Received: 14 September 2022 | Revised: 23 March 2023 | Accepted: 24 March 2023

DOI: 10.1002/pmic.202200096

Proteomics
Proteomics and Systems Biology

RESEARCH ARTICLE

Enabling cryo-EM density interpretation from yeast native cell extracts by proteomics data and AlphaFold structures

Christian Tüting[1,2,3] | **Lisa Schmidt**[1,2] | **Ioannis Skalidis**[1,2] | **Andrea Sinz**[4,5] | **Panagiotis L. Kastritis**[1,2,3,6]

[1]Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[2]Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[3]Biozentrum, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[4]Institute of Pharmacy, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[5]Center for Structural Mass Spectrometry, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany

[6]Institute of Chemical Biology, National Hellenic Research Foundation, Athens, Greece

Correspondence
Panagiotis L. Kastritis and Christian Tüting, Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Weinbergweg 22, 06120 Halle (Saale), Germany.
Email: panagiotis.kastritis@bct.uni-halle.de; christian.tueting@biochemtech.uni-halle.de

Funding information
Horizon Europe ERA Chair "hot4cryo", Grant/Award Number: 101086665; Federal Ministry for Education and Research, Grant/Award Numbers: 03Z22HN23, 03Z22HI2, 03COV04; European Regional Development Funds, Grant/Award Number: ZS/2016/04/78115; Deutsche Forschungsgemeinschaft, Grant/Award Numbers: RTG2467, 391498659

Abstract

In the cellular context, proteins participate in communities to perform their function. The detection and identification of these communities as well as in-community interactions has long been the subject of investigation, mainly through proteomics analysis with mass spectrometry. With the advent of cryogenic electron microscopy and the "resolution revolution," their visualization has recently been made possible, even in complex, native samples. The advances in both fields have resulted in the generation of large amounts of data, whose analysis requires advanced computation, often employing machine learning approaches to reach the desired outcome. In this work, we first performed a robust proteomics analysis of mass spectrometry (MS) data derived from a yeast native cell extract and used this information to identify protein communities and inter-protein interactions. Cryo-EM analysis of the cell extract provided a reconstruction of a biomolecule at medium resolution (~8 Å (FSC = 0.143)). Utilizing MS-derived proteomics data and systematic fitting of AlphaFold-predicted atomic models, this density was assigned to the 2.6 MDa complex of yeast fatty acid synthase. Our proposed workflow identifies protein complexes in native cell extracts from *Saccharomyces cerevisiae* by combining proteomics, cryo-EM, and AI-guided protein structure prediction.

Abbreviations: ACC1, acetyl-CoA-synthase; AI, artificial intelligence; API, application programming interface; CC, cross-correlation; CF-MS, co-fractionation mass spectrometry; CTF, contrast transfer function; cryo-EM, cryogenic electron microscopy; cryo-ET, cryogenic electron tomography; EM, electron microscopy; FAS, fatty acid synthase; FSC, Fourier shell correlation; GPU, graphics processing unit; KEGG, Kyoto encyclopedia of genes and genomes; LFQ, label-free quantification; MDa, Mega-Dalton; MS, mass spectrometry; MS/MS, tandem mass spectrometry; NCS, non-crystallography symmetry; PFK, phosphofruktokinase; SEC, size exclusion chromatography; SPA, single-particle analysis; XL-MS, cross-linking mass spectrometry.

KEYWORDS
AI-guided, computational analysis, cryo-EM, homogenates, protein structure prediction, structural proteomics

1 | INTRODUCTION

The cell displays an astounding heterogeneity, harboring diverse biomolecules at a wide range of concentrations. Proteins represent the largest group of cellular components, at 20%–30% (w/v), or 200–300 g/L [1]. Employing state-of-the art technology, protein structure analysis from cells is only feasible at low resolution by cross-linking mass-spectrometry (XL-MS) [2–4] or at higher resolution by in situ cryo-electron tomography (cryo-ET) [5]. However, these methods mainly analyze very abundant complexes such as ribosomes [6] or the nuclear pore complex [7]. To reduce complexity while enriching for less abundant biomolecules, the cell must be lysed and fractionated, using, for example, centrifugation to separate membranes and aggregates from soluble cell content. Such coarse fractionation may then be followed by chromatography, for example, size exclusion chromatography (SEC), to separate biomolecules by size while retaining their native assemblies. Efficient structural determination of such retrieved extracts has been demonstrated by the structural analysis of the MDa-sized pyruvate dehydrogenase complex from the thermophilic fungus *Chaetomium thermophilum* [8].

The coupling of SEC fractionation with mass spectrometry (MS), termed *co-fractionation mass spectrometry* (CF/MS) [9], provides detailed insights into higher-order protein complexes. This is because co-elution profiles of proteins might contain information about stable as well as transient interactions. Recently, CF/MS data analysis was empowered by artificial intelligence (AI) [10] and can lead to the identification of known and unknown protein complexes from red blood cells [11]. These protein complexes often assemble in larger functional units, termed protein communities [12, 13].

Functional units of metabolic complexes in particular, referred to as *metabolons* [14], show interaction with various binders, including scaffold proteins, membrane patches, nucleic acids, and others [15, 16]. Currently, only cryo-EM of fractionated cell extracts is able to visualize such complexity at relatively high resolution [8, 15, 17]. Usually, classical cryo-EM single-particle analysis (SPA) is employed to the acquired images from heterogeneous fractions; SPA is comprised of the following steps: (a) particle picking from acquired micrographs, (b) 2D classification of extracted single-particles, and (c) 3D reconstruction of a Coulomb potential map that can often reach atomic resolution in the case of purified and stable specimen [18]. The majority of these steps already employ AI to learn, identify and reconstruct these structures and are implemented in specialized user-friendly toolkits, that is, RELION [19], Xmipp/Scipion [20], or cryoSPARC [21] which are either freely available (RELION, Scipion) or accessible to academic users (cryoSPARC).

SPA of complex native cell extracts harbors several technical limitations compared to SPA of purified proteins. For near-atomic resolution, a smaller pixel size (e.g., 1.6 Å/px for ~3 Å at Nyquist frequency) is required during data collection [17] but for cell extracts a larger pixel size is preferable. This choice limits resolution but increases the number of total particles per micrograph, which is indeed favorable for low-abundant macromolecular complexes. One, if not the major, prerequisite for the reconstruction of a high-resolution map from cryo-EM data is a sufficiently large number of single particles [22]. If the particle shape is recognizable, either directly in the micrograph to be picked manually or during the 2D classification of, that is, blob-picked particles (unbiased circular picking based on contrast), the analysis can be streamlined accordingly [23]. Statistical occurrence of a particle might be correlated with the MS-derived protein abundance, guiding the identification as well as the 3D reconstruction of the selected protein complex [15]. Additionally, technological advances in cryo-EM [18] ensure that a low-abundant protein will be present in high enough copy numbers to allow for atomic resolution reconstruction—if the target particle signature can be identified. However, megadalton cryo-EM maps from flexible macromolecules derived from native cell extracts are often of medium resolution, and therefore, hard to interpret in the context of molecular models [24].

