaches. In addition, the study of membrane proteins enabled by crystallography and cryo-EM can be followed up with MS approaches. In particular, FP has the potential to provide protein interaction and dynamic data for membrane proteins [21,22] as it is challenging to obtain models/structures of all the productive complexes involved in cellular signaling and function. For example, HDX and FP have recently leveraged high-resolution structural information to address the dynamics of signaling and the mechanisms of GPCR-G-protein complex assembly [23]. In this study, the details of the dynamics of interconversions between known states, including temporal and mechanistic details of interactions, were revealed, providing a new window into biological function. Specifically, an intermediate in the reaction to form the GPCR-G-protein complex was observed on the hundreds of milliseconds timescale, well before the actual cellular signaling events.

Key in integrative structural studies of challenging systems is the provision of a flexible, capable analytical toolbox to integrate disparate types of structural data; this has the potential to provide high-resolution structural information even in the absence of success via conventional methods. A further challenge is assessing native structures in at least solution-based, if not their physiological, cellular, and/or tissue context. Successful completion of such integrative, ''in (cellular) context'' studies will require integration of multiple-types of experimental data using varying computational strategies, including the integration of high-resolution structural information spanning both local and global scales. An additional set of unmet challenges is to obtain a readout of protein dynamics to both expand and validate these models while attaining the readouts at time resolutions sufficient to capture signatures of intermediate states of signaling and dynamics. Mass spectrometry plays a major role in sampling dynamics of structural proteomes for both *in vitro* systems and on cellular scales [24].

Table 2. Major opportunity areas for structural footprinting.

- Protein complexes/ligand interactions not amenable to crystallography or Cryo-EM
- Membrane proteins in native-like states
- Intrinsically disordered proteins
- Amyloidogenic proteins
- Structural kinetics and intermediates of protein machines
- Structure of molecules in cells/tissue

Footprinting for Structural Biology

Footprinting, as a structural tool, is a chemistry-based approach, wherein small molecules are designed to covalently modify macromolecules at sites of structural interest [1,25]. Hydroxyl radical footprinting (HRF) is the most popular approach for macromolecule modification or cleavage, although its utility can be reduced by the presence of scavenging elements in solution, including buffering agents or stabilizers needed to poise the macromolecule or cell in the biochemical state of interest. These scavengers soak up the dose of OH radicals, requiring chemical approaches to increase OH radical concentrations or timescales to accumulate sufficient dose to efficiently detect products. This is undesirable as it increases the potential for

secondary radical generation, complicating the analysis of primary radical reactions with the macromolecules, leading to significant challenges in acquiring reproducible or accurate data [26]. Due to these limiting factors, high-flux sources of hydroxyl radicals, using photolysis of peroxide [27–29] or radiolysis of water [1,11,13], have long been popular as ways to achieve sufficient [OH] radical concentrations to optimize labeling for successful footprinting.

Synchrotron resources have long been used as sources of OH radicals for footprinting, the high dose (HD) of radicals available has particular advantages including higher coverage of modifications in challenging samples, improved labeling of membrane proteins or prions, while HD applications are essential for enhanced prospects for footprinting in organelles or cells. Finally, HD applications enable time-resolved applications in conjunction with rapid mixing or other fast reaction initiation schemes. Recently, the Center for Synchrotron Biosciences CSB completed the construction and commissioning of a new state-of-the-art beamline for synchrotron footprinting at the National Synchrotron Light Source-II, the 17-BM X-ray Footprinting of Biomaterials (XFP) beamline [30], which delivers world-leading flux densities and photon dose capabilities ($>10^{16}$ photons/s, and up to 500 W/mm absorbed dose achievable). This new beamline (Fig. 2) is also adjacent to other cutting-edge NSLS-II structural biology beamlines for crystallography, small-angle scattering, and biological X-ray imaging, providing an integrated "Structural Biology Village" environment. This modern resource for synchrotron footprinting provides the ideal opportunity to rectify some of the limitations of present sample handling devices and develop an integrated platform for footprinting, which when coupled to new data analysis tools, can drive adoption of the technique by the broader structural biology community and the adoption of radiolytic HRF at other synchrotron beamlines both in the United States and internationally [30–36].

