STRUCTURAL SNAPSHOT

Integrative structure determination of histones H3 and H4 using genetic interactions

Ignacia Echeverria[1,2,3] (iD), Hannes Braberg[1,2] (iD), Nevan J. Krogan[1,2,4] (iD) and Andrej Sali[2,3,5] (iD)

1 Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA, USA
2 Quantitative Biosciences Institute, University of California, San Francisco, CA, USA
3 Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA
4 Gladstone Institute of Data Science and Biotechnology, J. David Gladstone Institutes, San Francisco, CA, USA
5 Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, USA

Integrative structure modeling is increasingly used for determining the architectures of biological assemblies, especially those that are structurally heterogeneous. Recently, we reported on how to convert *in vivo* genetic interaction measurements into spatial restraints for structural modeling: first, phenotypic profiles are generated for each point mutation and thousands of gene deletions or environmental perturbations. Following, the phenotypic profile similarities are converted into distance restraints on the pairs of mutated residues. We illustrate the approach by determining the structure of the histone H3–H4 complex. The method is implemented in our open-source IMP program, expanding the structural biology toolbox by allowing structural characterization based on *in vivo* data without the need to purify the target system. We compare genetic interaction measurements to other sources of structural information, such as residue coevolution and deep-learning structure prediction of complex subunits. We also suggest that determining genetic interactions could benefit from new technologies, such as CRISPR–Cas9 approaches to gene editing, especially for mammalian cells. Finally, we highlight the opportunity for using genetic interactions to determine recalcitrant biomolecular structures, such as those of disordered proteins, transient protein assemblies, and host–pathogen protein complexes.

Introduction

Many proteins function by forming macromolecular assemblies that may also include other components. Structure determination of these assemblies is essential for a mechanistic understanding of their function. However, only a fraction of the structures of these assemblies have been obtained by traditional structural biology methods, such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and electron microscopy (EM). These methods require pure samples of the studied system, which are often difficult

Abbreviations

AF, AlphaFold; CG-MAP, chemical genetics miniarray profile; DMS, deep mutational scanning; EM, electron microscopy; IDPs, intrinsically disordered proteins; IMP, Integrative Modeling Platform; MIC, maximal information coefficient; NMR, nuclear magnetic resonance; pE-MAP, point-mutant epistatic miniarray profile; RMSD, root-mean-square deviation; SAXS, small-angle X-ray scattering; XL-MS, chemical cross-linking mass spectrometry.

to obtain, have low stability, or are conformationally heterogeneous.

Integrative/hybrid structure determination methods are powerful tools for determining the structures of macromolecular assemblies [1–3]. These methods can take as input experimental data from different sources, followed by computing a model whose properties satisfy the input information within the uncertainty of the data. Sources of input information commonly used for integrative modeling include the atomic structures of the components obtained from X-ray crystallography, NMR spectroscopy or comparative modeling, chemical cross-links obtained with mass spectrometry (XL-MS), solution scattering data from small-angle X-ray scattering (SAXS), and protein–protein interactions from affinity co-purification. Over the years, structures of protein complexes obtained using integrative methods have been used to explain the architecture [3–6] and evolutionary principles [3,6] of large assemblies, rationalize the effect of disease mutations [3,7], and describe the structural heterogeneity of flexible protein complexes [8–10].

There is a growing need for *in vivo* data that can be used for integrative structure determination [11]. Structural models based on *in vivo* data may be more representative of protein complexes in their native environment, thereby decreasing the risk of producing structures of nonfunctional states or missing relevant functional states. Thus, they may be more useful, including for understanding the role of structural changes that take place in disease pathogenesis. Data such as single-molecule Förster resonance energy transfer spectroscopy [12,13], protein footprinting [14], and XL-MS [15,16] can be collected *in vivo*. However, these methods are usually low throughput or produce sparse structural data.

Here, we review integrative structure determination using point-mutant epistatic miniarray profile (pE-MAP [17–19]) genetic interaction data, as implemented in our open-source *Integrative Modeling Platform* (IMP) program (http://integrativemodeling.org; [20]). First, we illustrate the approach by describing structure determination of the yeast histone H3–H4 complex based on ~ 500 000 pairwise genetic interactions between 350 histone mutants and a library of gene deletions. Second, we compare genetic interaction data to other data types commonly used for integrative modeling. Third, we describe how genetic interaction data are complementary to deep-learning protein structure prediction approaches. Fourth, we outline how the pE-MAP data could be obtained for mammalian cells. Finally, we describe how this approach can potentially be used to determine the structural ensembles of intrinsically disordered proteins (IDPs), transient protein assemblies, and host–pathogen protein assemblies.

Determining the structures of biomolecular assemblies using quantitative genetic interaction mapping

Genetic interactions report on how the effect of one mutation is affected by the presence of a second mutation. For example, when considering the effect of two mutations on a phenotype such as cell growth, a genetic interaction is classified as positive (epistatic or suppressive) when the combination of mutations has a lesser effect on cell growth than the multiplicative growth defect of the two single mutants, resulting in healthier cells. Conversely, a genetic interaction is classified as negative (synthetic sickness) if the double mutant displays a slower growth phenotype than is expected from the combination of individual mutation effects. Genetic interactions can be used to identify functional relationships among genes, including biological pathways [21–23] and protein complexes [24–26]. Genetic interaction profiles, defined as a set of genetic interactions between a given mutation (e.g., a point mutation) and a library of secondary mutations (e.g., gene deletions), often provide signatures of protein functions, allowing us to compare the profiles to hierarchically organize sets of proteins. For example, in genetic interaction maps in which mutations correspond to gene deletions, genes encoding proteins that are part of protein complexes or that function together in the same pathway often display similar genetic interaction profiles. Consequently, genetic interaction maps are usually analyzed by using standard clustering algorithms to predict gene functions [27–29].