In this work, we present an automated workflow that incorporates MS-based protein identification and database knowledge integration via computational analysis to visualize protein communities. These proteomics data enable structural identification and model building of a cryo-EM map derived from native cell extracts utilizing AlphaFold2-predicted monomeric protein structures. The derived low-resolution structure of the yeast fatty-acid-synthase (FAS) is correctly identified, and the derived model is in agreement with previously published data (Figure 1A).

2 | MATERIALS AND METHODS

2.1 | Re-analyzed data

All MS and cryo-EM data have been previously reported in Schmidt *et al.*[25] MS data have been deposited in the PRIDE repository under the accession PXD034431. Cryo-EM data are available from the EMPIAR database under the accession EMPIAR-11069. The atomic models were downloaded from the AlphaFold2 database [26] under the accession UP000002311_559292_YEAST_v3.

2.2 | Network analysis

The proteinGroups.txt file was obtained from the MS data and contains all proteins identified together with the respective label-free

quantification (LFQ). Only proteins identified in at least 50% out of 12 experiments were considered, and their mean LFQ values were calculated. For every protein identified, the STRING-identifier [27] and the respective protein interaction network, containing both physical and functional interactions, were fetched using the API interface. Only binary interactions with stoichiometries less than 1:10, based on LFQ intensities, were considered. This is because LFQ intensities infer relative abundance of the protein species, and this is translated in the relative abundance in the cryo-EM micrographs. Complexes that have members with such difference in relative abundance are challenging to capture within micrographs from native cell extracts, for example, the E3 of the PDHc [8]. The protein interaction network plot was generated using the python module NetworkX [28]. Edges were colored from green to red according to their "exp. score" value, a parameter indicating the confidence of a physical interaction between two nodes, as described in Szklarczyk *et al.*[27]

Statement Significance

Progress in the analysis of heterogeneous biochemical samples, specifically in cryo-EM of native cell extracts, allows the identification and characterization of protein interactions at the structural level. Here, we propose a robust workflow that incorporates information from proteomics experiments to guide the identification of protein complexes and leverages AlphaFold to predict their structures. Our workflow forgoes the protein backbone tracing step and is able to characterize large protein complexes at medium resolution.

2.3 | Image processing

The cryo-EM micrographs were analyzed using cryoSPARC 3.3.2 [21]. Raw movies were motion corrected (using "patch motion correction (multi)") and CTF-estimated (using "Patch CTF estimation (multi)"). Particles were picked using the "blob picker" module with a minimum particle diameter of 150 Å, and a maximum particle diameter of 300 Å. Single-particle images were extracted with a box size of 180 px. Retrieved single-particles were iteratively 2D classified, each with 400 classes during 2D classification. Asymmetric (C1) ab initio reconstructions were done for clear 2D classes, and classes containing "junk" single-particles (e.g., ice contaminants or broken/damaged single-particles) were discarded after each iteration of 2D classification. For the dome-shaped map, later identified as the fatty acid synthase complex (FAS), D3 symmetry was applied during homogenous refinement after prediction of symmetry utilizing ChimeraX [29]. From the initial symmetrized reconstruction, 20 2D projections (using the "Create Template" module of cryoSPARC) were generated. Template-based particle picking then followed, with a particle diameter of 300 Å, a low-pass filter of templates and micrographs set to 25 Å, and a minimum particle separation set to a distance of 1.25 diameters. Picked particles were extracted with a box size of 210 px. After 2D classification and selection of clear classes, an asymmetric (C1) ab initio reconstruction was calculated, followed by a symmetrized (D3) homogenous refinement.

2.4 | Unambiguous fitting of AlphaFold2 models

The proteins identified by MS were sorted by their LFQ intensities, and the most abundant 150 proteins were selected. Proteins were annotated according to the UniProt database [30], and protein names containing the keyword "ribosome" were removed. Fitting of the AlphaFold2 models in the density was performed with a modified local installation of ChimeraX (version 1.4.dev202202240543; Table S1A).[29] Each model was globally fitted 10,000 times with ChimeraX by random placement within the map density and local placement optimization (Table S1B). Not each random placement results in an accepted solution. All fits were saved and ranked by atom engulfment, which is defined as the fraction of protein atoms overlapping with the electron density at a certain contour level. Taking into account the best fitting solution, an atom engulfment threshold of 0.8 was applied. Additionally, a coarse analysis was carried out to identify steric clashes, using a threshold of Cα-clashing >25% of the total number of Cα atoms in the monomer.

2.5 | Resolving of Cα-clashing and real-space refinement

For calculating Cα-clashes between two given protein structures (in PDB format), the (x,y,z) coordinates of the Cα-atoms were isolated in a NumPy array [31], the distance between each atom was calculated with SciPy [32], and a clash was identified when atoms resided at a calculated distance of less than 3.65 Å to the nearest atom (Table S1C). Acceptance threshold for homomultimers was set to a mean Cα-clash value of ≤10.

For real-space refinement, PHENIX (version 1.20.1-4487) [33] was used with activated NCS constraints (asymmetric unit: FAS α β heterodimer), a map resolution of 6.5 Å, and a single macro cycle.

2.6 | Local resolution estimation

Local resolution maps for the FAS final reconstruction were generated with the "Local Resolution Estimation" module in cryoSPARC v3.1.1 [21] and visualized with ChimeraX v1.4 [29].

2.7 | Code availability

The software employed in this study was written in python and is available upon request.

