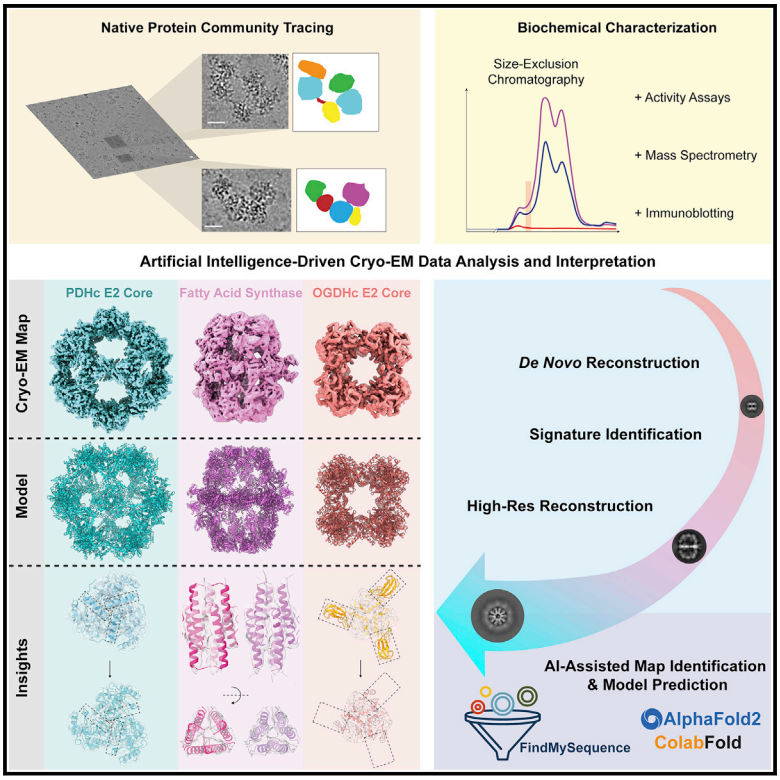
Structure

Cryo-EM and artificial intelligence visualize endogenous protein community members

Graphical abstract



Highlights

- Rapid high-resolution cryo-EM of cell extracts captures four active metabolons
- Simultaneous volume-based *de novo* identification of protein community members
- Multiple *ab initio* reconstructed, high-resolution cryo-EM maps from a single specimen
- AI-assisted atomic modeling reveals structural adaptations of captured native states

Authors

Ioannis Skalidis, Fotis L. Kyrilis, Christian Tüting, Farzad Hamdi, Grzegorz Chojnowski, Panagiotis L. Kastritis

Correspondence

panagiotis.kastritis@bct.uni-halle.de

In brief

Skalidis et al. present an AI-assisted, cryo-EM image analysis, model identification, and reconstruction pipeline. After the integration of traditional biochemical sample characterization and external database information, they obtain structural insights from multiple, simultaneously resolved atomic models of megadalton-range protein community members, directly derived from eukaryotic native cell extracts.

Skalidis et al., 2022, Structure *30*, 575–589
April 7, 2022
https://doi.org/10.1016/j.str.2022.01.001

CellPress

Structure

CellPress
OPEN ACCESS

Article

Cryo-EM and artificial intelligence visualize endogenous protein community members

Ioannis Skalidis,[1,2] Fotis L. Kyrilis,[1,2] Christian Tüting,[1] Farzad Hamdi,[1] Grzegorz Chojnowski,[3] and Panagiotis L. Kastritis[1,2,4,5,*]

[1]Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3a, 06120 Halle/Saale, Germany
[2]Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3, 06120 Halle/Saale, Germany
[3]European Molecular Biology Laboratory, Hamburg Unit, Notkestrasse 85, 22607 Hamburg, Germany
[4]Biozentrum, Martin Luther University Halle-Wittenberg, Weinbergweg 22, 06120 Halle/Saale, Germany
[5]Lead contact
*Correspondence: panagiotis.kastritis@bct.uni-halle.de
https://doi.org/10.1016/j.str.2022.01.001

SUMMARY

Cellular function is underlined by megadalton assemblies organizing in proximity, forming communities. Metabolons are protein communities involving metabolic pathways such as protein, fatty acid, and thioesters of coenzyme-A synthesis. Metabolons are highly heterogeneous due to their function, making their analysis particularly challenging. Here, we simultaneously characterize metabolon-embedded architectures of a 60S pre-ribosome, fatty acid synthase, and pyruvate/oxoglutarate dehydrogenase complex E2 cores *de novo*. Cryo-electron microscopy (cryo-EM) 3D reconstructions are resolved at 3.84–4.52 Å resolution by collecting <3,000 micrographs of a single cellular fraction. After combining cryo-EM with artificial intelligence-based atomic modeling and *de novo* sequence identification methods, at this resolution range, polypeptide hydrogen bonding patterns are discernible. Residing molecular components resemble their purified counterparts from other eukaryotes but also exhibit substantial conformational variation with potential functional implications. Our results propose an integrated tool, boosted by machine learning, that opens doors for structural systems biology spearheaded by cryo-EM characterization of native cell extracts.

INTRODUCTION

The cell is a very complex environment. Multiple processes have to be simultaneously coordinated and carried out in order for the cell to survive, perform its functions, and propagate. To achieve this, some kind of order through this chaos must be imposed. A major way that this is achieved is through the organization of the multiple proteins that take part in a specific process into larger assemblies. These large assemblies, dubbed "protein communities" (Gavin et al., 2006; Kuhner et al., 2009; Kastritis et al., 2017), facilitate the process they take part in (whether this is a signaling or metabolic pathway) by bringing the main players of the process closer together, linking intermediate products, and generally "compartmentalizing" the process in an environment where physical barriers to separate them are sometimes absent. In particular, enzymatic pathways that form protein communities are also dubbed "metabolons"(Srere, 1985). Community/metabolon-embedded proteins acquire different conformations to achieve function (Kyrilis et al., 2021b), and they can sometimes vastly differ in composition, structure, and function compared with studying them in isolation (Tüting et al., 2021). This is in contrast to their overexpressed and highly purified counterparts that are always preferred for structural studies due to their purity, homogeneity, and high quantity. So, the goal of modern structural biology should be to complement *in vitro* unambiguous data with the study of protein communities to more deeply understand endogenous cellular structure and function.

Extracts that are derived from cell lysis may provide us with this information, given that the proteins included are subjected to minimal perturbation. Former low-resolution studies (Han et al., 2009; Maco et al., 2011) and recent studies with cryo-EM (Kastritis et al., 2017; Kyrilis et al., 2019) have shown that this is certainly feasible, providing us with results that enrich our structural knowledge, which often omitted this "near-native" context. The concept of using cell extracts, combined with techniques such as gradient fractionation and differential centrifugation (Leighton et al., 1968), as a tool for the exploration and identification of sub-cellular organelles (Palade and Siekevitz, 1956) and protein complexes (Beaufay et al., 1974) is well established and has led to seminal discoveries, as in the pioneering works of De Duve (De Duve and Berthet, 1954) and Palade (Ehrenreich et al., 1973). Nevertheless, it is to be expected that when dealing with a sample as complex as a native cell extract, containing a plethora of different structural signatures, several biological and technical issues arise. How do we separate and identify the different structural signatures within protein communities?