Genetic interaction mapping can also be applied at a fine-grained scale by introducing point mutations in proteins to investigate the interactions between components within protein assemblies. We reasoned that point mutations that reside at or in close vicinity of a functional region (e.g., protein binding interfaces or active sites) within a macromolecule may have more similar phenotypes than point mutations that are distant in 3D space. Correspondingly, we hypothesized that the similarity of phenotypic profiles between a pair of point mutations in an assembly measured *in vivo* may be used to extract structural information about these assemblies. To this end, we used pE-MAP [17] to construct phenotypic profiles for point mutations crossed against single gene deletions or hypomorphic alleles. We designed distance restraints based on a pE-MAP dataset and the atomic structure of the complex between histones H3 and H4 extracted from the nucleosome X-ray structure [32]; the pE-MAP included 350 point mutations in histones H3 and H4 crossed

against an array of ~ 1370 gene deletion alleles. To set up the restraint, we quantified the similarity of each pairwise combination of phenotypic profiles using the maximal information coefficient (MIC) [30,31]. We observed that the similarities between the phenotypic profiles (i.e., MIC values) do not linearly correlate with the distances spanned by the mutated residues in the H3 and H4 structure (Fig. 1), but they are informative about an upper distance bound between the residues. This observation justified formulating a Bayesian scoring function to restrain the upper bound on the distance spanned by the pair of mutated residues, based on the pE-MAP data (Fig. 1). The distance restraints derived from the histone pE-MAP were then used for integrative structure modeling using our standard protocol implemented in the IMP program.

Structure of the histone H3–H4 complex computed using pE-MAP-derived restraints

The integrative modeling workflow implemented in IMP iterates through four stages (Fig. 2) [1,2,20,33]: (a) gathering all available experimental data and prior information (physical theories, statistical analyses, and other prior models); (b) translating information into representations of model components and a scoring function for ranking alternative models; (c) sampling models guided by the scoring function; and (d) validating the model. We now discuss each of these stages, using integrative modeling of the histones H3–H4 as an example [18].

First, the input information included the histones H3 and H4 pE-MAP dataset. Additionally, we used comparative models of H3 and H4 to mimic realistic integrative modeling.

Second, the molecular representation and scoring function were formulated based on the input information. Each subunit (i.e., H3 and H4) was represented as a rigid body based on the comparative models. The scoring function ranks alternative models based on the input information. Three scoring terms were used to restrain the structural degrees of freedom of the H3–H4 system. The defining and most important scoring term is extracted from the pE-MAP data. The H3–H4 pE-MAP results in 170 high MIC values (> 0.3) between subunits. We used these values to formulate a Bayesian scoring term that restrained the distances spanned by the mutated residues. The two other scoring terms encoded the sequence connectivity and excluded volume.

Third, structural models that satisfy the input information were computed using replica exchange Gibbs sampling, based on the Metropolis Monte Carlo algorithm [4,34,35], starting with random initial configurations of the components.

Finally, the uncertainty (precision) of the model was estimated to be 1 Å (Fig. 3); the model precision is defined as the variability among the ensemble of structural models that satisfy the input data within acceptable tolerances [36,37]. In addition, the usefulness of pE-MAP data for integrative structure modeling was demonstrated by the accuracy of the model of 3.8 Å; the accuracy is defined as the average Cα RMSD between each of the ensemble models and the X-ray structure. To further assess the information content of the pE-MAP data, we mapped the effect of the number of pE-MAP distance restraints on the accuracy of
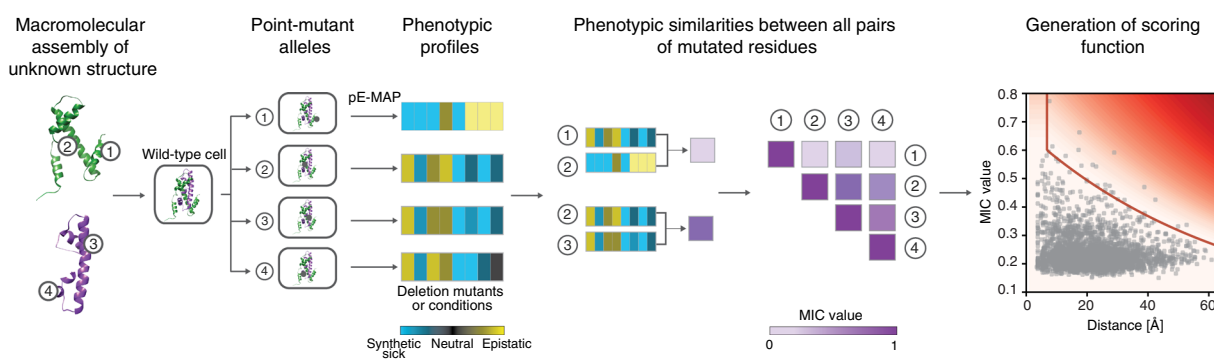