Integrating Footprinting and Stopped-Flow Kinetics

With the availability of footprinting beamlines capable of delivering sufficient flux densities to label proteins within microseconds in the context of flow-based exposure cells, studying single residue dynamics at the protein interface during the formation of a complex is clearly enabled on such timescales if the samples can be appropriately poised. This includes examining the structural transitions that govern the formation of a productive protein-protein binding interface or the protein dynamics that govern the interactions of a protein with its ligand partner. This approach provides a much-needed temporal window into the structures provided by X-ray crystallography and cryo-EM structural studies, which by their very nature, can

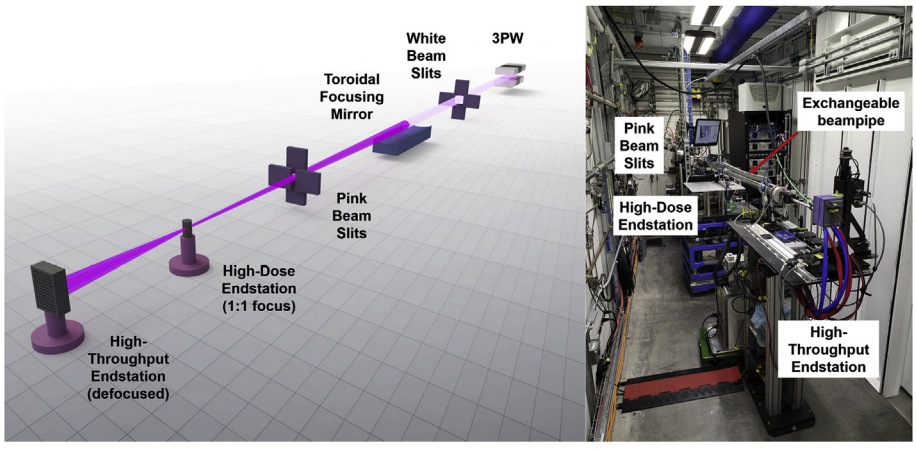


Fig. 2. XFP beamline at NSLS-II. Synchrotron white light (pale pink cylinder) is produced from a 3-pole wiggler (3PW) source at NSLS-II. White light passes through a set of beam-defining white-beam slits and then intercepts a toroidal mirror that focuses the X-ray light and rejects high-energy X-rays (>16 keV), producing "pink" beam (pink cylinder). The pink beam passes through the shield wall into the experimental hutch, where it intercepts beam defining pink beam slits and then achieves 1:1 focus at the high-dose footprinting endstation before diverging to a larger beam at the downstream high-throughput endstation. (right) photo of the XFP experimental hutch in the high-throughput X-ray footprinting configuration. The 3PW, white-beam slits, and toroidal focusing mirror are located upstream behind the shield wall and are not visible in this photo. The first element in the hutch is the pink-beam slits in a photon delivery chamber, followed by transport pipe to the high-dose and high-throughput endstations. Changes between the two endstations are accomplished by the removal or installation of an exchangeable beampipe, as required.

only reveal static snapshots of energetic intermediates/endpoints that may not fully represent or recapitulate aspects of the physiological transition. Furthermore, as HRF techniques can utilize native protein, HRF results have the potential to deliver a more physiologically relevant understanding of protein-protein and protein-ligand complex kinetic parameters in the absence of the conditions, mutations, crystal contacts, and constructs often required to stabilize proteins for structural studies. While competing techniques, such as HDX, are limited in their temporal resolution, essentially providing snapshots limited by the exchange rate of the label, the speed of OH radical generation using synchrotron radiation opens up new temporal vistas, enabling us to observe microsecond to millisecond (as well as longer) timescale events.

We recently employed the HRF technique in combination with stopped-flow kinetics to probe the structural and dynamic transitions involved in the formation of the β2-Adrenergic receptor (β2-AR) – Gs signaling complex (Fig. 3) after mixing the