Is there adequate abundance of a signature to provide meaningful data that will lead to the sought-after results? What kind of visualization and general investigation techniques can be applied that will maintain throughout the whole experimental process the structural state of community members and be amenable to high-resolution analysis?

Recent advancements in cryo-EM can provide us with the necessary tools to tackle the aforementioned questions. The "resolution revolution" has brought about an era of unprecedented structural discovery (Kühlbrandt, 2014a, 2014b), especially for highly enriched, endogenous protein complexes e.g., the mammalian brain V-ATPase, tau filaments responsible for neurodegenerative disorders, or the native ribosome-Sec61-OST complex cryo-EM structures (Braunger et al., 2018; Abbas et al., 2020; Zhang et al., 2020). In recent years we have experienced a boom of structural analysis of proteins and systems that was previously unattainable. From advancements in sample preparation and vitrification (Passmore and Russo, 2016), to hardware, such as electron sources, EM lenses, and direct electron detectors (Nakane et al., 2020; Yip et al., 2020), to image processing methods (de la Rosa-Trevin et al., 2016; Punjani et al., 2017; Tegunov and Cramer, 2019; Zivanov et al., 2020), the complex and exciting world of increasingly sophisticated cellular structures is now within reach (Beck and Baumeister, 2016; Danev et al., 2019). Single-particle cryo-EM has the intrinsic advantage of studying biomolecular complexes in a frozen-hydrated state, i.e., embedding random orientations of biomolecules inside vitreous ice. This way, any alterations to the protein of interest are not necessarily required (e.g., truncations, mutations, or other types of general modifications). In addition, multiple cryo-EM reconstructions have shown that observing numerous, large-scale conformational variations among structural states is possible (Saibil, 2000), e.g., during elongation in protein translation (Behrmann et al., 2015) or breathing motions of purified fatty acid synthase (Ciccarelli et al., 2013). Disordered regions of proteins in particular can now be better understood in the context of the metabolons (Skalidis et al., 2020), and advances in cryo-EM image processing methods may even allow their direct visualization (Punjani and Fleet, 2021).

Of course, advancements in instrumentation and sample preparation have provided us with a daunting amount of data that have to be analyzed so we can reach the desired end result. It is of particular note that sample complexity in cryo-EM does not pose a strict barrier for structure calculation, as shown, e.g., by the recent high-resolution cryo-EM structures of ribosomes from inside the cell (O'Reilly et al., 2020). However, data analysis is increasingly complex and may act as a bottleneck for high-resolution cryo-EM map reconstruction, especially for heterogeneous specimens. Traditional and machine-learning-based image processing algorithms and pipelines have been equally developed to address such challenges of structural and chemical heterogeneity (de la Rosa-Trevin et al., 2016; Punjani et al., 2017; Tegunov and Cramer, 2019; Zivanov et al., 2020). Structural biologists now have access to an array of computational tools that are streamlining the structure determination process from cryo-EM data, tools applicable not just to purified protein samples but also tools that have been recently applied to increasingly complex protein mixtures (recently reviewed in Kyrilis et al., 2021a).

Recent advances in artificial intelligence (AI)-driven protein structure prediction (Baek et al., 2021; Jumper et al., 2021) have shaken the field of structural biology and have opened new avenues to be investigated, especially in the context of cell extract studies. Researchers can now access powerful and reliable model prediction tools, removing the previous restraints of high-resolution requirements for the identification and reconstruction of low-abundance proteins, a development that contemporary explorers of native cell extracts can fully utilize, in order to acquire insights into native protein structure and function.

Protein communities and their component members have not yet been resolved to near-atomic resolution, especially in a highly complex environment, such as that of a native cell extract. In this work, we endeavor to coalesce available advancements in cryo-EM sample preparation, data collection schemes, image analysis tools, and AI-driven atomic model prediction algorithms into a rapid, optimized workflow for the investigation of endogenous, community/metabolon-embedded proteins at unprecedented resolution. We combine biochemical assays, mass spectrometry data, high-throughput image analysis, and atomic model reconstructions to simultaneously identify, characterize, and structurally analyze *de novo* multiple active endogenous metabolons captured by cryo-EM. Ultimately, our methods devise a pipeline for structural characterization of fractionated extracts, which will be critical for bridging structural and cell biology studies.

RESULTS

Isolation and enrichment of multiple metabolons and in-extract activity assays in a complex mixture

We have previously devised (Kastritis et al., 2017; Kyrilis et al., 2021b) a simple, one-step size-exclusion chromatography (SEC) approach of *Chaetomium thermophilum* native cell extracts that we also applied in this study and further improved protein complex separation (Figure 1A). Peaks in the elution profile correspond to complexes >200 kDa, with the first peak corresponding to complexes of several MDa and spanning fractions 1–8. Initial cryo-EM imaging of a fraction eluting at the MDa regime showed formation of protein communities after denoising (Bepler et al., 2020) (Figures 1B and S1). Cryo-EM imaging of this cellular fraction denotes the presence of a heterogeneous mixture of endogenous communities, revealing intriguing organizational principles of the pyruvate dehydrogenase complex (PDHc) metabolon (Figure S1). We then analyzed previously collected mass spectrometry (MS) data (Kastritis et al., 2017) and observed overall an increased abundance of proteins and complexes involved in forming the metabolons of (1) the branched-chain ketoacid dehydrogenase complexes (BCKDHc), (2) the PDHc and (3) oxo-acid dehydrogenase complex (OGDHc) (4) the 60S ribosome, and (5) the fatty acid synthase (FAS). Proteins related to the assembly and function of these five complexes comprise 24.8% of the total calculated protein abundance (Figure 1C) in a total of 1,281 proteins that were identified within this single fraction. Immunoblotting data confirmed the presence of the E2 cores of oxo-acid dehydrogenases (E2 of PDHc (E2p) and OGDHc (E2o), Figure 1D). In-fraction spectrophotometric assays verify the retrieval of active