Fig. 1. Converting quantitative genetic interactions into spatial restraints for integrative structure modeling. First, to generate a pE-MAP a collection of point mutations is constructed by systematic mutagenesis of the genes encoding the subunits of the macromolecular assembly of interest (mutations 1–4). Second, the point-mutant strains are crossed against a library of gene deletions. Third, the phenotypic profiles are obtained by measuring the genetic interaction scores. Fourth, the pairwise phenotypic profile similarities are transformed into single MIC values to quantify the similarity between phenotypic profiles. Finally, the MIC values are translated into an upper distance bound (red curve) scoring term. In the plot, the background color gradient reflects the dependence of the scoring term on the MIC value and distance (darker colors represent higher scores). Figure partially reproduced from Ref. [18].
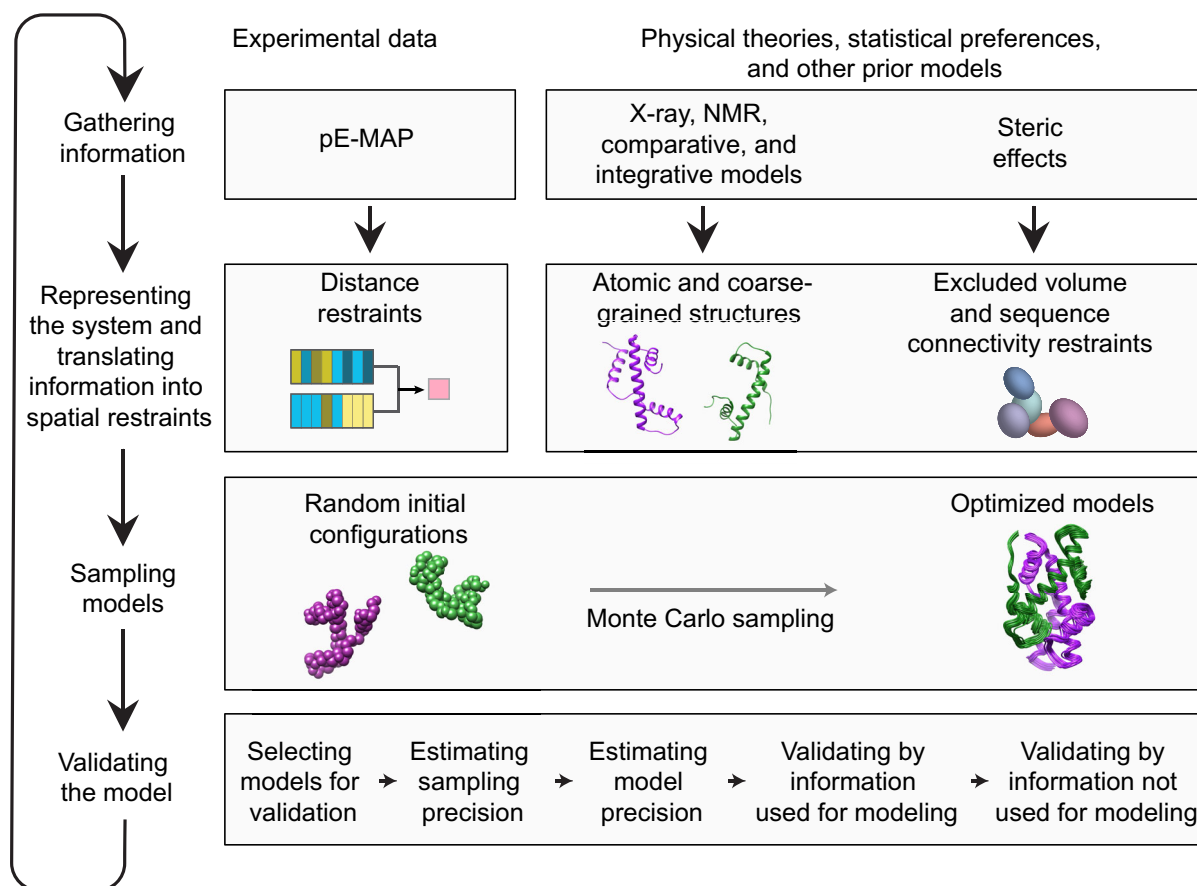
Fig. 2. Description of the integrative structure determination of histones H3 and H4. In this example, histones H3 and H4 are represented as rigid bodies. The scoring function consists of upper distance bound restraints that are derived from the pE-MAP data and restraints to account for sequence connectivity and excluded volume. The sampling searches for the structures that satisfy the spatial restraints indicated by the input information. The result is an ensemble of model structures that sufficiently satisfy the input information (e.g., within acceptable tolerances as indicated by the data). Validation of the model includes computing its precision and assessing the degree of consistency between the model and the input information used to and not used to compute it.

the models. The model accuracy improved from 12.7 to 3.8 Å when the number of pE-MAP-derived distance restraints increased from 68 (40% of all available restraints) to 170 (100% of all available restraints), indicating that the more pE-MAP data that are used, the more accurate is the model, as expected (Fig. 3C). Similarly, the precision of the model also improves as more data are used (Fig. 3D).

Comparison of restraints derived from pE-MAPs, coevolution, and biophysical methods

Integrative structure determination is based on the proposition that the complementarity between different sources of input information minimizes the drawbacks of sparse, noisy, ambiguous, and incoherent datasets. Consequently, broadening the types of input datasets and prior information used for integrative modeling would improve the accuracy, precision, and completeness of the resulting models [8,18]. Since the pE-MAP data rely on genotype-to-phenotype experiments, the spatial restraints derived from these datasets are orthogonal to biophysical data derived from, for example, *in vitro* samples. Thus, pE-MAP data have great potential for genetics-based structural modeling.