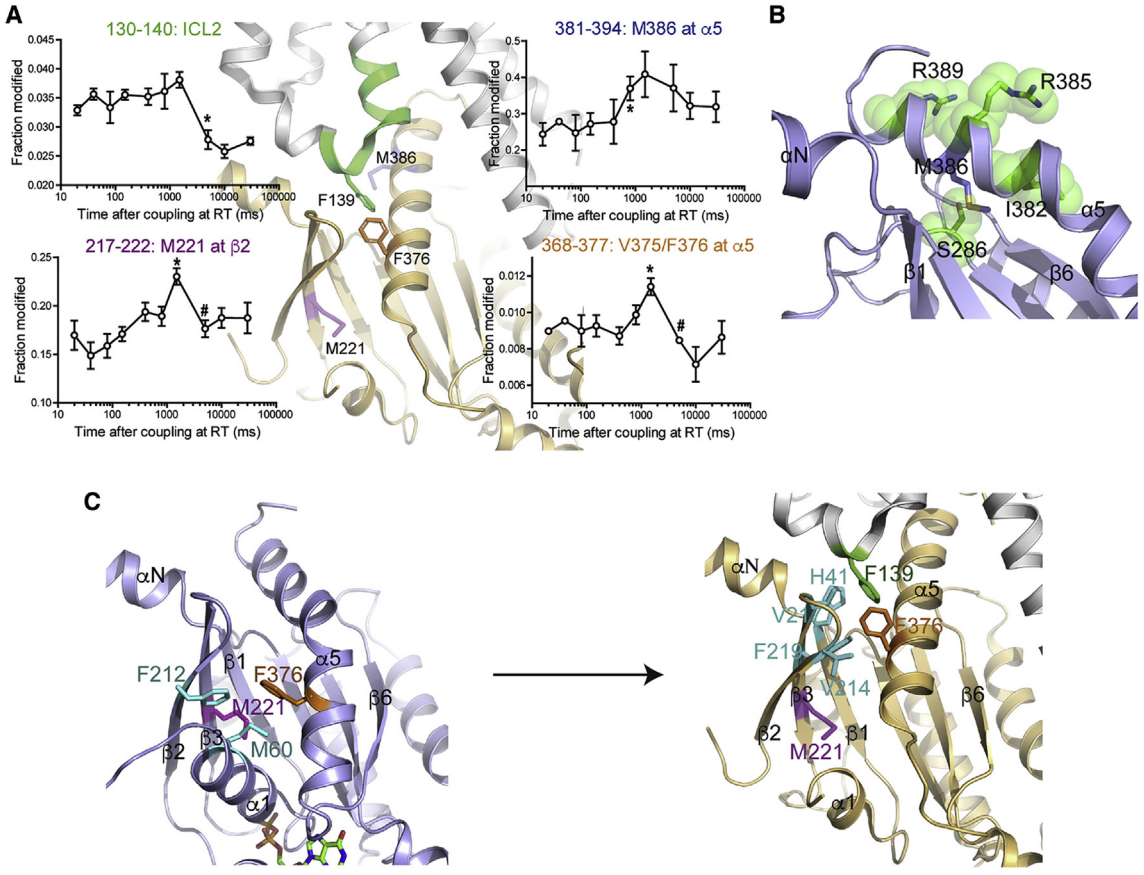


Fig. 3. Time-Resolved Analysis of GPCR-Gs Complex Formation by HRF-MS. Utilizing synchrotron-based HRF technologies in combination with hydrogen-deuterium mass spectrometry data, it is possible to both probe the formation of a protein complex, as well as the amino acid/peptide resolution changes that underlie the formation of a protein complex over timescales ranging from 10s of milliseconds to minutes. In this study, the changes in protein dynamics and solvent accessibility that underlie the formation of the β2-AR – Gs signaling complex were studied, yielding temporal data on the small scale changes that enable the GPCR to specifically recognize its cognate G protein and initiate allosteric/structural changes that enable the process of nucleotide exchange required for downstream signaling. (A) X-ray generated radiolytic oxidative modification profiles of selected peptides or residues from the β_2-AR or Gαs. Oxidative modification changes of Gαs upon incubation with the β_2-AR were analyzed. The modified peptides or residues are indicated as colored regions or sticks on the X-ray crystal structure of the β_2-AR – Gs complex (PDB: 3SN6). (B) The surrounding environment of M386. In the GDP-bound Gs structure, M386 is located within a pocket formed by four amino acids (green spheres) with limited solvent exposure. (C) Rearrangement of interactions with M221 and F376 of Gαs following the formation of the nucleotide-free β_2-AR – Gs complex. In the GDP-bound Gs structure, M221 and F376 form interactions with residues within β2-β3 strands and α1 helix (left), which are lost in the β_2-AR-bound nucleotide-free structure (PDB: 3SN6) (right). In the β_2-AR-bound nucleotide-free structure (PDB: 3SN6), F376 forms new interactions with F139 of the β_2-AR and amino acids in the αN/β1 hinge and β2/β3 loop (right). Adapted with permission [23].

individual species of G-protein and receptor. The structural detail of the formation of this complex is of great scientific interest as it elucidates aspects of heart rate and blood pressure signaling, is a target of a number of commonly prescribed blood pressure mediations, while also serving as a prototypical receptor for the understanding of G protein-coupled receptor signaling in general. This customized beamline front end of 17-BM allowed us to label (and thus observe) components of the β2-AR − Gs signaling complex at time points ranging from to milliseconds after mixing out to 10 min, providing a kinetic trace with amino acid resolution for several component regions of the receptor and G protein as they interact to form the signal initiating complex. This experimental methodology provides the first look into the residue level changes in β2-AR/Gs dynamics during complex formation, as well as providing a mechanistic understanding of Gαs subunit recognition by an activated receptor.

Integrations of Footprinting and SAXS

Advanced FP approaches can measure the relative or absolute solvent accessibility of side-chain residues in $Å^2$ via a protection factor (PF) analysis, and thus can provide a quantitative input to modeling exercises, where the predicted solvent accessibility from a model can be compared to experimentally derived solvent accessibility data from footprinting for scoring purposes [37,38]. FP, however, as it reports only local structural information, is not sufficient to understand the full context and scope of all relevant structural interactions. The synergism of FP with other complementary techniques such as SAXS, which provides overall protein shape information, can overcome the limitations of the individual techniques providing a more complete and comprehensive picture of a given system of interest.