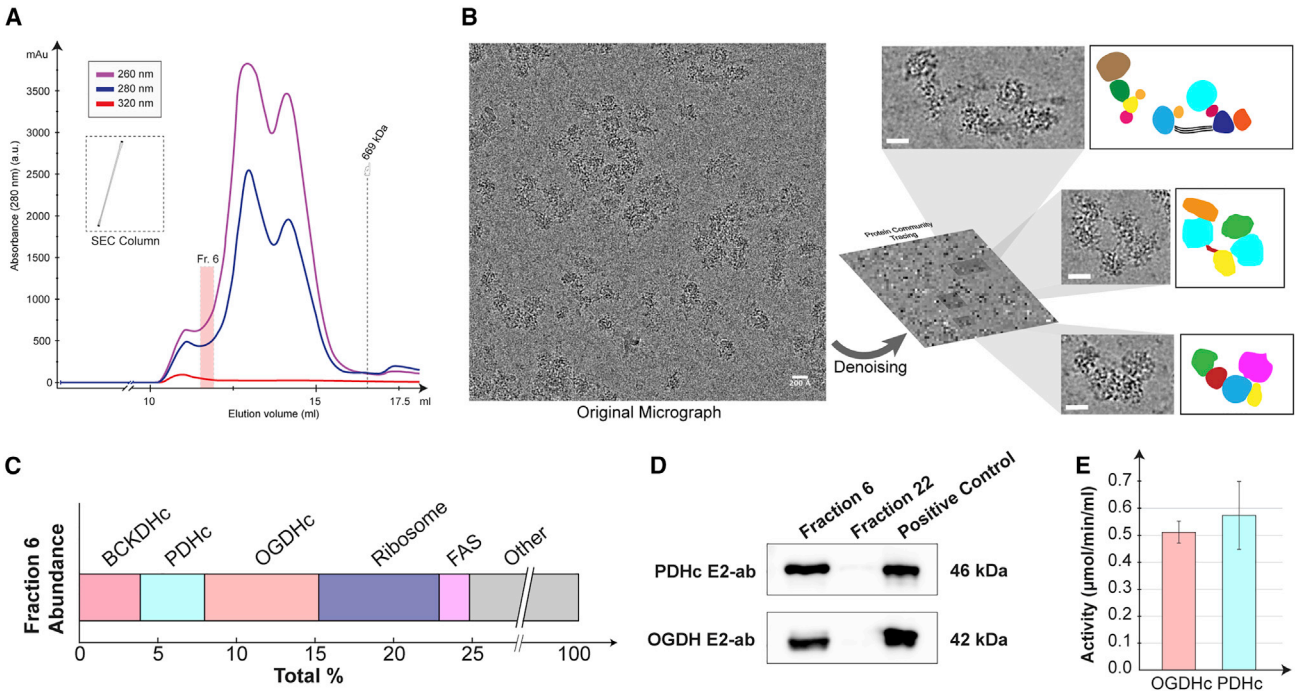


Figure 1. Biochemical characterization of native protein communities

(A) Size-exclusion chromatography profile of a native *C. thermophilum* cell extract. Fraction 6 is located in the MDa weight range.

(B) Cryo-EM of fraction 6 allows, after denoising (Bepler et al., 2020), detection and tracing of various protein community assemblies that remain intact during fractionation and vitrification. Scale bars: 20 nm.

(C) MS data allows for the calculation of the abundance of high-molecular-weight metabolons in the fraction.

(D) Immunodetection of PDHc E2 and OGDHc E2 proteins in fraction 6. Protein amount loaded was 1.0 µg per well and 0.02 µg for the positive control (recombinant protein). Molecular weight (on the right side) corresponds to recombinant proteins used for Ab production.

(E) In-fraction OGDHc and PDHc activity assays. Standard deviation is calculated for three technical triplicates.

oxidation of pyruvic and 2-oxoglutaric acids to acetyl-CoA and succinyl-CoA (Figure 1E). Overall, our simple fractionation approach of a cell extract enriches metabolons without compromising their activity and in addition retrieves megadalton (MDa) assemblies of protein communities of notable complexity.

Identification and recovery of heterogeneous cryo-EM structural signatures within a single cellular fraction

We vitrified fraction 6 of peak 1, optimizing recovery of structural signatures embedded in vitreous ice and ultimately collected a cryo-EM dataset of 2,808 movies. A pixel size of 1.5678 Å was applied to improve statistics of the recovery of the in-fraction metabolons, other communities, and their components at a 200-kV cryo-microscope equipped with a Falcon IIIEC direct electron detector. After template-free particle picking of these micrographs, heterogeneous 2D classes of single-particles were derived, utilizing a convolutional neural network (CNN)-based approach for classification, resulting in high-signal, prominent structural signatures of multiple complexes (Figure 2A). These complexes included, apart from the notable high-contrast central densities, lower-resolution densities at the periphery, indicating highly flexible binders, additional subunits, and conformational plasticity. *De novo*, asymmetric 3D reconstructions using CNNs of those 2D classes showed uniform and discrete Coulomb potential maps of four large assemblies (Figure 2B). These represented 3D signatures of near-cubic (signature 1), near-icosahedral (signature 2), near-dihedral (signature 3) and asymmetric-like (signature 4) MDa particles that, after refinement with CNNs without symmetry, reached ~20 Å resolution (Fourier shell correlation [FSC] = 0.5).

Statistical analysis of the large number of single particles used for the *ab initio* reconstructions shows that 0.65%, 0.93%, 1.20%, and 7.30% of total single-particles (n = 276,339) belonged in each signature, respectively (Figure 2C). This observation indicates that efficient asymmetric 3D reconstruction *de novo* can be realized even for complex biochemical samples, i.e., native cell extracts, using state-of-the-art image- processing software, e.g., cryoSPARC (Punjani et al., 2017). To identify the 3D structural signatures, we employed two complementary strategies: First, we used the Omokage search (Suzuki et al., 2015) to look for similarity hits based on global shapes of the signatures, independent of biochemical and stoichiometric information. In the case of signatures 1, 2 and 3, comparison between hits recovered in the top 10 against the bottom 10 within the top 100 correlation scores shows a statistically significant difference (Figure 2D), providing us with a confident identification of the structural signatures. To unambiguously identify signature 4, a downstream validation that efficiently annotated the signature was performed. A combination of (1) similarity between 2D projections of identified signature and a recovered map from the top-10 hits from the Omokage search and (2) MS abundance data allows for a confident signature identification (Figure S3).