We have compared the precision and accuracy of models obtained using pE-MAPs to those of models obtained using data from XL-MS and coevolution analysis. These data types can also be converted to distance restraints [4,38,39]. In general, the accuracy and/ or precision of the models improve when pE-MAP
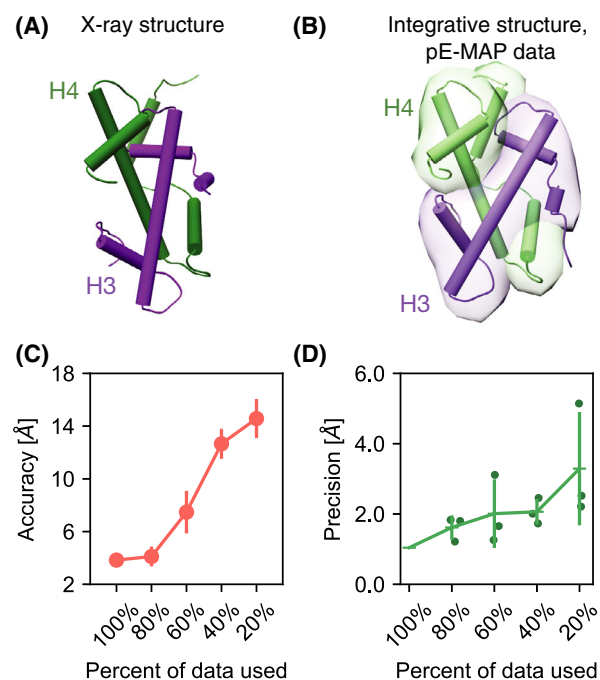
Fig. 3. Integrative structure of the histone H3 and H4 complex. (A) The X-ray structure of the histone H3–H4 dimer (PDB 1ID3, [32]). (B) Representative structure computed using integrative modeling embedded in the localization probability density. The localization probability density map represents the probability of any volume element being occupied by a protein, given the model ensemble. (C) Accuracy of models in the ensemble based on the full pE-MAP dataset and resampled datasets with only fractions of the data; the accuracy is defined as the average Cα RMSD between each of the structures in the ensemble and the X-ray structure. The standard deviations are shown as error bars. (D) Model precisions are based on the full and resampled pE-MAP datasets; the model precision is defined as the average RMSD between all solutions in the ensemble. Each dot represents the model precision of each of three independent realizations; error bars correspond to the standard deviation. Figure partially reproduced from Ref. [18].

and other data types are used together, demonstrating the premise of integrative structure determination. However, differences between these three data types highlight their relative strengths and synergy. For example, whereas a cross-link between two residues may provide more direct evidence of structural proximity than the corresponding pE-MAP pair, the number of potential cross-links is constrained by the number of reactive residues. In contracts, introducing a single point mutation increases the number of potential pE-MAP restraints quadratically. Whereas the number of cross-links can be increased by using cross-linkers that target different residue types [8], the effect of point mutations in a pE-MAP can be increased by specifically introducing surface point mutations and/or

targeting sites known to be functionally important, and by choosing substitutions likely to perturb protein–protein interactions. Coevolution-derived distance restraints are a promising way to obtain information about protein–protein interactions at a residue level [40–42]. However, the success of this approach is determined by the input sequence alignment depths and the identification of pairs of physically interacting proteins in genomes with multiple paralogs [41–43]. Recently, genetic interactions measured using deep mutational scanning (DMS) [44–47] and experimental evolution [48] approaches have also been successful in determining the structures of protein monomers and small protein complexes, highlighting the applicability of genetics data for structural biology.

Opportunities for integrative structure determination using structural models of subunits computed using deep-learning approaches

Recent advances in deep learning have revolutionized protein structure prediction; for example, AlphaFold (AF) [49] and RoseTTAfold [50] can often produce protein models with atomic accuracies comparable to those of experimental methods [51,52]. Deep-learning approaches use neural networks to benefit from the evolutionary information encoded in multiple sequence alignments as well as physical and geometric information derived from known protein structures to compute spatial relationships between the amino acid residues in a protein chain. Additionally, these deep-learning approaches show promise in modeling small protein complexes using neural networks trained with monomeric and multimeric structures [53–55]. *De novo* structural models of domains, subunits, and potentially subcomplexes can be used as building blocks for integrative modeling of entire protein assemblies. In particular, these building blocks can be combined with genetic interactions and other sources of information to compute the structural ensembles of protein assemblies in multiple conformational, compositional, and oligomeric states. While the deep-learning approaches rely on a multiple sequence alignment that recapitulates the natural sequence variation that potentially sparsely samples the sequence landscape under a variety of selection pressures, high-throughput genetic perturbation experiments used to determine genetic interactions characterize the sequence–function landscape under controlled experimental conditions. Importantly, genetic perturbation experiments with a phenotypic readout (pE-MAP) or selection experiments (DMS) that target specific cellular functions allow us to study proteins in the context of their biological functions. Notably, these experiments can

be performed under varying conditions to inform about different functional states accessed by the proteins. Thus, we propose that genetic interaction mapping is complementary to deep-learning approaches, providing additional information to predict the full range of biologically relevant structures of protein assemblies, including those that could be modulated by binding patterns or external factors. Finally, we expect that future deep-learning methods will be improved to allow incorporating experimental information specific to the modeled system [56], thus bridging the gap with integrative structure modeling [1,2].

Opportunities for genetic interaction mapping using CRISPR–Cas9-mediated gene editing

Recent advances in CRISPR–Cas9 approaches have enabled the prospect of large-scale precision gene editing [57] that expand the scale and scope of mutation libraries. For example, CRISPR–Cas9 gene editing can potentially be used to efficiently generate chemical genetics miniarray profiles (CG-MAP) [58]. In this approach, phenotypic profiles are generated by subjecting point-mutants to different environmental perturbations such as temperature and chemical stresses [58]. We have proven that the genetic profiles obtained using CG-MAPs provide structure–function relationships that can be used for integrative structure determination [18] and are generally cheaper and less laborious to generate than pE-MAPs.