Consistent with the data integration philosophy articulated in Fig. 1, the iSPOT (integration of Scattering, footPrinting, and dOcking simulaTion) software [20] has been developed to integrate crystallographic or other high-resolution structural models of macromolecules with SAXS and FP data to resolve macromolecular structures (Fig. 4). iSPOT combines multiple sources of structural data, and it builds on approaches pioneered by others, such as IMP, HADDOCK, and CNS [39–43]. CNS uses information from X-ray crystallography or NMR, and can also integrate data from EM, while HADDOCK can integrate NMR, SAXS and EM data. The IMP approach developed in the Sali group was initially focused on modeling using low-resolution EM data but has the capability to incorporate many different types of data (such as cryo-EM, X-ray crystallography and chemical cross-linking) for enabling the modeling of very large molecular complexes, such as the nuclear pore [43]. In contrast to these three methods, a key focus of iSPOT is the orthogonality and complementarity of FP and SAXS: the latter sensitive to protein shape and overall arrangement and the former sensitive to residue-specific solvent exposure. iSPOT fills a gap in modeling as it can address protein-protein complexes that are too large for NMR, too dynamic for crystallization, and/or too small/dynamic for cryo-EM. While iSPOT is the first approach to include FP data for structural modeling, the addition of HRF data has been shown to improve the overall performance of the Rosetta modeling for protein structure prediction as well [44].

Computational docking is central to iSPOT modeling by providing a basis-set of conformations to fit the SAXS and FP data simultaneously. One such method is straightforward rigid-body docking, where the known structures of individual components are treated as two rigid particles and fed into many available software packages, such as ClustPro [45]. The binding of two interacting proteins, however, often induces conformational changes. Under this circumstance, flexible docking is required to account for binding-induced structural changes. In fact, when such flexibility occurs at the amino acid level, the side chains at the interface can be optimized during the docking process, as implemented in HADDOCK [39]. In other scenarios, the conformational flexibility can go beyond the local side-chain changes, e.g., involving the tightening of secondary structure and packing at the interface due to the binding. As such, these induced-fit changes need to be considered and can be addressed. For iSPOT, coarse-grained molecular dynamics simulations are implemented to allow large-scale induced-fit structural changes [20,42]. Even larger-scale conformational changes, such as allosteric dynamics that travel beyond the protein interface, are quite challenging to accommodate; this has been one limiting factor of current integrative modeling approaches. Nonetheless, as computing power increases, integrative modeling is on the path to accurately and reliably characterize both rigid-body and flexible induced-fit protein-protein docking approaches.

Our success in integrating SAXS and FP datasets to model protein-protein interacting complexes [20,42] includes the determination of the multi-domain structure of the human estrogen receptor (Fig. 5a) [46]. Although the individual structures of the DNA and ligand-binding domains were known at high resolution, all previous attempts to solve the structure of the complex at high resolution were unsuccessful [47]. To understand the structure of the complex, we first measured its experimental protection factors (PF), which are rates of hydroxyl radical reactivity measured by experiment and then normalized by intrinsic reactivity of the individual amino acid