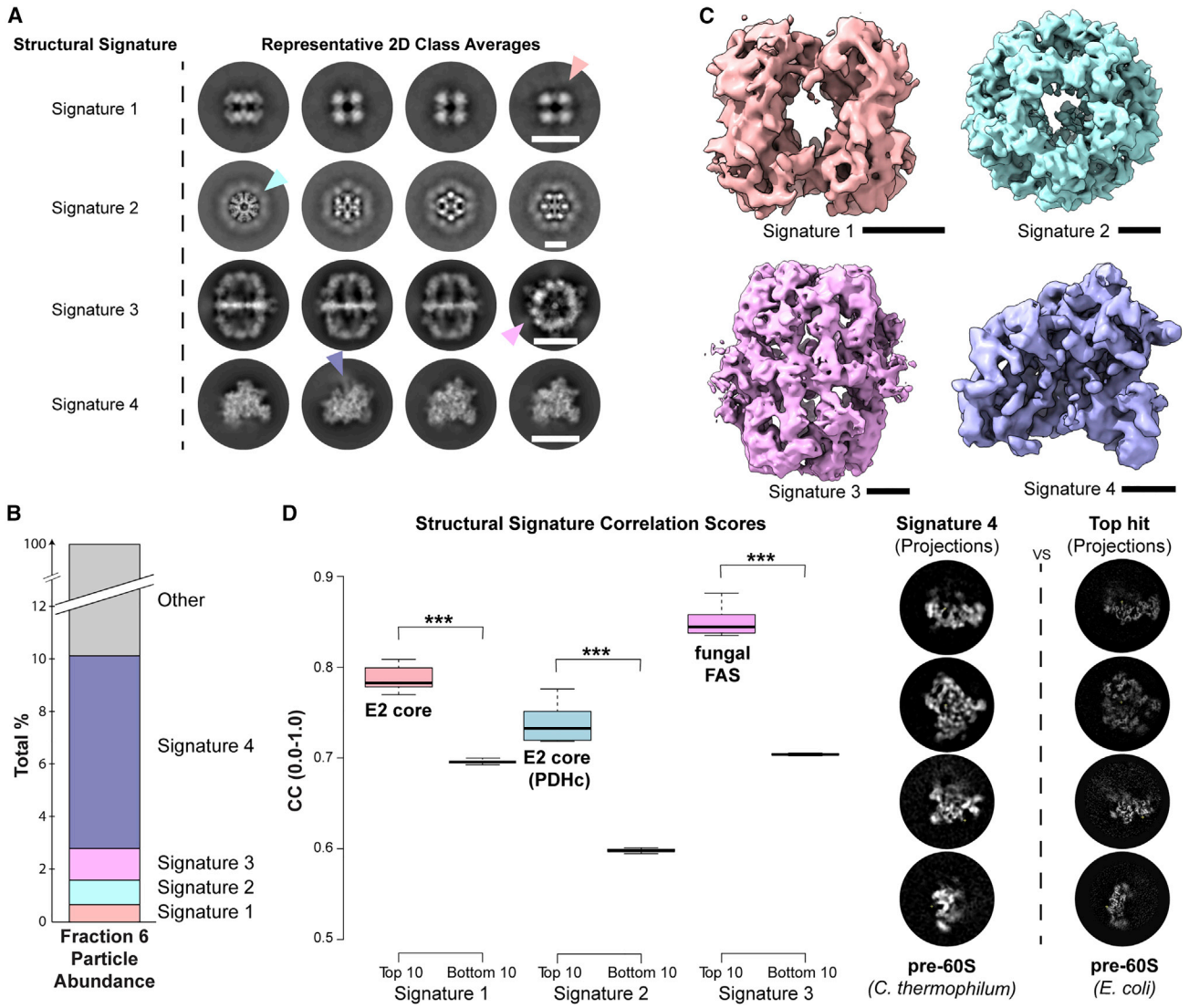


Figure 2. Signature detection and identification

(A) Representative 2D class averages of the most prominent in-fraction structural signatures. Arrows denote diffused densities of lower signal, highlighting structural flexibility or potential binders. Scale bars: 20 nm.

(B) Particle abundance for each signature compared wit total particles initially picked.

(C) *Ab initio* reconstruction of each signature. Scale bars: 10 nm.

(D) Cross-correlation comparison among top 10 and bottom 10 of the top 100 hits returned from the Omokage search (Suzuki et al., 2015) for each signature. 2D projections of signature 4 compared with the matching top-10 hits returned from the Omokage search. p values for the top-10 and bottom-10 comparisons: Signature 1: 3.09×10^{-14}, Signature 2: 2.20502×10^{-12}, Signature 3: 4.23128×10^{-17} ($p < 0.05$).

To summarize, we discovered as top hits and simultaneously reconstructed the *C. thermophilum* community members of (1) hybrid OGDHc/BCKDHc (signature 1) and (2) PDHc (signature 2) E2 core complexes, (3) FAS (signature 3), and the pre-60S ribosome (signature 4) (Figure 2D). Additional validation for the identity of these complexes is provided by the relative abundance of the component proteins in the MS data (Figure 1C). Of note, the pre-60S from a thermophilic eukaryote has not been structurally characterized to date; in addition, only an overexpressed, truncated version of the OGDHc core has been reported (Nagy et al., 2021), whereas here, the native asymmetric and cubic reconstruction of the hybrid OGDHc/BCKDHc core is communicated.

High-resolution reconstructions of endogenous community members from native cell extracts

We recovered refined maps for all four structural signatures that reached 3.84–4.52 Å at the reported FSC = 0.143 threshold (Figures 3 and 4) by applying standard refinement-based procedures on each dataset of classified single particles. All these maps exhibit features that allow their unambiguous identification by repeating the 1D-point searches performed above for their *ab initio* reconstructed Coulomb potential maps (Figure 2C). It is of particular interest that all maps resolve secondary structure elements, including α-helical pitch (Figure 3A), multimeric interface contact points mediated by α-helices (Figure 3B), separation of

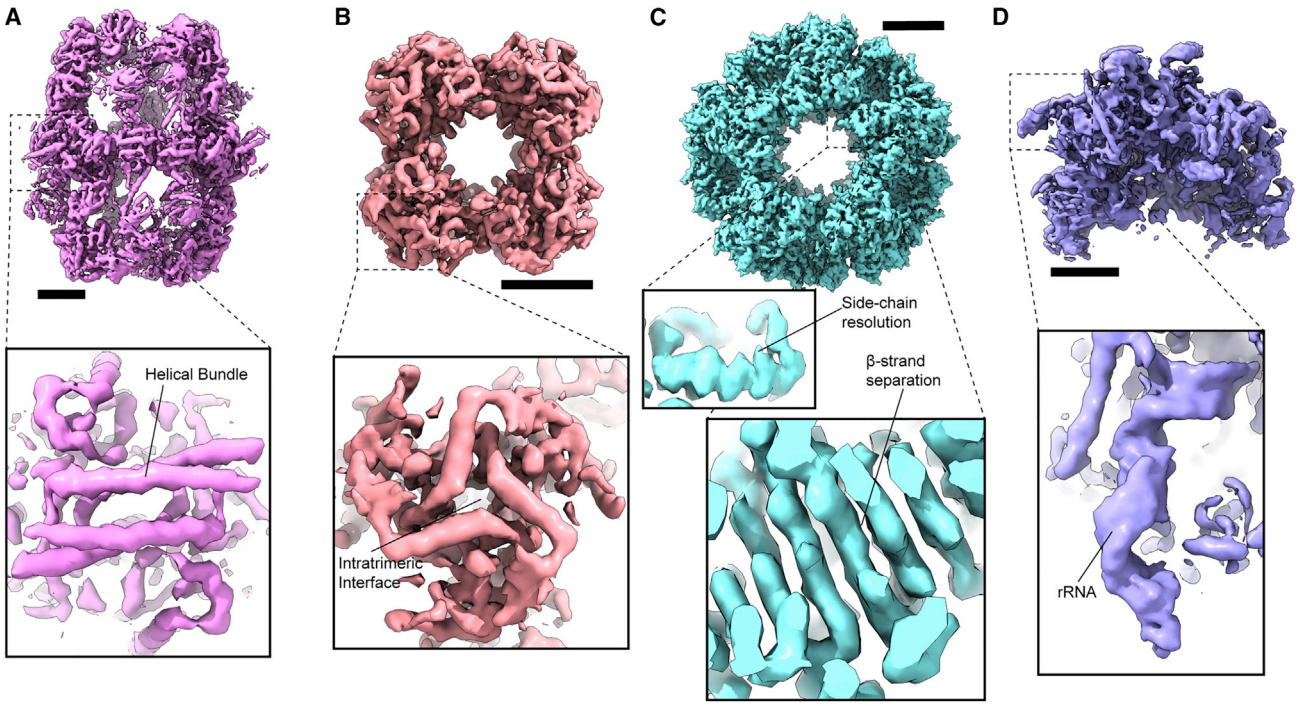


Figure 3. High-resolution signature reconstructions and visible features

(A) Reconstruction of FAS complex. α-Helical bundles and pitch are clearly visible.

(B) Reconstruction of the OGDHc E2 core, where the intratrimeric interfaces at the edge of the core are recapitulated.

(C) Reconstruction of the PDHc E2 core. High-resolution structural features, such as side-chain densities and β-strand separation, are identifiable.