One exciting application of integrative modeling based on pE-MAP and CG-MAP data is *in vivo* structure determination of mammalian protein assemblies. Even though human combinatorial CRISPR–Cas9 perturbations have been used to generate genetic interaction maps using gene knockouts or knockdowns [26,59,60], these approaches currently do not generate point mutations at sufficiently high efficiency to incorporate in pE-MAPs or CG-MAPs. However, as gene editing methods improve [61,62], they will likely open up for point-mutant genetic interaction mapping in mammalian cells [11]. In this context, unbiased analyses of genetic interactions could potentially serve a twofold purpose: systematic identification of protein pathways and assemblies, followed by structure determination of these assemblies. Additionally, the precise and systematic incorporation of point mutations into proteins or protein assemblies of interest will enable us to characterize the structural changes associated with disease alleles and their functional effects.

Most genetic interaction mapping initiatives use growth rate or cell fitness as the phenotypic readout, which can conceal the diversity of phenotypes emerging from the combinatorial expression of genes. To maximize structural information that can be recovered from comparing phenotypic profiles, genetic interactions could be quantified using alternative phenotypic readouts that are targeted to the system of interest. For example, phenotypic readouts such as reporter gene expression [63–65], single-cell transcriptomic [66,67], or high-content imaging [68] might provide system-specific signals that can be converted into distance restraints for integrative structure modeling. Moreover, as different phenotypic readouts may not necessarily provide the same information [69], we hypothesize that they can be incorporated into integrative modeling as complementary sources of information or interpreted as high-dimensional cell states (e.g., manifold [70]).

Opportunities for integrative determination of the structural ensembles of disordered proteins

The H3–H4 pE-MAP data reveal functional relationships between histone residues and between histone residues and other associated complexes. The distribution of MIC values for the histone tail–core and tail–tail mutation pairs is comparable to that of the core–core mutation pairs. Consequently, distance restraints for the histone tails may also be derived from the pE-MAP data. Genetic profiles from the histone tail mutations result in 390 distance restraints, which we expect will narrow the model space accessible to these tails. Thus, integrative modeling of IDPs based on genetic interaction data is likely feasible.

Structural characterization of IDPs remains challenging largely because of the rapidly interchanging conformations in the unfolded ensemble. Although challenging, accurate description of some IDP conformational ensembles has been possible using integrative approaches based on experimental data from NMR spectroscopy, SAXS, and single-molecule FRET [71,72]. Successful examples of IDPs modeling using integrative approaches rely on a Bayesian definition of the most probable ensemble of structural models, given the input information [73,74]. Data derived from *in vivo* measurements of genetic interactions would greatly complement the data types currently used to model IDPs and could be used to help elucidate the effects of macromolecular crowding in the structures and dynamics of IDPs. Furthermore, the pE-MAP data collected *in vivo* could be used to probe whether or not posttranslational modifications (PTMs) or other cellular perturbations regulate IDPs' functions by shifting the distribution of conformations in the structural ensembles.

Opportunities for integrative structure determination of transient protein assemblies

Using pE-MAP for integrative structure determination provides new opportunities to determine the structures of protein assemblies that are difficult to isolate or purify, such as transiently stable associations. Transient interactions involve proteins that dissociate readily and are often modulated by physiological conditions, environmental perturbations, or PTMs; they may also include binding interfaces that are flexible [75]. Thus, the conditions modulating these transient interactions might not be reproducible *in vitro*.

We showed that the H3–H4 pE-MAP can identify relationships between individual modifiable histone residues and their cognate enzymes, which are unlikely to be stably associated [18]. In principle, the current pE-MAP can be extended to include point mutations in these enzymes and use these genetic profiles to derive distance restraints between the histones' tails and their cognate enzymes. Describing the structures of transient interactions requires an ensemble representation that captures the interconverting conformational states of the proteins and the alternative specific and nonspecific binding configurations. To this end, protein–protein binding events can be characterized by creating equilibrium ensembles that include different specific and nonspecific protein assemblies using methods such as Monte Carlo or Brownian dynamics simulations [76,77]. Following, the structures of the specific transient protein assemblies and their binding affinities can be obtained by using a scoring function that includes scoring terms derived from genetic interaction measurements, coarse-grained energy functions, and other sources of information such as coevolution. As described for modeling of IDPs, independent high-resolution experimental data are necessary to validate the structures of transient protein assemblies obtained using distance restraints derived from pE-MAP data. Experiments using paramagnetic relaxation enhancement [76,78,79] and solid-state NMR [80,81] have proven successful to visualize transient protein–protein interactions at an atomic resolution and can be used for validation.

Opportunities for integrative structure determination of host–pathogen macromolecular assemblies

We have achieved progress in developing methods for integrative structure determination of host–pathogen complexes, primarily based on cross-linking data [82–84]. However, structural characterization of host–pathogen protein assemblies remains challenging, largely due to their compositional and conformational heterogeneity. For instance, a high proportion of pathogen proteins are intrinsically disordered or contain intrinsically disordered regions [85–87], making these proteins and the protein assemblies not amendable to traditional structural biology approaches. Consequently, structure determination of host–pathogen protein assemblies would greatly benefit from an orthogonal source of information, especially of data collected *in vivo*.