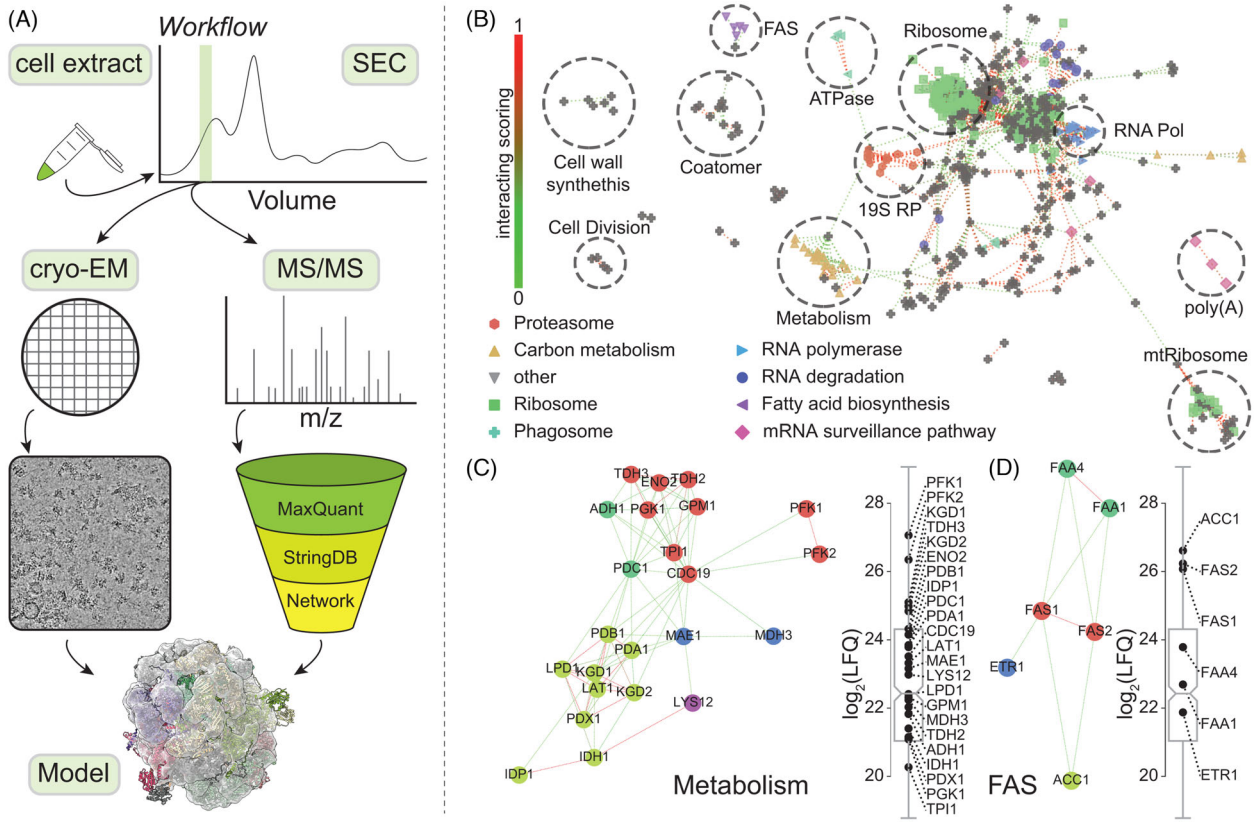

FIGURE 1 Modeling pipeline and network analysis of yeast native cell extract. (A) Modeling workflow—A native cell extract is fractionated by size exclusion chromatography (SEC) and high molecular weight (MW) fractions are analyzed by cryo-EM and mass spectrometry (MS/MS). The mass spectrometry results are processed using a pipeline that includes protein identification using MaxQuant, interaction analysis using the STRING database (StringDB), along with network annotation using STRING and KEGG information (Network). These results are combined with those from single particle analysis (SPA) of native cell extracts to produce a plausible atomic model within a medium resolution density map. (B) Network analysis—The network interactions are based on annotated interactions in STRING. Nodes are colored by KEGG pathways and edges are colored based on their direct interaction confidence derived from STRING (exp. score). Isolated groups are highlighted and named. (C) Metabolism group—The nodes are colored based on their metabolic pathway: glycolysis (red), citrate cycle (light green), fermentation (dark green), malate shuttling (blue), and other (purple). Edges are colored based on the direct interaction confidence (refer to B). All proteins identified in the sample are plotted as a boxplot in terms of relative abundance as measured by the label-free quantification (LFQ) score [35]. Members of the Metabolism group are highlighted as dots. (D) Fatty acid synthesis group—The nodes are colored based on their reactions: fatty acid synthase (red), acetyl-CoA synthesis (light green), long-fatty-acid ligases (dark green), and enoyl-ACP-reductase (blue). Edges are colored based on the exp. score (refer to B). The relative abundance, based on label-free quantification (LFQ) values reported by MaxQuant, of all proteins identified in the sample are plotted as a boxplot and members of the FAS group are highlighted as dots. The boxplot minima represent the 25th percentile, the maxima represent the 75th percentile, the notch indicates the data's median, and the whiskers extend to the minimum and maximum value within a 1.5 interquartile range.

3 | RESULTS

3.1 | High molecular weight protein complexes from native cell extracts—Identifying protein communities by combining MS and database knowledge

Two high-molecular weight fractions of yeast native cell extracts were previously analyzed by Schmidt et al. [25] to retrieve the endogenous L-A helper virus and various states of translating ribosomes [25]. Here, we re-analyzed the reported MS results as well as 12,795 cryo-EM micrographs that were acquired at a pixel size of 3.177 Å/pix to capture the cell extract content beyond these abundant biomolecules (Figure S1A).

A total of 585 unique proteins were identified by MS after combining the data derived from the two high-molecular weight fractions. For an unbiased identification of likely protein communities, the STRING database [27] was used. The STRING database integrates results for protein interactions and includes not only stable but also transient interactors, for example, allosteric regulators like kinases. STRING analysis resulted in 3304 binary interactions where only 59 out of 585 identified proteins were found to be singletons (meaning no interactor identified).

To further classify identified proteins, KEGG pathway analysis was performed [34]. In total, eight principal classes were recognized: (1) ribosomes, (2) proteasome, (3) carbon metabolism (including glycolysis, ethanol fermentation, and Krebs cycle among others), (4) fatty acid biosynthesis, (5) RNA polymerases, (6) RNA degradation, (7) phago-

some, and (8) mRNA surveillance (Table S2). These classes cover 40% of all identified proteins highlighting the high complexity of the sample.

The protein–protein interaction analysis, utilizing the standard STRING aggregated score, revealed a densely packed network with highly interconnected proteins. Nevertheless, an apparent grouping of proteins is visible (Figure 1B). A large and heterogenous group, referred to as Metabolism, includes proteins involved in cytosolic glycolysis, the malate shuttling mechanism, the mitochondrial pyruvate dehydrogenase complex, and the Krebs cycle (Figure 1C). Relative abundance values were estimated using the LFQ intensity score [35] (see Section 2), also previously used for deriving stoichiometric data for human protein–protein interactions (PPIs) [36]. Diverse abundance values (LFQ score) and low interaction values (exp. score [27]) indicate that the entire group probably does not form a stable metabolon but is composed of different subcomplexes. This is comparable with previously identified structures of co-eluting pyruvate and α-ketoglutarate dehydrogenase complexes [37, 38]. Judged by the exp. score, protein–protein interactions are visible for mitochondrial proteins but cytosolic proteins have lower exp. score, that is, a reduced probability of interaction. A notable exception is phosphofructokinase (PFK): Both its subunits (α, β) are present with very high abundance and direct interaction scores (Figure 1C).

Another interesting group includes proteins related to fatty acid synthase (FAS; Figure 1D). Notably, apart from the canonical α and β subunits of the FAS (FAS1, FAS2), the acetyl-CoA carboxylase (ACC1) is also identified, which is involved in the production of malonyl-CoA, a substrate for FAS, along with two ligases and one reductase. Correlating with the MS-derived protein abundances, the α (FAS2) and β (FAS1) subunits of FAS have the same copy number, indicating a 1:1 stoichiometry in the complex, perfectly corresponding to the known structure of fungal FAS [39], where they form an A_6B_6 complex. The carboxylase is present at a similar abundance but due to its megadalton size [40] might co-elute, while the ligases and reductase are of much lower abundance, indicating a transient interaction. This is also underpinned by pre-existing experimental evidence, that is, that only the α and β subunits of FAS form a stable complex, while the interaction of ACC1 and other carboxylases is of a transient nature [12, 41].