(D) Among other features, in the reconstruction of the pre-60S ribosomal subunit densities belonging to the rRNA structural elements are visible. Scale bars: 5 nm.

β-strands (Figure 3C), and other features of recapitulated resolution (e.g., unambiguously distinguishing their protomers or the RNA component in the case of the pre-60S; Figure 3D).

In addition to the overall calculated high resolution, local-resolution estimation shows that few features of the maps are significantly better resolved than others (Figures 4A–4D). This result points to an overall resolution homogeneity across reconstructed maps. Local-resolution estimation for all reconstructed maps (Figures 4A–4D) indicates regions of resolution that approximate 3.4 Å and, considering pixel size utilized was 1.5678 Å, reach Nyquist. This observation allows one to differentiate rigid and flexible regions of the reconstructed signatures reported. For example, in the case of FAS (Figure 4A), the inner dome densities exhibit higher resolution, often <4.0 Å, whereas external densities display significant flexibility (Figure 4A). Despite the high flexibility observed, intermediate resolutions (~6 Å) are sufficient to allow structure-function insights with integrative modeling. This is especially true for all resolved reconstructions (Figures 4A–4D). The number of particles ending up in each of the derived reconstructions is also noteworthy: The lowest-resolution reconstruction, the pre-60S, includes 35,773 particles, a phenomenon that indicates high plasticity and/or chemical heterogeneity. On the other hand, it is rather unexpected that, with only 1,819 single particles and the application of octahedral symmetry, a map is retrieved for the E2o/b core with clear improvements in recovery of densities, e.g., corresponding to α-helices. Therefore, these results show that reconstructions from cell extracts can be retrieved at near-atomic resolutions even for rare protein-community members.

Distinct localization of the PPT acetyl-CoA-binding domain in the FAS metabolon of *C. thermophilum*

To understand functional aspects of each of the structural signatures reconstructed, a comparison with available data for similar complexes was performed. In the case of the FAS from *C. thermophilum* (Kastritis et al., 2017), the structure was previously resolved to lower resolution, where the acetyl-CoA-binding domain of phosphopantetheinyl transferase (PPT) was not observed (Figure 5A). In the current reconstruction, the domain is clearly present, and resolution allows one to unambiguously resolve its model and relative position to the FAS dome. The acetyl-CoA-binding domain has been resolved in yeast, e.g., in the work of Singh et al. (Singh et al., 2020). Overall map similarity to yeast FAS allowed the retrieval of significant differences at the densities matching to the acetyl-CoA-binding domain. This domain is highly conserved and resides within the sequence corresponding to the PPT domain, located at the C terminus of the α-subunit. To resolve the conformational difference, we real-space-refined the acetyl-CoA-binding domain of *C. thermophilum* in its density and compared its architecture in relation to the FAS dome with the yeast ortholog. Results show that the acetyl-CoA-binding domain undergoes significant rotation and translation with respect to the dome (Figure 5B), mediated by a solvent-accessible helix-turn-helix region spanning residues Asn1709-Arg1737 in *C. thermophilum* (Figure 5C) and residues Ser1730-Lys1761 in yeast (Figure 5D). In this respect, both *C. thermophilum* and yeast domains display the same local conformation but, additional flexibility is manifested in the

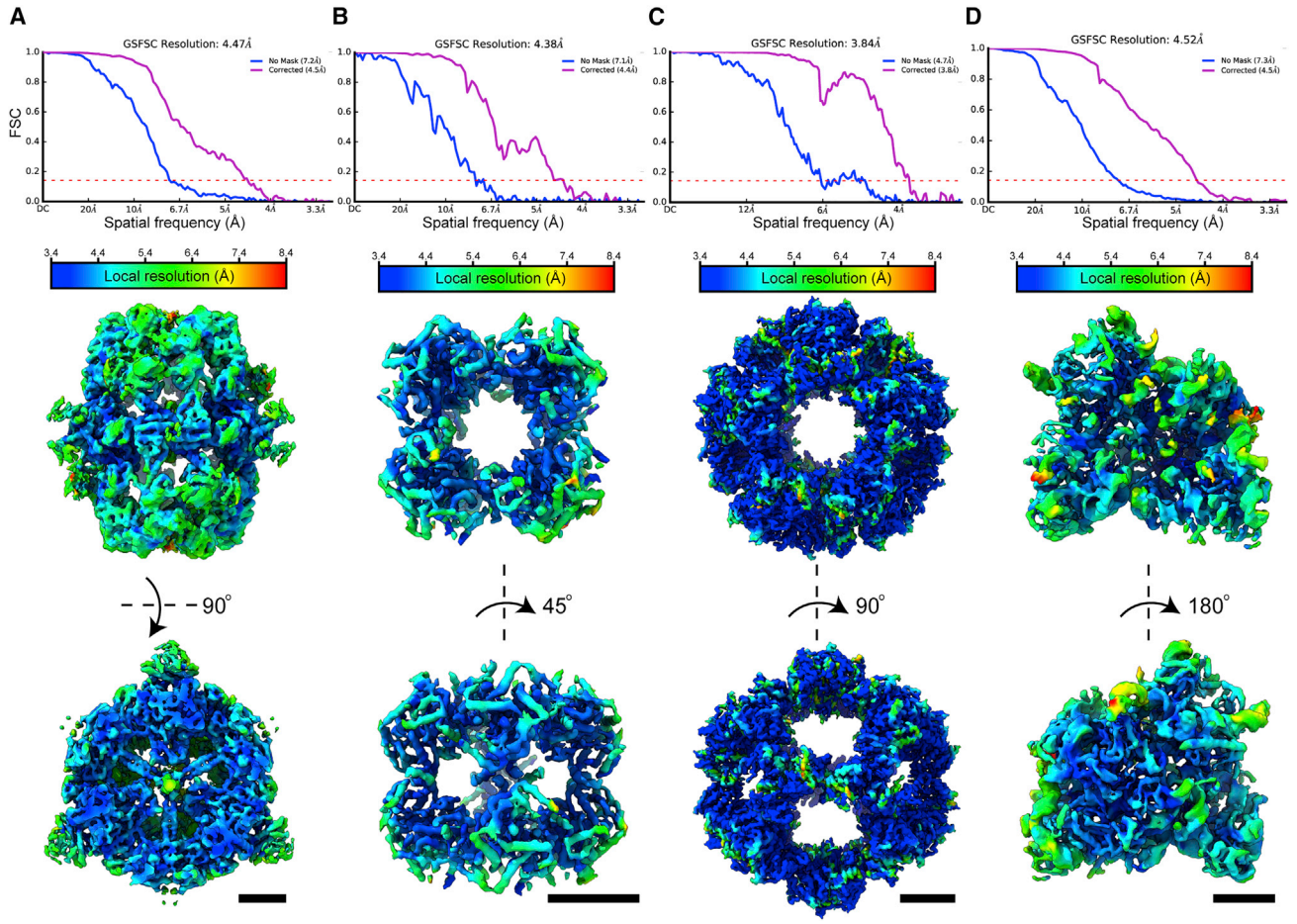


Figure 4. FSC plots and local resolution distributions for all reconstructed maps

(A) FSC resolution plot and local resolution distribution for FAS. Central core of the complex demonstrates an overall higher resolution compared with the external, more flexible densities.