The structure of host–pathogen protein assemblies and their role in infection can be characterized by measuring the phenotypic effects of introducing mutations in the host and pathogen proteins and using genetic interaction profiling of selected host genes. Using pE-MAP in infected cells also opens the possibility of studying intraviral interactomes in the host's context. Intraviral protein interactions are crucial for the viral structure as well as replication and transcription of viral genomes [88–91]. However, the role of other virus–virus protein interactions, if any, is still not well characterized [92,93]. Another application of integrative structure determination based on genetic interactions is defining the structural preferences of large, multidomain pathogen proteins under different conditions/environments. One such example is the multifunctional Nsp3 proteins of coronaviruses, which are known to be essential components of the replication/transcription complex. Nsp3 proteins have also been implicated in disrupting the expression of innate immunity genes, formation of double-membrane vesicles, and inhibition of IFN production, among others. Additionally, Nsp3s have been reported to interact with several other viral proteins, suggesting a pleiotropic role [94]. The full structure of Nsp3 is currently unknown, and there are still several uncharacterized domains. An integrative modeling approach based on pE-MAP data may help elucidate the spatial organization of the domains and their functional roles.

In conclusion, we anticipate that genetic interactions involving viral protein will serve to identify pathogenicity factors, characterize their structures, understand their functions, and help develop antiviral therapies.

Conflict of interest

The Krogan Laboratory has received research support from Vir Biotechnology and F. Hoffmann-La Roche. NJK has consulting agreements with the Icahn School of Medicine at Mount Sinai, New York, Maze Therapeutics, and Interline Therapeutics. He is a shareholder in Tenaya Therapeutics, Maze Therapeutics, and Interline Therapeutics, and is financially compensated by GEm1E Lifesciences, Inc. and Twist Bioscience Corp.

Author contributions

All authors contributed to writing the manuscript.

Data availability statement

All input data, scripts, and output results are available in Ref. [18] and at http://integrativemodeling.org/systems/pemap.

References

1 Rout MP, Sali A. Principles for integrative structural biology studies. *Cell*. 2019;**177**:1384–403.

2 Sali A. From integrative structural biology to cell biology. *J Biol Chem*. 2021;**296**:100743.

3 Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature*. 2018;**555**:475–82.

4 Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, et al. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics*. 2014;**13**:2927–43.

5 Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, et al. Molecular architecture of the yeast mediator complex. *Elife*. 2015;**4**:e08719.

6 Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, et al. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci USA*. 2012;**109**:1380–7.

7 Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, et al. Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol Cell*. 2015;**59**:794–806.

8 Gutierrez C, Chemmama IE, Mao H, Yu C, Echeverria I, Block SA, et al. Structural dynamics of the human COP9 signalosome revealed by cross-linking mass spectrometry and integrative modeling. *Proc Natl Acad Sci USA*. 2020;**117**:4088–98.

9 Molnar KS, Bonomi M, Pellarin R, Clinthorne GD, Gonzalez G, Goldberg SD, et al. Cys-scanning disulfide crosslinking and bayesian modeling probe the transmembrane signaling mechanism of the histidine kinase, PhoQ. *Structure*. 2014;**22**:1239–51.

10 Zhou CY, Stoddard CI, Johnston JB, Trnka MJ, Echeverria I, Palovcak E, et al. Regulation of Rvb1/Rvb2 by a domain within the INO80 chromatin remodeling complex implicates the yeast Rvbs as protein assembly chaperones. *Cell Rep*. 2017;**19**:2033–44.

11 Braberg H, Echeverria I, Kaake RM, Sali A, Krogan NJ. From systems to structure – using genetic data to model protein structures. *Nat Rev Genet*. 2022;1–13. Online ahead of print. https://doi.org/10.1038/s41576-021-00441-w

12 Long Y, Stahl Y, Weidtkamp-Peters S, Postma M, Zhou W, Goedhart J, et al. In vivo FRET–FLIM reveals cell-type-specific protein interactions in Arabidopsis roots. *Nature*. 2017;**548**:97–102.

13 Meyer BH, Martinez KL, Segura J-M, Pascoal P, Hovius R, George N, et al. Covalent labeling of cell-surface proteins for in-vivo FRET studies. *FEBS Lett*. 2006;**580**:1654–8.

14 Espino JA, Jones LM. Illuminating biological interactions with in vivo protein footprinting. *Anal Chem*. 2019;**91**:6577–84.

15 Kaake RM, Wang X, Burke A, Yu C, Kandur W, Yang Y, et al. A new in vivo cross-linking mass spectrometry platform to define protein–protein interactions in living cells. *Mol Cell Proteomics*. 2014;**13**:3533–43.

16 Chavez JD, Mohr JP, Mathay M, Zhong X, Keller A, Bruce JE. Systems structural biology measurements by in vivo cross-linking with mass spectrometry. *Nat Protoc*. 2019;**14**:2318–43.

17 Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, et al. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell*. 2013;**154**:775–88.

18 Braberg H, Echeverria I, Bohn S, Cimermancic P, Shiver A, Alexander R, et al. Genetic interaction mapping informs integrative structure determination of protein complexes. *Science*. 2020;**370**:eaaz4910.

19 Braberg H, Moehle EA, Shales M, Guthrie C, Krogan NJ. Genetic interaction analysis of point mutations enables interrogation of gene function at a residue-level resolution: exploring the applications of high-resolution genetic interaction mapping of point mutations. *BioEssays*. 2014;**36**:706–13.

20 Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol*. 2012;**10**:e1001244.

21 Martin H, Shales M, Fernandez-Piñar P, Wei P, Molina M, Fiedler D, et al. Differential genetic

interactions of yeast stress response MAPK pathways. *Mol Syst Biol.* 2015;**11**:800.

22 van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell.* 2010;**143**:991–1004.

23 Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell.* 2005;**123**:507–19.

24 Bandyopadhyay S, Kelley R, Krogan NJ, Ideker T. Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput Biol.* 2008;**4**:e1000065.

25 Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature.* 2007;**446**:806–10.