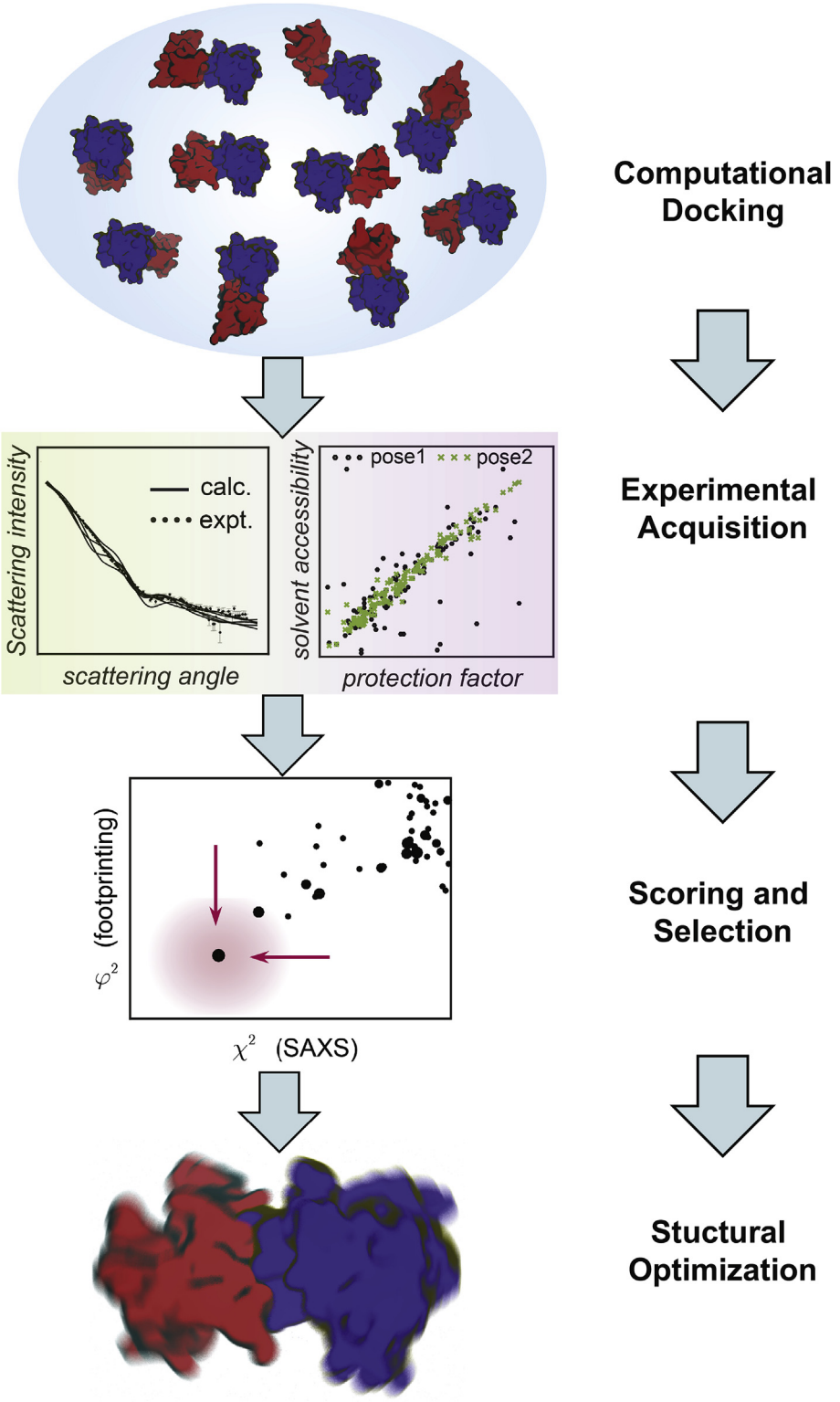


Fig. 4. The iSPOT workflow. It consists of four components: (a) computational protein-protein docking, (b) experimental SAXS and footprinting data acquisition, (c) scoring and selection, and (d) structural model optimization. Reprinted with permission [42].

side chains, for twenty independent side-chain sites in the complex. We then plotted the solvent-accessible surface area for the individual DNA and ligand-binding domains from crystallography vs. the PF for the 20 residues. Fourteen of these 20 sites obeyed the expected linear relationship of PF and solvent accessible surface area, indicating that the PF values accurately reflected the crystal structure's orientations for these residues. The other six did not obey the known relationship, all were much more protected (Higher PF) than expected, and were thus identified as potential interfacial residues. Instructively, three of the residues formed a tight patch on the DNA-binding domain and the other three formed a tight patch on the ligand-binding domain. Thus, an important interaction point between the two domains was clearly identified; this assumption was used to drive a molecular docking strategy (Fig. 4) that generated thousands of potential conformations that maintained this interface point. Scoring schemes were then employed to compare experimental FP and SAXS data to define a subset of the simulations that were self-consistent with all the data (Fig. 4). The novel 3-D arrangement of the ligand and DNA binding domains revealed a previously unknown interface at high resolution (~3–4 Å) and can be used to drive the development of small molecules to interact with and modulate estrogen receptor function.

In the above case, high-resolution structures of DNA- and ligand-binding domains were available and were essentially correctly docked by the application of SAXS, FP, and computation. However, this combination of techniques mediated by iSPOT can also be quite valuable even when virtually no structural information is available or when the macromolecules are present in an ensemble of distributed structures. For example, we have recently combined SAXS and FP data to describe the ensemble-structures of the estrogen receptor's N-terminal transactivation domain (NTD) [51], which is an intrinsically disordered protein (IDP) (Fig. 5b). In this latter case, FP protection factors, which are highly correlated with solvent accessible surface area (SA), were calculated for multiple residues in the NTD; these data reflected the average SA for all members of the ensemble [37,38,48]. Coupling these constraints with analysis of SAXS data [20,49–51], we determined that this IDP has a novel structured element of long-range contact, whose release is likely relevant to receptor function. Overall, we anticipate that the proposed multitechnique integrative modeling approach is well suited for studies of IDPs.

Challenges

One of the key challenges in these integrative studies is to coordinate experimental data acquisition of individual datasets, *i.e.*, from HRF and SAXS separately. For example, one of our recent samples is very prone to protein aggregation as the protein itself degrades within several hours after size-exclusion chromatography (SEC). For tackling this issue, we have performed SEC purification on-site, enabling identical, fresh protein samples to be