(B) FSC resolution plot and local resolution distribution for OGDHc E2 core. External densities demonstrate more flexibility compared with the inner part of the complex's core.

(C) FSC resolution plot and local resolution distribution for PDHc. The reconstruction demonstrates high and uniform resolution distribution.

(D) FSC resolution plot and local resolution distribution for the pre-60S ribosomal subunit. More flexible external densities corresponding to the rRNA components are recognizable. GSFSC, gold standrd FSC. Scale bars: 5 nm.

C. thermophilum sequence, specifically residues Thr1717-Ser1726 (Figures 5C and 5D). This local conformation change allows the acetyl-CoA-binding domain to explore distinct local, rigid movements, possibly facilitating direct interaction with the acyl carrier protein (ACP) for the critical ACP 4′-phosphopantetheine modification to occur.

Persistent vicinal E2p core densities may act as anchor points for the E3BP

Through this reconstruction, we can observe in the PDHc core new map optical densities on the inside of the E2p trimers (Figure 5E). These densities are in close proximity to a charged region of E2p, and a similar density was observed in a mesophilic counterpart (Forsberg et al., 2020). This density was hypothesized to belong to the E3 binding protein (E3BP) and its interaction interface with the E2 (Kyrilis et al., 2021b). Here, we not only validate the observed anchor point, but also derive another reconstruction from a thermophilic species sharing the same feature. The common presence of this additional density in two distinct fungal species points to a possible conservation of the E2-E3BP interface across fungi. Note that this density was retrieved in the hereby reported *C. thermophilum* PDHc E2p core map and was not previously observed in its lower-resolution, symmetrically resolved cryo-EM counterpart (Kyrilis et al., 2021b). This result suggests that, with increasing resolutions of protein community members, additional densities might appear even in symmetric reconstructions. Careful evaluation of recovered densities should be performed to avoid overinterpretation of symmetry artifacts; these same densities can also correspond to the presence of macromolecules with a different symmetry, as in the case of the PDHc core.

Discovery of the low-abundant, native α-keto-acid dehydrogenase E2 core

A cubic reconstruction of a protein community member was retrieved after image processing, which matched the E2 core of eukaryotic α-keto-acid dehydrogenases. Notably, reconstruction

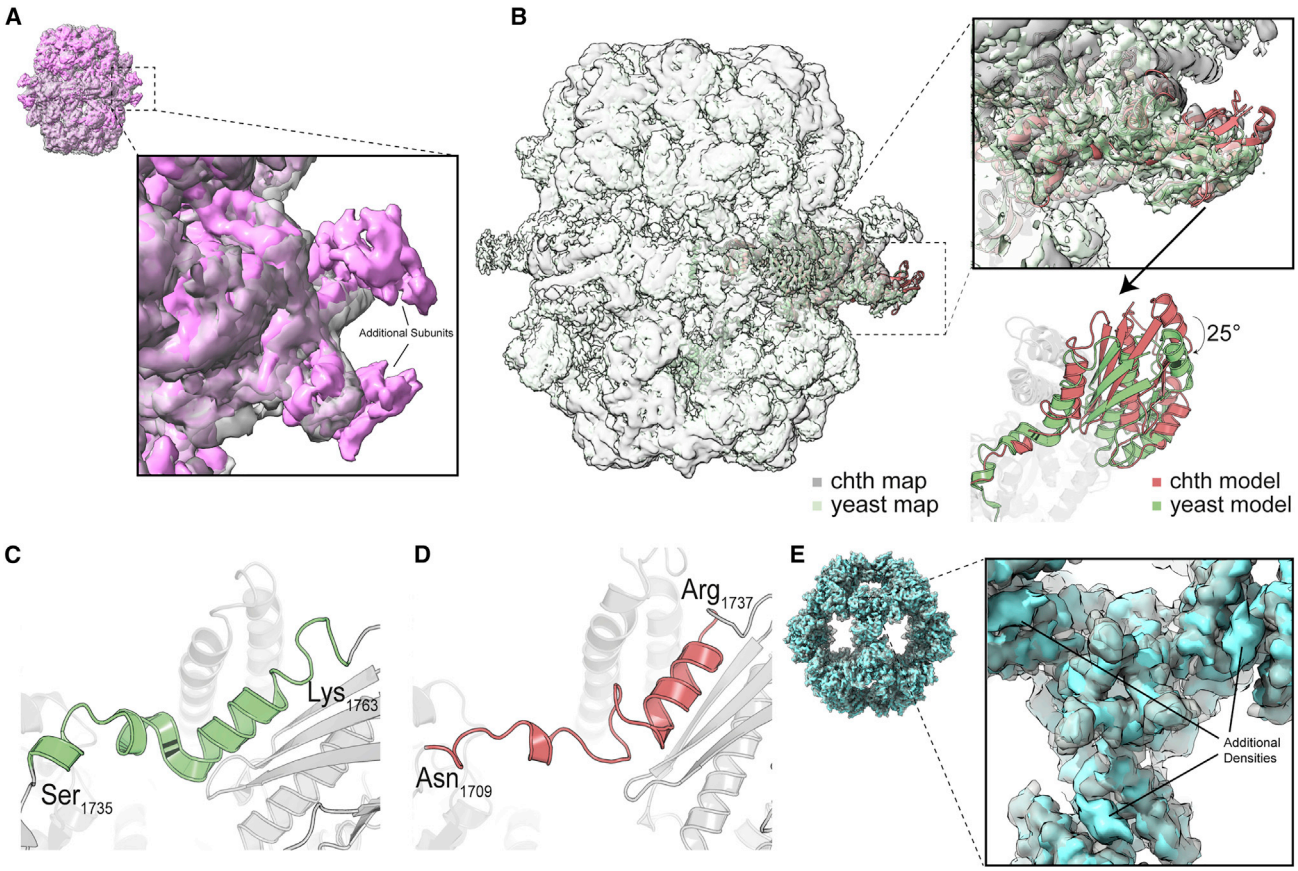


Figure 5. Structural insights into the FAS and PDHc reconstructed maps

(A) Compared with previously resolved *C. thermophilum* FAS, new densities can be identified.

(B) Comparison of the acetyl-CoA-binding PPT domains of the *C. thermophilum* (chth) and *S. cerevisiae* (yeast) reveals the lateral movement of the domain.

(C) Flexible linker region of the acetyl-CoA-binding PPT domain of yeast.

(D) Flexible linker region of the acetyl-CoA-binding PPT domain of *C. thermophilum*. Longer length of unstructured region may provide explanation for the domain's higher flexibility.

(E) Comparison with the previously resolved *C. thermophilum* PDHc E2 core reveals the existence of additional densities on the inside of the core, possibly indicating the anchor points of the E3BP.

included 1,819 single particles, a surprisingly limited number for reconstructing a protein community member. Nevertheless, a reconstruction at 4.38 Å was derived, which clearly showed an E2-like protein assembled in octahedral symmetry in 24 copies. This E2-like protein includes the ordered C-terminal (C-ter) domain, similar to the one previously used to determine the truncated *in vitro* reconstituted counterparts from human OGDHc C-ter E2o (Nagy et al., 2021) and bovine BCKDHc C-ter E2b (Kato et al., 2006). Our map reveals an endogenous α-keto-acid dehydrogenase community member whose core component retains the same architectural principles as the overexpressed C-ter constructs. The map may equally well accommodate the placement of E2o and E2b with high cross-correlations (CC), i.e., CC_{maxE2o} = 0.91 and CC_{maxE2b} = 0.80. Although qualitatively, E2o fits with slightly higher CC values (Figure 6A), the values within the distribution of fits are not significantly higher compared with the corresponding core from the E2b. Conclusively, the core component of the α-keto-acid dehydrogenase community retains a cubic architecture with no inner visible densities, unlike the PDHc E2p core.