26 Du D, Roguev A, Gordon DE, Chen M, Chen S-H, Shales M, et al. Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat Methods.* 2017;**14**:577.

27 Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. *Science.* 2004;**303**:808–13.

28 Beltrao P, Cagney G, Krogan NJ. Quantitative genetic interactions reveal biological modularity. *Cell.* 2010;**141**:739–45.

29 Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science.* 2016;**353**:aaf1420.

30 Albanese D, Filosi M, Visintainer R, Riccadonna S, Jurman G, Furlanello C. Minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. *Bioinformatics.* 2013;**29**:407–8.

31 Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al. Detecting novel associations in large data sets. *Science.* 2011;**334**:1518–24.

32 White CL, Suto RK, Luger K. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J.* 2001;**20**:5207–18.

33 Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, et al. Determining the architectures of macromolecular assemblies. *Nature.* 2007;**450**:683–94.

34 Rieping W, Habeck M, Nilges M. Inferential structure determination. *Science.* 2005;**309**:303–6.

35 Swendsen RH, Wang JS. Replica Monte Carlo simulation of spin glasses. *Phys Rev Lett.* 1986;**57**:2607–9.

36 Saltzberg DJ, Viswanath S, Echeverria I, Chemmama IE, Webb B, Sali A. Using Integrative Modeling Platform to compute, validate, and archive a model of a protein complex structure. *Protein Sci.* 2021;**30**:250–61.

37 Viswanath S, Chemmama IE, Cimermancic P, Sali A. Assessing exhaustiveness of stochastic sampling for integrative modeling of macromolecular structures. *Biophys J.* 2017;**113**:2344–53.

38 Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS One.* 2011;**6**:e28766.

39 Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife.* 2014;**3**:e03430.

40 Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contain information about protein-protein interaction. *J Mol Biol.* 1997;**271**:511–23.

41 Bitbol A-F, Dwyer RS, Colwell LJ, Wingreen NS. Inferring interaction partners from protein sequences. *Proc Natl Acad Sci USA.* 2016;**113**:12180–5.

42 Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc Natl Acad Sci USA.* 2009;**106**:67–72.

43 Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *Elife.* 2014;**3**:e02030.

44 Newberry RW, Leong JT, Chow ED, Kampmann M, DeGrado WF. Deep mutational scanning reveals the structural basis for α-synuclein activity. *Nat Chem Biol.* 2020;**16**:653–9.

45 Jones EM, Lubock NB, Venkatakrishnan AJ, Wang J, Tseng AM, Paggi JM, et al. Structural and functional characterization of G protein–coupled receptors with deep mutational scanning. *Elife.* 2020;**9**:e54895.

46 Rollins NJ, Brock KP, Poelwijk FJ, Stiffler MA, Gauthier NP, Sander C, et al. Inferring protein 3D structure from deep mutation scans. *Nat Genet.* 2019;**51**:1170–6.

47 Schmiedel JM, Lehner B. Determining protein structures using deep mutagenesis. *Nat Genet.* 2019;**51**:1177–86.

48 Stiffler MA, Poelwijk FJ, Brock KP, Stein RR, Riesselman A, Teyra J, et al. Protein structure from experimental evolution. *Cell Syst.* 2020;**10**:15–24.e5.

49 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;**596**:583–9.

50 Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science.* 2021;**373**:871–6.

51 Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN. High-accuracy protein structure prediction in CASP14. *Proteins.* 2021;**89**:1687–99.

52 Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold 2 applications. *bioRxiv*. 2021 [PREPRINT]. https://doi.org/10.1101/2021.09.26.461876

53 Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. 2021 [PREPRINT]. https://doi.org/10.1101/2021.10.04.463034

54 Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *bioRxiv*. 2021 [PREPRINT].

55 Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, et al. Computed structures of core eukaryotic protein complexes. *Science*. 2021;**374**:eabm4805.

56 Rosenbaum D, Garnelo M, Zielinski M, Beattie C, Clancy E, Huber A, et al. Inferring a continuous distribution of atom coordinates from Cryo-EM images using VAEs. *arXiv [csCE]*. 2021. https://doi.org/10.48550/arXiv.2106.14108

57 Roy KR, Smith JD, Vonesch SC, Lin G, Tu CS, Lederer AR, et al. Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat Biotechnol*. 2018;**36**:512–20.

58 Shiver AL, Osadnik H, Peters JM, Mooney RA, Wu PI, Henry KK, et al. Chemical-genetic interrogation of RNA polymerase mutants reveals structure-function relationships and physiological tradeoffs. *Mol Cell*. 2021;**81**:2201–15.e9.

59 Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, et al. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat Methods*. 2017;**14**:573–6.

60 Stockman VB, Ghamsari L, Lasso G, Honig B, Shapira SD, Wang HH. A high-throughput strategy for dissecting mammalian genetic interactions. *PLoS One*. 2016;**11**:e0167617.

61 Ma L, Boucher JI, Paulsen J, Matuszewski S, Eide CA, Ou J, et al. CRISPR-Cas9–mediated saturated mutagenesis screen predicts clinical drug resistance with improved accuracy. *Proc Natl Acad Sci USA*. 2017;**114**:11751–6.

62 Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019;**576**:149–57.

63 Gordon DE, Watson A, Roguev A, Zheng S, Jang GM, Kane J, et al. A quantitative genetic interaction map of HIV infection. *Mol Cell*. 2020;**78**:197–209.e7.

64 Zhao R, Davey M, Hsu Y-C, Kaplanek P, Tong A, Parsons AB, et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell*. 2005;**120**:715–27.

65 Kang JH, Chung J-K. Molecular-genetic imaging based on reporter gene expression. *J Nucl Med*. 2008;**49**(Suppl 2):164S–79S.