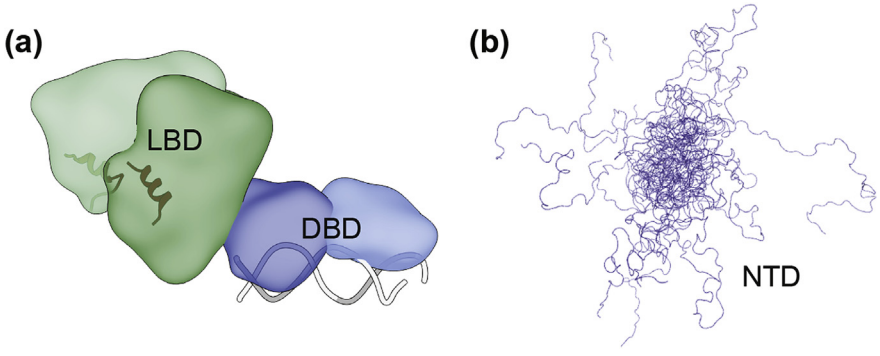


Fig. 5. Example of data integration via iSPOT. (a) Multidomain architecture of human estrogen receptor (ER) homodimer (ligand-binding domain or LBD in green and DNA-binding domain or DBD in blue). Neither SAXS or FP alone has been able to unambiguously depict a meaningful picture of the LDB-DBD complex, although the combination of structural information from FP (on a set of 20 residues, six of which are solvent-protected at the domain interface) and SAXS (shape and spatial distribution) is shown to successfully determine the ensemble-structures that are subsequently validated, structurally and functionally. Modified with permission [46]. (b) Ensemble-structures of the intrinsically disordered region of ER's N-terminal transactivation domain or NTD, where the FP probes a set of 16 residues along the protein amino acid sequence, joining forces with SAXS data and molecular dynamics simulations for the structural-ensemble characterization of the NTD as an IDP. Modified with permission [51].

employed for both HRF and SAXS data acquisition simultaneously, in which SAXS was performed at the NSLS-II LiX (16-ID) beamline, adjacent to the XFP (17-BM) beamline. This physical proximity is a key advantage to our data acquisition strategy and will be implemented with dedicated analytical coordination prior to our follow-up iSPOT-based structure determinations. Overall the NSLS-II's "Biology Village" project can be leveraged to promote novel integrated structural assessments of interesting biological systems, providing a one-stop-shop for the streamlining of multimodal structural analysis.

Conclusions

Structural biology has entered a new era, enabling us to answer complex questions of dynamics and multiprotein complex assembly. Integrated structural biology and continuing technology development will, going forward, enable the most challenging systems to be tackled. The extension of these studies to the physiological environment, to better understand the structural basis of critical functions in the cell is the next frontier.

Acknowledgements

This research was supported by grants R01-GM-114056 from the National Institute of General Medicine Sciences (NIH), the Mt. Sinai Healthcare Foundation. Support for the Center for Synchrotron Biosciences is through P30-EB-009998. Funding for development of 17-BM was provided by the National Science Foundation, Division of Biological Infrastructure (grant No. 1228549). 17-BM at the National Synchrotron Light Source II is a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. We thank Tiffany Bowman, Brookhaven National Laboratory Graphic Design, and Ricarda Laasch at the NSLS-II User Office for assistance with Fig. 2 graphics.

Conflict of interest

MRC is a scientific advisory board member and a shareholder of GenNext technologies and is also a Founder and Chief Scientific Officer of Neo Proteomics. JK is a consultant for Neo Proteomics.

Received 30 September 2019;
Received in revised form 7 February 2020;
Accepted 10 February 2020
Available online 21 February 2020

Keywords:
proteins;
mass spectrometry;
structural biology;
footprinting;
hybrid methods

References

[1] J. Kiselar, M.R. Chance, High-resolution hydroxyl radical protein footprinting: biophysics tool for drug discovery, Annu. Rev. Biophys. 47 (2018) 315–333.
[2] U. Pieper, R. Chiang, J.J. Seffernick, S.D. Brown, M.E. Glasner, L. Kelly, et al., Target selection and annotation for the structural genomics of the amidohydrolase and enolase superfamilies, J. Struct. Funct. Genom. 10 (2009) 107–125.
[3] Structural Genomics C, China Structural Genomics C, Northeast Structural Genomics C, Graslund S, Nordlund P, Weigelt J, et al. Protein production and purification, Nat. Methods 5 (2008) 135–146.
[4] L. Cerofolini, M. Fragai, E. Ravera, C.A. Diebolder, L. Renault, V. Calderone, Integrative approaches in structural biology: a more complete picture from the combination of individual techniques, Biomolecules 9 (2019).
[5] J.S. Kim, B. Afsari, G.S. Chirikjian, Cross-validation of data compatibility between small angle X-ray scattering and cryo-electron microscopy, J. Comput. Biol. 24 (2017) 13–30.
[6] J. Gault, C.V. Robinson, Cracking complexes to build models of protein assemblies, ACS Cent. Sci. 5 (2019) 1310–1311.
[7] A.C. Leney, A.J. Heck, Native mass spectrometry: what is in the name? J. Am. Soc. Mass Spectrom. 28 (2017) 5–13.
[8] J.E. Fajardo, R. Shrestha, N. Gil, A. Belsom, S.N. Crivelli, C. Czaplewski, et al., Assessment of chemical-crosslink-assisted protein structure modeling in CASP13, Proteins 87 (2019) 1283–1297.
[9] P.I. Koukos, A. Bonvin, Integrative modelling of biomolecular complexes, J. Mol. Biol. (2019), https://doi.org/10.1016/j.jmb.2019.11.009.
[10] J. Zheng, T. Strutzenberg, B.D. Pascal, P.R. Griffin, Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry, Curr. Opin. Struct. Biol. 58 (2019) 305–313.
[11] J.Q. Guan, S.C. Almo, E. Reisler, M.R. Chance, Structural reorganization of proteins revealed by radiolysis and mass spectrometry: G-actin solution structure is divalent cation dependent, Biochemistry 42 (2003) 11992–12000.
[12] J.G. Kiselar, R. Mahaffy, T.D. Pollard, S.C. Almo, M.R. Chance, Visualizing Arp2/3 complex activation mediated by binding of ATP and WASp using structural mass spectrometry, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 1552–1557.
[13] K. Takamoto, M.R. Chance, Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes, Annu. Rev. Biophys. Biomol. Struct. 35 (2006) 251–276.
[14] S. Maeda, A. Koehl, H. Matile, H. Hu, D. Hilger, G.F.X. Schertler, et al., Development of an antibody