3.2 | Template-free cryo-EM reconstruction

For an unbiased picking of particles for cryo-EM, a blob picker module with a target particle size of 150–300 Å (Figure 2A,B) was used, corresponding to the approximate size of FAS and ribosomes, which represent the most abundant protein complexes in the sample. The picked particles were extracted with a box of 180 px (~570 Å), to capture not only the core proteins, but also potential additional binding partners. A total of 27,727,854 particles were picked and extracted (Figure 2B). Following particle extraction, the resulting large number of single-particle images have to be 2D classified and refined, but the number of 2D classes assigned during classification presents a significant limitation. Usually, a number of 50–200 2D classes is recommended by cryoSPARC, but based on available computing power, up to 400 classes are possible on modern workstations. Iterative 2D classifications were required to effectively identify clear signatures (Figure 2C). Very clear 2D classes appeared after multiple rounds of 2D classification (Figure 2D) but the total number of particles contained in each final class was low. To increase final particle numbers, ab initio 3D reconstructions utilizing the particles contained in these classes were generated and, based on these, 2D templates for template-based picking were created. With this approach, particle picks were considerably increased (Figure 2E). After the template-based picking, 6,923,214 particles could be identified, of which 7895 were eventually selected, again through iterative 2D classifications. A final reconstruction (D3 symmetry) resulted in a cryo-EM map of 8.0 Å resolution (FSC = 0.143) (Figures 2F and S1B). The cryo-EM density map shows an overall uniform resolution despite the missing views (Figure S1C), with additional localized lower resolution features (Figure S1D). An additional challenge posed by the current reconstruction is the limited coverage of particle views in the 2D class averages (Figure S1A). This effect of FAS particles has been previously observed [15] but due to the particles' shape and symmetry this does not pose a limitation to achieve sufficient resolution for further map analysis.

3.3 | AI-guided protein structure modeling – Identification and refinement of FAS

The dome-shaped reconstruction that was retrieved (Figure 2F) could not be built with current refinement tools; due to the resolution of ~8 Å, AI tools like FindMySequence [42], DeepTracer [43], or ModelAngelo [44] are not applicable. To address this issue, the 150 most abundant proteins were selected, ribosomal proteins were excluded, and modeling was focused on less abundant protein signatures. Ribosomes display distinct structural signatures in the raw micrograph data, 2D classes, and 3D reconstruction and can be easily distinguished from other, less abundant, protein signatures in native cell extract [25, 37]. The remaining 61 proteins were systematically fitted in the derived reconstruction (Figure 2F). To this end, all 61 protein structures were retrieved from the AlphaFold2 database and fitted 10,000 times each in the density to capture a variety of solutions (Figure 3A). The best-fitting solutions were isolated and grouped into predicted homomers. To assess the quality of these homomeric complexes, the number of Cα-clashes between subunits was calculated and ranked according to both average and total clash number. A high clash number indicates an ambiguous placement of a protein (Figure 3B). Only 18 homomers fulfill the criteria. To further statistically analyze these assemblies, the map coverage at various contour levels was calculated (Figure 3C). The majority of homomers explained only a minor part of the map (<10% map coverage at a contour level of 2.0), but two proteins stood out: FAS1 and FAS2, each explaining nearly half of the map.

Even though these two could be the correct hits, all heteromeric assemblies were generated and analyzed analogously to the predicted homomers: Only seven heteromeric assemblies fulfilled the low clash values criteria (maximum 2.5 % of Cα clashing; data not shown), and their map coverage was calculated (Figure 3D). Interestingly, the fitted heteromeric complex consisting of FAS1 and FAS2 was one of those

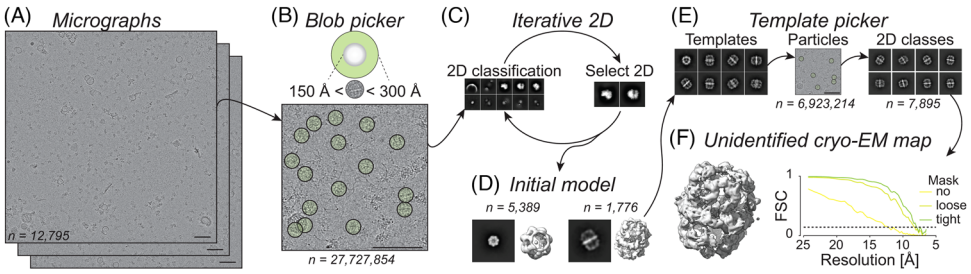

FIGURE 2 Template-free single-particle-analysis workflow. Subpanels (A–F) show the steps of the workflow that include: (A) Motion Correction and CTF Estimation—The 12,795 acquired micrographs underwent motion correction, and their CTF value was estimated. (B) Template-Free Particle Picking—A blob picker with a minimum diameter of 150 Å and a maximum diameter of 300 Å was applied to the micrographs, resulting in 27,727,854 picked and extracted particles. (C) 2D Classification and Class Average Selection—The picked particles underwent iterative rounds of 2D classification and class average selection (Select 2D) to generate high signal-to-noise-ratio 2D classes. (D) Ab Initio Map Reconstruction—Ab initio maps were reconstructed from noise-free 2D classes. 2D projections were generated and served as input for template picking (E). (E) Particle picking using 2D Projections—The 2D projections (templates) were used to pick particles with a similar appearance, resulting in 6,923,214 picked particles and 8 noise-free 2D classes containing a total of 7895 particles. (F) Cryo-EM Map Reconstruction—A cryo-EM map of 8.0 Å resolution (gold-standard FSC = 0.143, dashed line) was reconstructed after applying D3 symmetry.

hits. The complex was fully covered by the density at high contour levels unlike the other predicted heteromeric complexes (Figure 3D).

The unambiguously fitted and AI-generated monomeric protein structures resulted in a dodecameric assembly of FAS1 and FAS2 (each in six copies) (Figure 3). This was further cross-validated by the initial network analysis, where a direct interaction between these two proteins was visible (Figure 1D). To reduce clashes or close contacts between the fitted monomers, a simple real-space-refinement with default values was performed (Figure 3E) as it also optimizes clashes within the interfaces. Without any manual interference during model identification and model building, the protein-protein interactions in the FAS complex were effectively recapitulated. The central homo-hexameric α-helical interaction of FAS1 and the dimeric interaction of FAS1 and FAS2 are clearly visible and captured in both the reconstructed map and the atomic model.

4 | DISCUSSION

Modern approaches in structural biology are able to identify and structurally characterize protein communities in a near-native state [15, 24]. Due to the nature of the native environment, the samples are highly heterogenous, and multiple disciplines must be combined to eventually decipher the structural information. These include, among various others, different EM techniques, including tomography [45, 46], SPA [8, 17, 37, 47], single cell lysis during sample preparation for cryo-EM [48], and MS [49].