66 Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell*. 2016;**167**:1853–66.e17.

67 Schraivogel D, Gschwind AR, Milbank JH, Leonce DR, Jakob P, Mathur L, et al. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods*. 2020;**17**:629–35.

68 Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S, et al. Quantitative genetic-interaction mapping in mammalian cells. *Nat Methods*. 2013;**10**:432–7.

69 Michaut M, Bader GD. Multiple genetic interaction experiments provide complementary information useful for gene function prediction. *PLoS Comput Biol*. 2012;**8**:e1002559.

70 Norman TM, Horlbeck MA, Replogle JM, Ge AY, Xu A, Jost M, et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science*. 2019;**365**:786–93.

71 Gomes G-NW, Krzeminski M, Namini A, Martin EW, Mittag T, Head-Gordon T, et al. Conformational ensembles of an intrinsically disordered protein consistent with NMR, SAXS, and single-molecule FRET. *J Am Chem Soc*. 2020;**142**:15697–710.

72 Pietrek LM, Stelzl LS, Hummer G. Hierarchical ensembles of intrinsically disordered proteins at atomic resolution in molecular dynamics simulations. *J Chem Theory Comput*. 2020;**16**:725–37.

73 Crehuet R, Buigues PJ, Salvatella X, Lindorff-Larsen K. Bayesian-maximum-entropy reweighting of IDP ensembles based on NMR chemical shifts. *Entropy*. 2019;**21**:898.

74 Brookes DH, Head-Gordon T. Experimental inferential structure determination of ensembles for intrinsically disordered proteins. *J Am Chem Soc*. 2016;**138**:4530–8.

75 Fuxreiter M. Fuzziness in protein interactions—a historical perspective. *J Mol Biol*. 2018;**430**:2278–87.

76 Kim YC, Tang C, Clore GM, Hummer G. Replica exchange simulations of transient encounter complexes in protein–protein association. *Proc Natl Acad Sci USA*. 2008;**105**:12855–60.

77 Raveh B, Karp JM, Sparks S, Dutta K, Rout MP, Sali A, et al. Slide-and-exchange mechanism for rapid and selective transport through the nuclear pore complex. *Proc Natl Acad Sci USA*. 2016;**113**:E2489–97.

78 Tang C, Iwahara J, Marius Clore G. Visualization of transient encounter complexes in protein–protein association. *Nature*. 2006;**444**:383–6.

79 Volkov AN, Worrall JAR, Holtzmann E, Ubbink M. Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR. *Proc Natl Acad Sci USA*. 2006;**103**:18945–50.

80 Dannatt HRW, Felletti M, Jehle S, Wang Y, Emsley L, Dixon NE, et al. Weak and transient protein interactions determined by solid-state NMR. *Angew Chem Int Ed Engl*. 2016;**55**:6638–41.

81 Liu Z, Gong Z, Dong X, Tang C. Transient protein–protein interactions visualized by solution NMR. *Biochim Biophys Acta*. 2016;**1864**:115–22.

82 Kwon Y, Kaake RM, Echeverria I, Suarez M, Karimian Shamsabadi M, Stoneham C, et al. Structural basis of CD4 downregulation by HIV-1 Nef. *Nat Struct Mol Biol*. 2020;**27**:822–8.

83 Schulze-Gahmen U, Echeverria I, Stjepanovic G, Bai Y, Lu H, Schneidman-Duhovny D, et al. Insights into HIV-1 proviral transcription from integrative structure and dynamics of the Tat:AFF4:P-TEFb:TAR complex. *Elife*. 2016;**5**:e15910.

84 Kaake RM, Echeverria I, Kim SJ, Von Dollen J, Chesarino NM, Feng Y, et al. Characterization of a A3G-VifHIV-1-CRL5-CBFβ structure using a cross-linking mass spectrometry pipeline for integrative modeling of host-pathogen complexes. *Mol Cell Proteomics*. 2021;**20**:100132.

85 Pushker R, Mooney C, Davey NE, Jacqué J-M, Shields DC. Marked variability in the extent of protein disorder within and between viral families. *PLoS One*. 2013;**8**:e60724.

86 Xue B, Williams RW, Oldfield CJ, Goh GK-M, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept Lett*. 2010;**17**:932–51.

87 Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*. 2012;**30**:137–49.

88 Pan J, Peng X, Gao Y, Li Z, Lu X, Chen Y, et al. Genome-wide analysis of protein-protein interactions and involvement of viral proteins in SARS-CoV replication. *PLoS One*. 2008;**3**:e3299.

89 Lee S, Salwinski L, Zhang C, Chu D, Sampankanpanich C, Reyes NA, et al. An integrated approach to elucidate the intra-viral and viral-cellular protein interaction networks of a gamma-herpesvirus. *PLoS Pathog*. 2011;**7**:e1002297.

90 von Brunn A, Teepe C, Simpson JC, Pepperkok R, Friedel CC, Zimmer R, et al. Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PLoS One*. 2007;**2**:e459.

91 Fossum E, Friedel CC, Rajagopala SV, Titz B, Baiker A, Schmidt T, et al. Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog*. 2009;**5**:e1000570.

92 Hagen N, Bayer K, Rösch K, Schindler M. The intraviral protein interaction network of hepatitis C virus. *Mol Cell Proteomics*. 2014;**13**:1676–89.

93 Osterman A, Stellberger T, Gebhardt A, Kurz M, Friedel CC, Uetz P, et al. The hepatitis E virus intraviral interactome. *Sci Rep*. 2015;**5**:13872.

94 Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antiviral Res*. 2018;**149**:58–74.