fragment that stabilizes GPCR/G-protein complexes, Nat. Commun. 9 (2018) 3712.
[15] D.M. Rosenbaum, V. Cherezov, M.A. Hanson, S.G. Rasmussen, F.S. Thian, T.S. Kobilka, et al., GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function, Science 318 (2007) 1266–1273.
[16] M. Khoshouei, R. Danev, J.M. Plitzko, W. Baumeister, Revisiting the structure of hemoglobin and myoglobin with cryo-electron microscopy, J. Mol. Biol. 429 (2017) 2611–2618.
[17] M. Khoshouei, M. Radjainia, W. Baumeister, R. Danev, Cryo-EM structure of haemoglobin at 3.2 A determined with the Volta phase plate, Nat. Commun. 8 (2017) 16099.
[18] R.I. Koning, A.J. Koster, T.H. Sharp, Advances in cryo-electron tomography for biology and medicine, Ann. Anat. 217 (2018) 82–96.
[19] V. Lucic, A. Rigort, W. Baumeister, Cryo-electron tomography: the challenge of doing structural biology in situ, J. Cell Biol. 202 (2013) 407–419.
[20] W. Huang, K.M. Ravikumar, M. Parisien, S. Yang, Theoretical modeling of multiprotein complexes by iSPOT: integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking, J. Struct. Biol. 196 (2016) 340–349.
[21] S. Gupta, J. Chai, J. Cheng, R. D'Mello, M.R. Chance, D. Fu, Visualizing the kinetic power stroke that drives proton-coupled zinc(II) transport, Nature 512 (2014) 101–104.
[22] T. Orban, B. Jastrzebska, S. Gupta, B. Wang, M. Miyagi, M.R. Chance, et al., Conformational dynamics of activation for the pentameric complex of dimeric G protein-coupled receptor and heterotrimeric G protein, Structure 20 (2012) 826–840.
[23] Y. Du, N.M. Duc, S.G.F. Rasmussen, D. Hilger, X. Kubiak, L. Wang, et al., Assembly of a GPCR-G protein complex, Cell 177 (2019) 1232–1242 e11.
[24] N. de Souza, P. Picotti, Mass spectrometry analysis of the structural proteome, Curr. Opin. Struct. Biol. 60 (2019) 57–65.
[25] D.T. Johnson, L.H. Di Stefano, L.M. Jones, Fast photochemical oxidation of proteins (FPOP): a powerful mass spectrometry-based structural proteomics tool, J. Biol. Chem. 294 (2019) 11969–11979.
[26] G. Xu, M.R. Chance, Hydroxyl radical-mediated modification of proteins as probes for structural proteomics, Chem. Rev. 107 (2007) 3514–3543.
[27] B.C. Gau, J.S. Sharp, D.L. Rempel, M.L. Gross, Fast photochemical oxidation of protein footprints faster than protein unfolding, Anal. Chem. 81 (2009) 6563–6571.
[28] J.S. Sharp, J.M. Becker, R.L. Hettich, Analysis of protein solvent accessible surfaces by photochemical oxidation and mass spectrometry, Anal. Chem. 76 (2004) 672–683.
[29] D.M. Hambly, M.L. Gross, Laser flash photolysis of hydrogen peroxide to oxidize protein solvent-accessible residues on the microsecond timescale, J. Am. Soc. Mass Spectrom. 16 (2005) 2057–2063.
[30] A. Asuru, E.R. Farquhar, M. Sullivan, D. Abel, J. Toomey, M.R. Chance, et al., The XFP (17-BM) beamline for X-ray footprinting at NSLS-II, J. Synchrotron Radiat. 26 (2019) 1388–1399.
[31] A. Baud, L. Ayme, F. Gonnet, I. Salard, Y. Gohon, P. Jolivet, et al., SOLEIL shining on the solution-state structure of biomacromolecules by synchrotron X-ray footprinting at the Metrology beamline, J. Synchrotron Radiat. 24 (2017) 576–585.
[32] J. Bohon, R. D'Mello, C. Ralston, S. Gupta, M.R. Chance, Synchrotron X-ray footprinting on tour, J. Synchrotron Radiat. 21 (2014) 24–31.
[33] J. Bohon, M. Sullivan, J. Dvorak, D. Abel, J. Toomey, M. Chance, Development of the XFP beamline X-ray footprinting at NSLS-II, in: 12th International Conference on Synchrotron Radiation Instrumentation, AMerican Institute of Physics, Inc, 2016.