While MS is precise and can identify proteins even at very low abundances, these data must be set in the context of protein communities. Here, we demonstrate how network visualization and grouping—based on relative protein abundances, KEGG and STRING information—can identify these communities in a single chromatographic fraction, even without the inclusion of elution profile information. In the future, our protocol will be able to incorporate more data, for example, MS results from multiple consecutive fractions. Such data could act as input to identify more protein communities and differentiate these from random co-elution using advanced techniques like deep learning algorithms for more precise data interpretation [11, 50]. Using cryo-EM to analyze these highly heterogenous samples adds another layer of knowledge, but also increases complexity. To tackle this, we used MS data, AlphaFold2-predicted protein structures, and were successfully able to unambiguously recapitulate the FAS complex. This demonstrates that, even without reaching near-atomic resolution (<3.8 Å) during map reconstruction, it is feasible to retrieve models of protein complexes in a robust manner. This result complements our recent work [37] where we showed unambiguous identification of a protein complex at ~4.5 Å resolution.

In terms of model quality, AlphaFold2-generated structures are most often of high quality in their ordered regions (provided a high-quality multiple sequence alignment is available)—and usually these are well-resolved areas even in medium resolution cryo-EM density maps retrieved from cell extracts. With systematic fitting of the AI-predicted models, all possible accurate fits can be identified while false-positives can be removed through classification, scoring, and validation. Based solely on MS data we successfully identified the correct dodecameric complex consisting of FAS1 and FAS2. The network plot itself is a powerful tool to cross-validate the derived result. Using this information too early in our proposed workflow might lead to misguided preconceptions if, for example, a completely novel interaction is actually present in the cryo-EM map. Eliminating the pre-filtering of possible interactions is a clearly favorable approach to reduce an eventual bias to a minimum. Further optimization can be performed by improving interface energetics, for example, computed using tools such as HADDOCK [51].

The proposed workflow is generally applicable to any given cryo-EM map if a set of potential target proteins is available, as medium resolution (~8 Å) maps hold enough information to accurately fit AlphaFold models. Other approaches require the knowledge of the proteins that form the target assembly, for example, HADDOCK [52, 53], EMBuild [54], or CoEVxIMP [55, 56], or higher resolution in the final cryo-EM

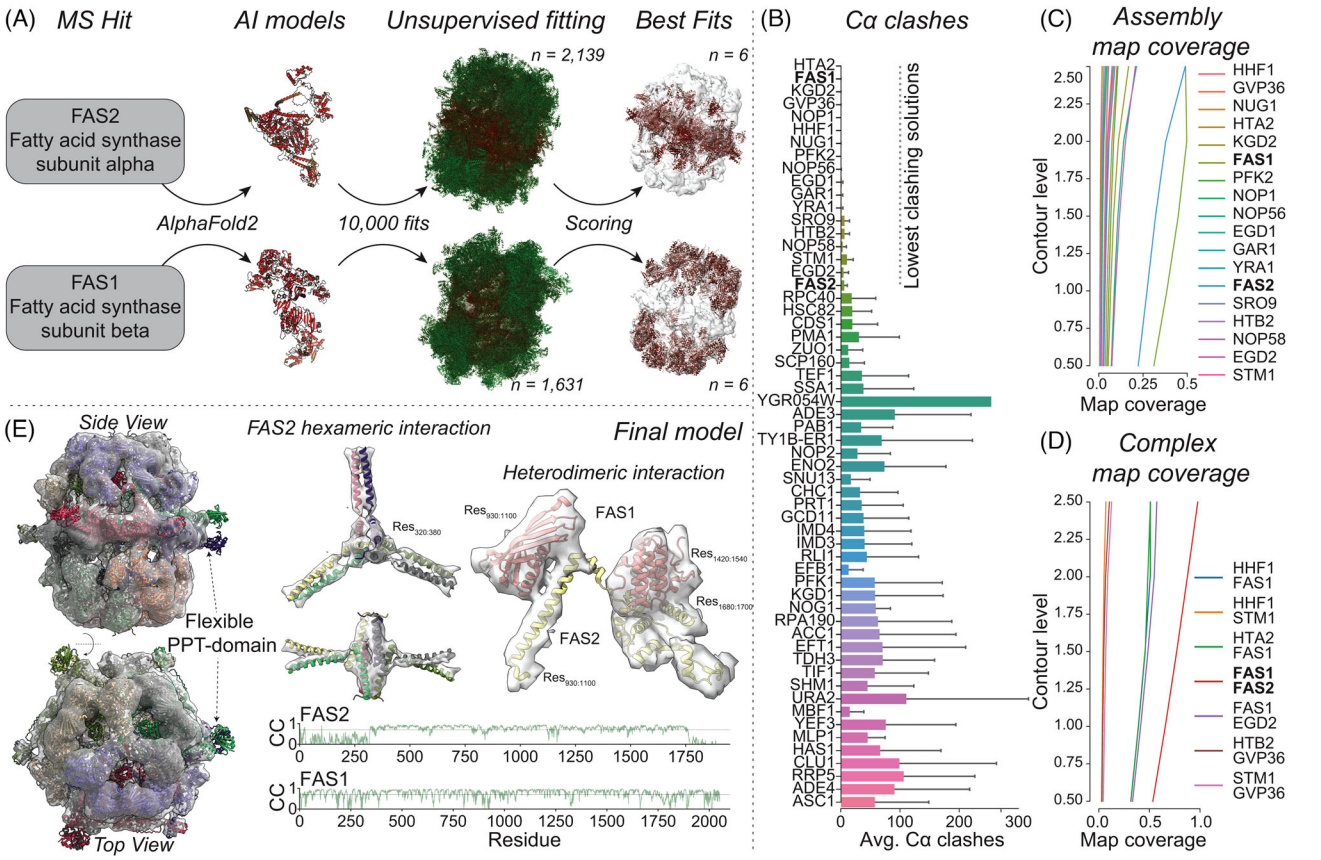

FIGURE 3 Model building based on AI-generated structures. (A) Monomeric model fitting—The figure shows an example of the fitting process for both FAS subunits. A selection of the 150 most abundant proteins from mass spectrometry results was made, excluding ribosomal proteins. From the remaining 61 proteins, the AI-predicted structure was retrieved from the AlphaFold database and fitted 10,000 times in the unidentified cryo-EM map shown in Figure 2. Only fits with high model-map correlation scores were retained, and the high-ranked monomeric fits were combined into a multimeric assembly. (B) Cα clash calculation—For each multimeric assembly, the clashes between each subunit's Cα were calculated. The mean values are shown as bar plots with error bars representing the standard deviation. (C) Map coverage—The coverage of the map by the low-clashing solutions (average clash <25) was calculated at various contour levels. FAS1 and FAS2 each cover approximately 50% of the map at a contour level of 2.0. (D) Complex map coverage—The coverage of the complex map was calculated for all non-Cα-clashing solutions after combining all 18 selected proteins. Only the combination of FAS1 and FAS2 forms a non-clashing solution, resulting in 100% coverage at a contour level of 2.5. (E) Final complex structure—The final complex structure was refined in the previously unidentified map. The dodecamer, comprised of six copies of FAS1 and FAS2, explains the unknown map visually and statistically. Despite the low resolution, the central hexamerization wheel of FAS2 and the binary interaction between both subunits are captured and covered by the atomic model. Both models exhibit a high map-to-model cross-correlation (CC). However, the N-terminal acyl carrier domain and the C-terminal transferase domain are not covered by the map density due to their known high flexibility and low local resolution of the cryo-EM map.