A thermophilic pre-60S complex is composed of 22 core ribosomal proteins and exhibits substantial flexibility

A reconstruction of the thermophilic pre-60S was derived at 4.52 Å (FSC = 0.143) using 35,773 single particles. Volume-based cross-correlations indicated a reasonable fit with both *Homo sapiens* and *Saccharomyces cerevisiae* pre-60S (Wu et al., 2016; Liang et al., 2020) (Figure S4). The arch-like architecture of the reconstruction points to a nucleolar assembly state for *C. thermophilum*, as previously observed for the yeast equivalent (Kater et al., 2017). Resolution allowed identification of the rRNA component and confident localization of 22 ribosomal proteins in the EM map (18 from the yeast [Wu et al., 2016] and 4 from the human pre-60S [Liang et al., 2020]) (Figure 6B). Following the recent nomenclature system (Ban et al., 2014), ribosomal proteins organizing the *C. thermophilum* pre-60S are uL4, uL5, uL6, uL13, uL15, uL18, uL22, uL24, uL29, uL30, eL6, eL8, eL13, eL14, eL15, eL18, eL20, eL21, eL29, eL32, eL33, and eL36. Mapping the gene ontology (GO) terms for the identified proteins, a few (uL5, uL18, and eL6) have been annotated in yeast to participate in large ribosomal assembly (GO:0000027),

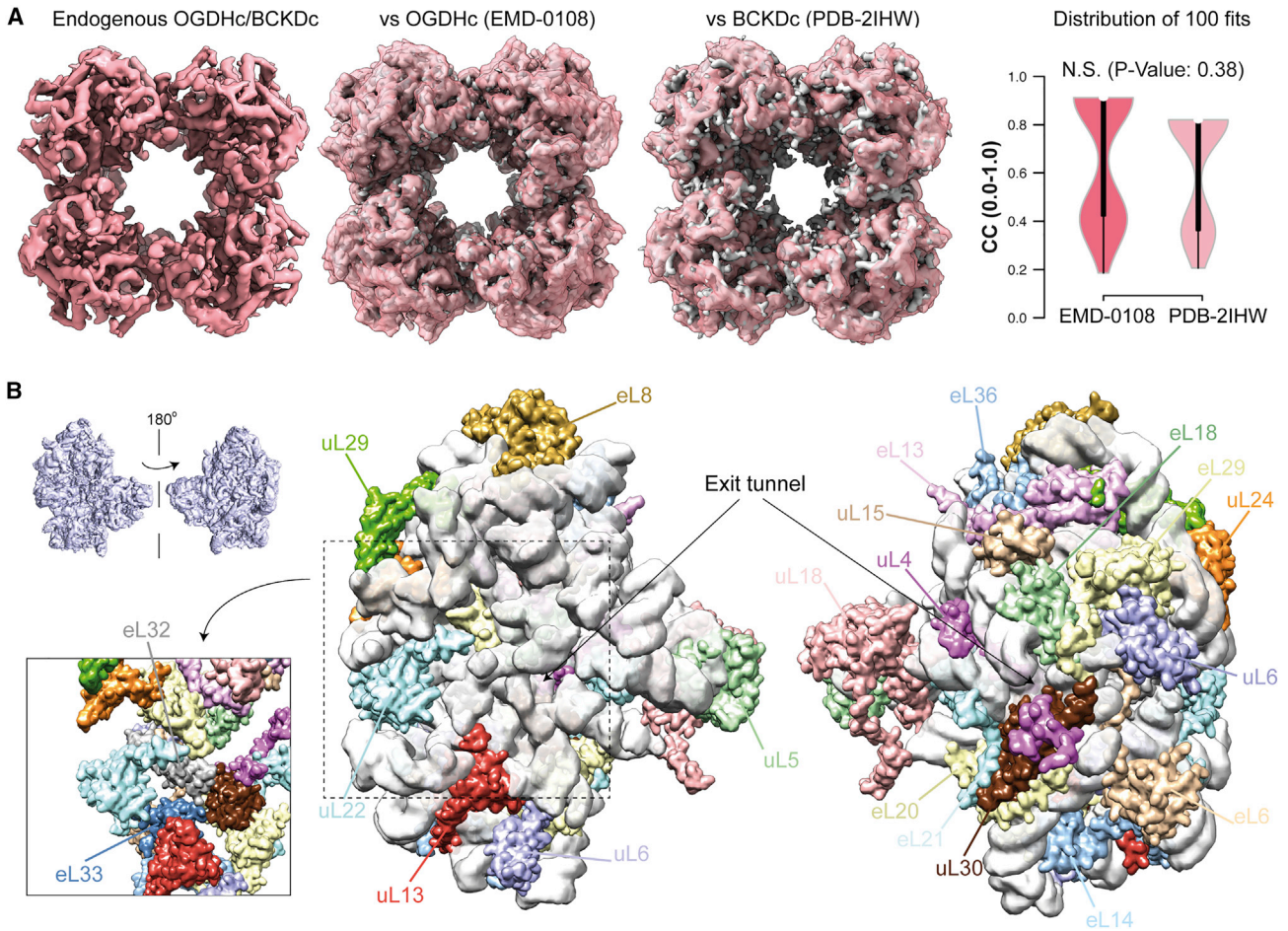


Figure 6. Structural insights into the hybrid OGDHc/BCKDHc E2 core and pre-60S ribosomal subunit reconstructed maps

(A) Distribution of fits with overexpressed OGDHc and BCKDHc E2 core maps (gray) does not allow the unambiguous identification of the endogenous reconstructed hybrid OGDHc/BCKDHc map (salmon).

(B) Identifiable ribosomal subunits taking part in the assembly of the endogenous pre-60S ribosomal subunit. The exit tunnel for newly synthesized polypeptide chain is one of the first features visible during the 60S assembly sequence.

while several (uL24, uL30, eL14, and eL33) are annotated to participate in its biogenesis (GO:0042273). Annotation of proteins is presented in Figure 6B, showing that an immature pre-60S is formed that organizes around the exit tunnel, pointing to the potential importance underlying its early formation.