[34] S. Gupta, R. Celestre, C.J. Petzold, M.R. Chance, C. Ralston, Development of a microsecond X-ray protein footprinting facility at the Advanced Light Source, J. Synchrotron Radiat. 21 (2014) 690–699.
[35] S. Gupta, J. Feng, M. Chance, C. Ralston, Recent advances and applications in synchrotron X-ray protein footprinting for protein structure and dynamics elucidation, Protein Pept. Lett. 23 (2016) 309–322.
[36] S. Gupta, M. Sullivan, J. Toomey, J. Kiselar, M.R. Chance, The Beamline X28C of the Center for Synchrotron Biosciences: a national resource for biomolecular structure and dynamics experiments using synchrotron footprinting, J. Synchrotron Radiat. 14 (2007) 233–243.
[37] W. Huang, K.M. Ravikumar, M.R. Chance, S. Yang, Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: a protection factor analysis, Biophys. J. 108 (2015) 107–115.
[38] P. Kaur, J. Kiselar, S. Yang, M.R. Chance, Quantitative protein topography analysis and high-resolution structure prediction using hydroxyl radical labeling and tandem-ion mass spectrometry (MS), Mol. Cell. Proteomics 14 (2015) 1159–1168.
[39] C. Dominguez, R. Boelens, A.M. Bonvin, HADDOCK: a protein-protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.
[40] D. Russel, K. Lasker, B. Webb, J. Velazquez-Muriel, E. Tjioe, D. Schneidman-Duhovny, et al., Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies, PLoS Biol. 10 (2012), e1001244.
[41] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, et al., Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr D Biol Crystallogr 54 (1998) 905–921.
[42] A. Hsieh, L. Lu, M.R. Chance, S. Yang, A practical guide to iSPOT modeling: an integrative structural biology platform, Adv. Exp. Med. Biol. 1009 (2017) 229–238.
[43] F. Alber, S. Dokudovskaya, L.M. Veenhoff, W. Zhang, J. Kipper, D. Devos, et al., Determining the architectures of macromolecular assemblies, Nature 450 (2007) 683–694.
[44] M.L. Aprahamian, E.E. Chea, L.M. Jones, S. Lindert, Rosetta protein structure prediction from hydroxyl radical protein footprinting mass spectrometry data, Anal. Chem. 90 (2018) 7721–7729.
[45] D. Kozakov, D.R. Hall, B. Xia, K.A. Porter, D. Padhorny, C. Yueh, et al., The ClusPro web server for protein-protein docking, Nat. Protoc. 12 (2017) 255–278.
[46] W. Huang, Y. Peng, J. Kiselar, X. Zhao, A. Albaqami, D. Mendez, et al., Multidomain architecture of estrogen receptor reveals interfacial cross-talk between its DNA-binding and ligand-binding domains, Nat. Commun. 9 (2018) 3520.

[47] P. Yi, Z. Wang, Q. Feng, G.D. Pintilie, C.E. Foulds, R.B. Lanz, et al., Structure of a biologically active estrogen receptor-coactivator complex on DNA, Mol. Cell 57 (2015) 1047–1058.
[48] M. Gustavsson, L. Wang, N. van Gils, B.S. Stephens, P. Zhang, T.J. Schall, et al., Structural basis of ligand interaction with atypical chemokine receptor 3, Nat. Commun. 8 (2017) 14135.
[49] K.M. Ravikumar, W. Huang, S. Yang, Fast-SAXS-pro: a unified approach to computing SAXS profiles of DNA, RNA, protein, and their complexes, J. Chem. Phys. 138 (2013), 024112.
[50] S.C. Yang, S. Park, L. Makowski, B. Roux, A rapid coarse residue-based computational method for X-ray solution scattering characterization of protein folds and multiple conformational states of large protein complexes, Biophys. J. 96 (2009) 4449–4463.
[51] Y. Peng, S. Cao, J. Kiselar, X. Xiao, Z. Du, A. Hsieh, et al., A metastable contact and structural disorder in the estrogen receptor transactivation domain, Structure 27 (2019) 229–240 e4.