maps (<6 Å) with clear secondary structure separation, for example, Pathwalker/ROSETTA [57, 58]. Additionally, they require information about the sequence and structure. Additionally, such algorithms require flexible refinement procedures—therefore, docking hundreds of proteins and their pairs within the map is computationally challenging. Our workflow does not have these limitations, as the sequence information is directly derived from experimental MS data and the rigid fitting is computationally less demanding. A lower limit in protein size should be set, as for example, a 25 kDa protein can most likely be fitted in various rotations in a given low-resolution map. Additionally, an overall domain shape must be recognizable, setting an upper resolution limit to approximately 15 Å; however, these estimates can be quantitatively approached in the future by further benchmarking the proposed workflow. Moreover, a derived, highly heterogeneous, macromolecular assembly, which would include not just protein density (for example, ribosomes which consist of rRNA for a major part), might also negatively influence fitting results. Identifying and modeling of ribonucleoproteins in cryo-EM maps is nevertheless possible with other tools, such as DRRAFTER [59]. It should be noted that FAS, even though in the MDa molecular weight range, is highly symmetric allowing an identification based on only two polypeptide chains. More complex cryo-EM density maps might require adaptations in the fitting and scoring procedures.

Another major limitation of our workflow is how the parameters are set during particle picking. Blob-picked particles must fulfill a certain diameter threshold to be actually selected. Variability in diameter, which is actually feasible to perform during particle picking, can greatly influence chemical and structural heterogeneity of retrieved particles.

Also, a large majority of picked particles can either be noise or damaged [60], and must be separated in lengthy, manually curated 2D classifications. Some 2D class averages might never be retrieved due to particle heterogeneity: phosphofructokinase was never observed in the data but the molecule should have been seen as its diameter is in the range of blob sizes set during picking.

Future developments in cryo-EM should include "on-the-fly" analysis combined with advanced AI-assisted algorithms that cover all steps in a cryo-EM pipeline (from image analysis to model building). Single solutions for individual steps exist (i.e., DeepCryoPicker [61], DeepPicker [62], phenix autobuild [63], or others [64, 65]) but all these require specialized knowledge which limits usability for the majority of scientists. Future developments should aim at making these systems available and easy-to-use (simple point-and-click). These systems can include generalizable neural networks that eventually discriminate noise, contamination and signal, and ultimately provide an output of a molecular structure.

AUTHOR CONTRIBUTIONS

Christian Tüting designed the model and the computational framework and analyzed all the data. Christian Tüting, Lisa Schmidt, and Ioannis Skalidis wrote the submitted manuscript with the input of all authors. Christian Tüting and Panagiotis L. Kastritis conceived the original idea. Panagiotis L. Kastritis and Andrea Sinz revised the manuscript and Panagiotis L. Kastritissecured funding.

ACKNOWLEDGMENTS

The authors thank the members of the Kastritis laboratory, especially Fotis Kyrilis and Farzad Hamdi for cryo-EM sample preparation and data collection, and Kevin "Safe-Space" Janson for valuable input. They also thank Christian Ihling for providing access to the Center of Structural Mass Spectrometry. This work was supported by the European Union through funding of the Horizon Europe ERA Chair "hot4cryo" project number 101086665 (to P.L.K.), the Federal Ministry for Education and Research (BMBF, ZIK program) (Grant nos. 03Z22HN23, 03Z22HI2 and 03COV04 to P.L.K.), the European Regional Development Funds for Saxony-Anhalt (grant number EFRE: ZS/2016/04/78115 to P.L.K.), funding by Deutsche Forschungsgemeinschaft (DFG) (RTG2467, project number 391498659), and the Martin-Luther University of Halle-Wittenberg.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All raw data, used for image preparation, is attached (Supplementary File 1).

ORCID

Christian Tüting https://orcid.org/0000-0001-6209-4012
Lisa Schmidt https://orcid.org/0000-0002-1445-8149
Ioannis Skalidis https://orcid.org/0000-0002-3077-9014
Andrea Sinz https://orcid.org/0000-0003-1521-4899
Panagiotis L. Kastritis https://orcid.org/0000-0002-1463-8422

REFERENCES

1. Brown, G. C. (1991). Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *Journal of Theoretical Biology*, *153*(2), 195–203. https://doi.org/10.1016/s0022-5193(05)80422-9
2. Götze, M., Iacobucci, C., Ihling, C. H., & Sinz, A. (2019). A simple cross-linking/mass spectrometry workflow for studying system-wide protein interactions. *Analytical Chemistry*, *91*(15), 10236–10244. https://doi.org/10.1021/acs.analchem.9b02372
3. O'reilly, F. J., & Rappsilber, J. (2018). Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nature Structural & Molecular Biology*, *25*(11), 1000–1008. https://doi.org/10.1038/s41594-018-0147-0
4. Piersimoni, L., Kastritis, P. L., Arlt, C., & Sinz, A. (2022). Cross-linking mass spectrometry for investigating protein conformations and protein–protein interactions—A method for all seasons. *Chem. Rev.*, *122*(8), 7500–7531. https://doi.org/10.1021/acs.chemrev.1c00786
5. Turk, M., & Baumeister, W. (2020). The promise and the challenges of cryo-electron tomography. *Febs Letters*, *594*(20), 3243–3261. https://doi.org/10.1002/1873-3468.13948
6. O'reilly, F. J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., Hagen, W. J. H., Cramer, P., Stülke, J., Mahamid, J., & Rappsilber, J. (2020). In-cell architecture of an actively transcribing-translating expressome. *Science*, *369*(6503), 554–557. https://doi.org/10.1126/science.abb3758
7. Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turonová, B., Zimmerli, C E., Buczak, K., Schmidt, F. H., Margiotta, E., Mackmull, M.-T., Hagen, W. J. H., Hummer, G., Kosinski, J., & Beck, M. (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science*, *376*(6598), eabm9506. https://doi.org/10.1126/science.abm9506
8. Kyrilis, F. L., Semchonok, D. A., Skalidis, I., Tüting, C., Hamdi, F., O'reilly, F. J., Rappsilber, J., & Kastritis, P. L. (2021). Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. *Cell Reports*, *34*(6), 108727. https://doi.org/10.1016/j.celrep.2021.108727
9. Havugimana, P. C., Goel, R. K., Phanse, S., Youssef, A., Padhorny, D., Kotelnikov, S., Kozakov, D., & Emili, A. (2022). Scalable multiplex co-fractionation/mass spectrometry platform for accelerated protein interactome discovery. *Nature Communications*, *13*(1), 4043. https://doi.org/10.1038/s41467-022-31809-z
10. Fossati, A., Li, C., Uliana, F., Wendt, F., Frommelt, F., Sykacek, P., Heusel, M., Hallal, M., Bludau, I., Capraz, T., Xue, P., Song, J., Wollscheid, B., Purcell, A. W., Gstaiger, M., & Aebersold, R. (2021). PCprophet: A framework for protein complex prediction and differential analysis using proteomic data. *Nature Methods*, *18*(5), 520–527. https://doi.org/10.1038/s41592-021-01107-5
11. Sae-Lee, W., Mccafferty, C. L., Verbeke, E. J., Havugimana, P. C., Papoulas, O., Mcwhite, C. D., Houser, J. R., Vanuytsel, K., Murphy, G. J., Drew, K., Emili, A., Taylor, D. W., & Marcotte, E. M. (2022). The protein organization of a red blood cell. *Cell Reports*, *40*(3), 111103. https://doi.org/10.1016/j.celrep.2022.111103
12. Gavin, A.-C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dümpelfeld, B., Edelmann, A., Heurtier, M.-A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.-M., Schelder, M., Schirle, M., … Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, *440*(7084), 631–636. https://doi.org/10.1038/nature04532
13. Gavin, A.-C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A.-M., Cruciat, C.-M., Remor, M., Höfert, C., Schelder, M., Brajenovic, M., Ruffner, H., Merino, A., Klein, K., Hudak, M., Dickson, D., … Superti-Furga, G. (2002). Functional

organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, *415*(6868), 141–147. https://doi.org/10.1038/415141a
14. Srere, P. A. (1985). The Metabolon. *Trends in Biochemical Sciences*, *10*(3), 109–110. https://doi.org/10.1016/0968-0004(85)90266-X
15. Kastritis, P. L., O'reilly, F. J., Bock, T., Li, Y., Rogon, M. Z., Buczak, K., Romanov, N., Betts, M. J., Bui, K. H., Hagen, W. J., Hennrich, M. L., Mackmull, M.-T., Rappsilber, J., Russell, R. B., Bork, P., Beck, M., & Gavin, A.-C. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Molecular Systems Biology*, *13*(7), 936. https://doi.org/10.15252/msb.20167412
16. Kastritis, P. L., & Gavin, A.-C. (2018). Enzymatic complexes across scales. *Essays in Biochemistry*, *62*(4), 501–514. https://doi.org/10.1042/EBC20180008
17. Tüting, C., Kyrilis, F. L., Müller, J., Sorokina, M., Skalidis, I., Hamdi, F., Sadian, Y., & Kastritis, P. L. (2021). Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nature Communications*, *12*(1), 6933. https://doi.org/10.1038/s41467-021-27287-4
18. Chua, E. Y. D., Mendez, J. H., Rapp, M., Ilca, S. L., Tan, Y. Z., Maruthi, K., Kuang, H., Zimanyi, C. M., Cheng, A., Eng, E. T., Noble, A. J., Potter, C. S., & Carragher, B. (2022). Better, faster, cheaper: Recent advances in cryo–electron microscopy. *Annual Review of Biochemistry*, *91*, 1–32. https://doi.org/10.1146/annurev-biochem-032620-110705
19. Kimanius, D., Dong, L., Sharov, G., Nakane, T., & Scheres, S. H. W. (2021). New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal*, *478*(24), 4169–4185. https://doi.org/10.1042/BCJ20210708
20. Strelak, D., Jiménez-Moreno, A., Vilas, J. L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., De Isidro-Gómez, F. P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y. C., Jiménez, J., Martínez, M., Harastani, M., Jonic, S., … Sorzano, C. O. S. (2021). Advances in Xmipp for cryo–electron microscopy: From Xmipp to Scipion. *Molecules (Basel, Switzerland)*, *26*(20), 6224. https://doi.org/10.3390/molecules26206224
21. Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, *14*(3), 290–296. https://doi.org/10.1038/nmeth.4169
22. Henderson, R. (2004). Realizing the potential of electron cryo-microscopy. *Quarterly Reviews of Biophysics*, *37*(1), 3–13. https://doi.org/10.1017/s0033583504003920
23. Sigworth, F. J. (2016). Principles of cryo-EM single-particle image processing. *Microscopy (Oxford)*, *65*(1), 57–67. https://doi.org/10.1093/jmicro/dfv370
24. Kyrilis, F. L., Belapure, J., & Kastritis, P. L. (2021). Detecting protein communities in native cell extracts by machine learning: A structural biologist's perspective. *Front. Mol. Biosci.*, *8*, 660542. https://doi.org/10.3389/fmolb.2021.660542
25. Schmidt, L., Tüting, C., Kyrilis, F. L., Hamdi, F., Semchonok, D. A., Hause, G., Meister, A., Ihling, C., Shah, P. N. M., Stubbs, M. T., Sinz, A., Stuart, D. I., & Kastritis, P. L. (2022). Delineating organizational principles of the endogenous L-A virus by cryo-EM and computational analysis of native cell extracts. *BioRxiv*, https://doi.org/10.1101/2022.07.15.498668
26. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2
27. Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & Von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074
28. Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). *Exploring network structure, dynamics, and function using NetworkX*. Paper presented at the Proceedings of the 7th Python in Science Conference, Pasadena, CA USA. http://conference.scipy.org/proceedings/SciPy2008/paper_2/
29. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943
30. Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Da Silva, A., Denny, P., Dogan, T., Ebenezer, T., Fan, J., Castro, L. G., … Teodoro, D. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100
31. Harris, C. R., Millman, K. J, Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2
32. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J, Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2
33. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., Mccoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, *75*(Pt 10), 861–877. https://doi.org/10.1107/S2059798319011471
34. Kanehisa, M. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27
35. Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., & Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Molecular & Cellular Proteomics*, *13*(9), 2513–2526. https://doi.org/10.1074/mcp.M113.031591
36. Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A., Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A., & Mann, M. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell*, *163*(3), 712–723. https://doi.org/10.1016/j.cell.2015.09.053
37. Skalidis, I., Kyrilis, F. L., Tüting, C., Hamdi, F., Chojnowski, G., & Kastritis, P. L. (2022). Cryo-EM and artificial intelligence visualize endogenous protein community members. *Structure (London, England)*, *30*(4), 575–589.e6 e576. https://doi.org/10.1016/j.str.2022.01.001
38. Kyrilis, F. L., Meister, A., & Kastritis, P. L. (2019). Integrative biology of native cell extracts: A new era for structural characterization of life processes. *Biological Chemistry*, *400*(7), 831–846. https://doi.org/10.1515/hsz-2018-0445
39. Lomakin, I. B., Xiong, Y., & Steitz, T. A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell*, *129*(2), 319–332. https://doi.org/10.1016/j.cell.2007.03.013
40. Wei, J., & Tong, L. (2015). Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer. *Nature*, *526*(7575), 723–727. https://doi.org/10.1038/nature15375

41. Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S.-L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., … Tyers, M. (2002). Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, *415*(6868), 180–183. https://doi.org/10.1038/415180a
42. Chojnowski, G., Simpkin, A. J., Leonardo, D. A., Seifert-Davila, W., Vivas-Ruiz, D. E., Keegan, R. M., & Rigden, D. J. (2022). *findMySequence*: A neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. *IUCrJ*, *9*(Pt 1), 86–97. https://doi.org/10.1107/S2052252521011088
43. Pfab, J., Phan, N. M., & Si, D. (2021). DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(2), e2017525118. https://doi.org/10.1073/pnas.2017525118
44. Jamali, K., Kimanius, D., & Scheres, S. (2022). ModelAngelo: Automated model building in cryo-EM maps. *arXiv preprint*, https://doi.org/10.48550/arXiv.2210.00006
45. Beck, M., & Baumeister, W. (2016). Cryo-electron tomography: Can it reveal the molecular sociology of cells in atomic detail? *Trends in Cell Biology*, *26*(11), 825–837. https://doi.org/10.1016/j.tcb.2016.08.006
46. Erdmann, P. S., Hou, Z., Klumpe, S., Khavnekar, S., Beck, F., Wilfling, F., Plitzko, J. M., & Baumeister, W. (2021). In situ cryo-electron tomography reveals gradient organization of ribosome biogenesis in intact nucleoli. *Nature Communications*, *12*(1), 5364. https://doi.org/10.1038/s41467-021-25413-w
47. Verbeke, E. J., Mallam, A. L., Drew, K., Marcotte, E. M., & Taylor, D. W. (2018). Classification of single particles from human cell extract reveals distinct structures human cell extract reveals distinct structures. *Cell Reports*, *24*(1), 259–268.e3 e253. https://doi.org/10.1016/j.celrep.2018.06.022
48. Kemmerling, S., Arnold, S. A., Bircher, B. A., Sauter, N., Escobedo, C., Dernick, G., Hierlemann, A., Stahlberg, H., & Braun, T. (2013). Single-cell lysis for visual analysis by electron microscopy. *Journal of Structural Biology*, *183*(3), 467–473. https://doi.org/10.1016/j.jsb.2013.06.012
49. Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.-D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., Naimy, S., Rahbek-Gjerdrum, L. M., Dyring-Andersen, B., Bulkescher, J., Lukas, C., Eckert, M. A., Lengyel, E., Gnann, C., Lundberg, E., … Mann, M. (2022). Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nature Biotechnology*, *40*(8), 1231–1240. https://doi.org/10.1038/s41587-022-01302-5
50. Meyer, J. G. (2021). Deep learning neural network tools for proteomics. *Cell Rep Methods*, *1*(2), 100003. https://doi.org/10.1016/j.crmeth.2021.100003
51. Neijenhuis, T., Van Keulen, S. C., & Bonvin, A. M. J. J. (2022). Interface refinement of low- to medium-resolution cryo-EM complexes using HADDOCK2.4. *Structure (London, England)*, *30*(4), 476–484.e3 e473. https://doi.org/10.1016/j.str.2022.02.001
52. Van Zundert, G. C. P., Melquiond, A. S. J., & Bonvin, A. M. J. J. (2015). Integrative modeling of biomolecular complexes: HADDOCKing with Cryo-electron microscopy data. *Structure (London, England)*, *23*(5), 949–960. https://doi.org/10.1016/j.str.2015.03.014
53. Trellet, M., van Zundert, G., & Bonvin, A. (2020). Protein-protein modeling using cryo-EM restraints. *Methods in Molecular Biology*, *2112*, 145–162. https://doi.org/10.1007/978-1-0716-0270-6_11
54. He, J., Lin, P., Chen, J., Cao, H., & Huang, S.-Y. (2022). Model building of protein complexes from intermediate-resolution cryo-EM maps with deep learning-guided automatic assembly. *Nature Communications*, *13*(1), 4066. https://doi.org/10.1038/s41467-022-31748-9
55. Mccafferty, C. L., Taylor, D. W., & Marcotte, E. M. (2021). Improving integrative 3D modeling into low- to medium-resolution electron microscopy structures with evolutionary couplings. *Protein Science*, *30*(5), 1006–1021. https://doi.org/10.1002/pro.4067
56. Webb, B., Viswanath, S., Bonomi, M., Pellarin, R., Greenberg, C. H., Saltzberg, D., & Sali, A. (2018). Integrative structure modeling with the Integrative Modeling Platform. *Protein Science*, *27*(1), 245–258. https://doi.org/10.1002/pro.3311
57. Chen, M., Baldwin, P. R., Ludtke, S. J., & Baker, M. L. (2016). De Novo modeling in cryo-EM density maps with Pathwalking. *Journal of Structural Biology*, *196*(3), 289–298. https://doi.org/10.1016/j.jsb.2016.06.004
58. Dimaio, F., Tyka, M. D., Baker, M. L., Chiu, W., & Baker, D. (2009). Refinement of protein structures into low-resolution density maps using rosetta. *Journal of Molecular Biology*, *392*(1), 181–190. https://doi.org/10.1016/j.jmb.2009.07.008
59. Kappel, K., Liu, S., Larsen, K. P., Skiniotis, G., Puglisi, E. V., Puglisi, J. D., Zhou, Z. H, Zhao, R., & Das, R. (2018). De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature Methods*, *15*(11), 947–954. https://doi.org/10.1038/s41592-018-0172-2
60. D'imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M., & Kühlbrandt, W. (2019). Protein denaturation at the air-water interface and how to prevent it. *Elife*, *8*, e42747. https://doi.org/10.7554/eLife.42747
61. Al-Azzawi, A., Ouadou, A., Max, H., Duan, Y., Tanner, J. J., & Cheng, J. (2020). DeepCryoPicker: Fully automated deep neural network for single protein particle picking in cryo-EM. *BMC Bioinformatics [Electronic Resource]*, *21*(1), 509. https://doi.org/10.1186/s12859-020-03809-7
62. Wang, F., Gong, H., Liu, G., Li, M., Yan, C., Xia, T., Li, X., & Zeng, J. (2016). DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *Journal of Structural Biology*, *195*(3), 325–336. https://doi.org/10.1016/j.jsb.2016.07.006
63. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W, Read, R. J., & Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D, Biological Crystallography*, *64*(Pt 1), 61–69. https://doi.org/10.1107/S090744490705024X
64. Zhu, Y., Ouyang, Q., & Mao, Y. (2017). A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics [Electronic Resource]*, *18*(1), 348. https://doi.org/10.1186/s12859-017-1757-y
65. Sorzano, C. O. S., Recarte, E., Alcorlo, M., Bilbao-Castro, J. R., San-Martín, C., Marabini, R., & Carazo, J. M. (2009). Automatic particle selection from electron micrographs using machine learning techniques. *Journal of Structural Biology*, *167*(3), 252–260. https://doi.org/10.1016/j.jsb.2009.06.011

SUPPORTING INFORMATION

Additional supporting information may be found online https://doi.org/10.1002/pmic.202200096 in the Supporting Information section at the end of the article.

How to cite this article: Tüting, C., Schmidt, L., Skalidis, I., Sinz, A., & Kastritis, P. L. (2023). Enabling cryo-EM density interpretation from yeast native cell extracts by proteomics data and AlphaFold structures. *Proteomics*, *23*, e2200096. https://doi.org/10.1002/pmic.202200096