De novo protein sequence identification from reported cryo-EM maps is feasible

We previously reported that a global shape similar to existing structures in PDB allowed for an identification of presented protein macromolecular complexes by using Omokage (Figure 2D). However, an orthogonal measure of protein identification within cryo-EM maps could be the possibility of interpreting the density by residue-specific matches. This would mean that the map is annotated using its intrinsic characteristics completely *de novo* without a similarity search in known structural databases. In addition, results from similarity searches may be ambiguous, as distinct sequences may form complexes of very similar topology. Map resolutions of, e.g., 4.38 Å, as is the case for signature 1, do not allow for unambiguous shape-based interpretation. Indeed, we observed this in the case of the octahedrally symmetric oxo-acid dehydrogenase complex core, which has a comparable global shape to C-terminal domains of human OGDHc core (E2o) (Nagy et al., 2021) and bovine BCKDHc core (E2b) (Kato et al., 2006). Corresponding *C. thermophilum* sequences from UniProt (UniProt, 2021), G0SAX9 and G0S0D3, respectively, have ~22% sequence identity. To resolve this ambiguity, a complete workflow was devised (Figure 7). To begin with, both homotrimeric structures were predicted with AlphaFold2 (Jumper et al., 2021) by using the ColabFold advanced notebook (Mirdita et al., 2021). Resulting models were fitted into the EM map and refined in real space. The model's backbones and signature's map were then used as input to the findMySequence (Chojnowski et al., 2022) program to identify corresponding sequences in the *C. thermophilum* proteome, not accounting for the MS data. In both cases, the program selected the E2o sequence, achieving a significantly better score for the corresponding E2o model (E-value 67.9e-30 versus 1.7e-3 for E2b). This result also signifies the absence of E2b from the final particles, which may be attributed to various causes, including, but not limited

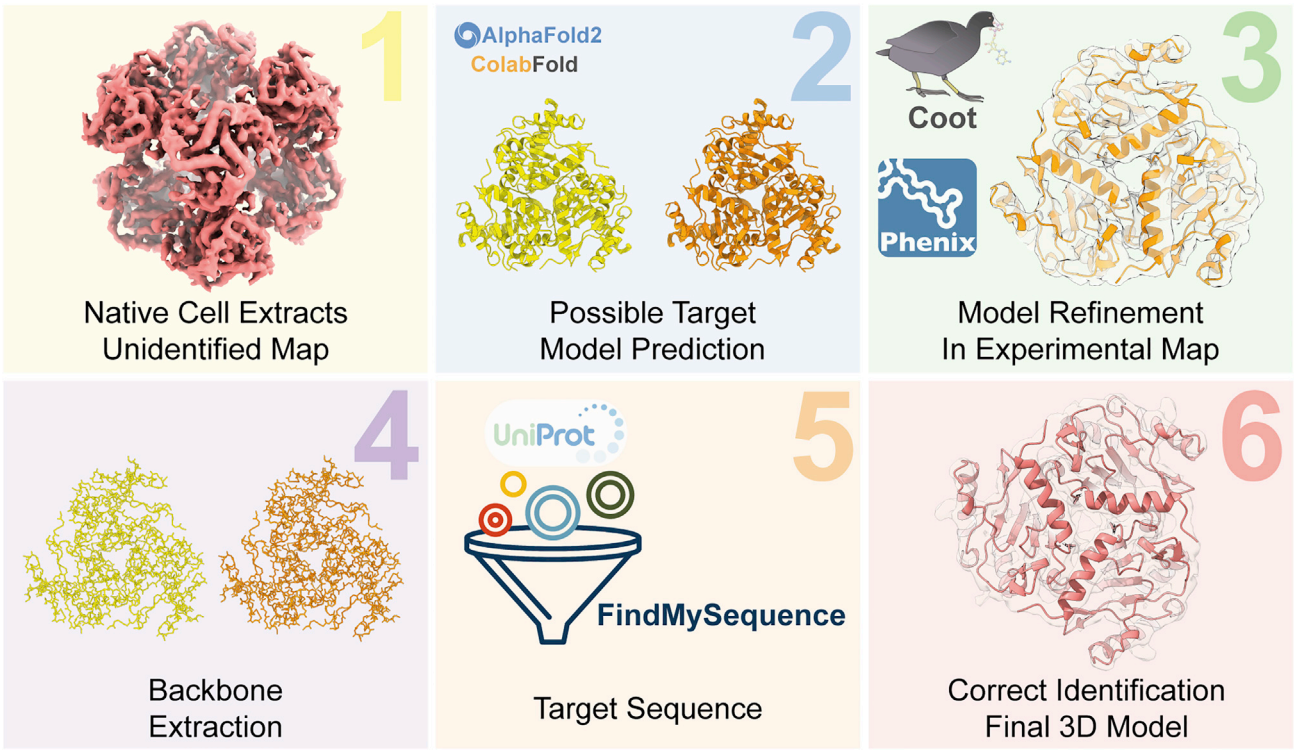


Figure 7. Scheme illustrating the workflow employed for the *de novo* identification and reconstruction of the OGDHc E2 core model derived from native cell extracts

to, higher peptide flyability in the MS, distinct surface properties of the BCKDHc metabolon resulting in preferential presence in the vitreous ice, metabolon stability and/or flexibility that would impact its presence in the final high-resolution classes during image processing, transient interactions of its community members that would lead to dissociation during freezing, etc. However, differentiating the previously unknown E2o core from the E2b shows that a distinction can be performed at the sequence level even at resolutions >3 Å, up to ~4.5 Å. Similar analysis for the FAS and the PDHc reconstructions validate and further highlight the possibility of *de novo* map interpretation of lower-resolution reconstructions from endogenous cell extract material (see STAR Methods).

Cryo-EM AI-assisted atomic modeling allows identification, improvement, and understanding of protein community members

Derived resolution for the protein community members of PDHc, the native α-keto-acid dehydrogenase E2 core and FAS allows atomic model building, especially with the recent major advances in structure prediction utilizing AI (Baek et al., 2021; Jumper et al., 2021). Therefore, atomic models were built based on structures predicted using AlphaFold2 (Jumper et al., 2021) within ColabFold (Mirdita et al., 2021), an advanced interface that permits modeling multimeric structures. Results show that atomic models are of sufficient quality (Figures S5 and S6 and Table S1), and combined with the previously reported protocol for *de novo* sequence identification (Figure 7), allowing unambiguous fitting to the corresponding EM reconstructions.

In detail, for PDHc and OGDHc core structures, reliable AlphaFold2 predictions of complete trimeric sub-complexes were obtained, which assisted in the accurate modeling of the most important intramolecular interfaces, devoid of steric clashes. Interestingly, comparison of the AlphaFold2 PDHc core predicted trimeric model and the final model refined in the cryo-EM map reveals that the AlphaFold2 model still has the folded N-terminal (N-ter) helix on top of the E2 core (Figure 8A). This structural element is absent from the experimental cryo-EM map, showing that its folded state must be absent or transiently present during CoA acetylation or in an inactive E2 core enzyme, possibly in an organism-specific manner. In the active PDHc enzyme, this secondary structure element is not observed (Figure 8A).

For OGDHc, the AlphaFold2 trimeric model was correctly resolved, especially the structural component of the E2 core (Figure 8B). However, the initial AlphaFold2 model has in close proximity the lipoyl domain (LD) bound to the E2. Note that the LD is carrying the lipoyl chain within the large metabolon for the succinylation reaction to occur within the E2 core active sites. Therefore, at one point, the interaction must be somehow present. Although AlphaFold2 predicts the relative proximity of the LD with the E2 core with low confidence, their direct interaction allows the insertion of the lipoyl moiety directly in the E2 active site. However, this conformation of the LD, in proximity of the E2 is not captured in the experimental endogenous BCKDHs or in the derived cryo-EM data. This conformation denotes either a bound conformation for the catalytic intermediate of succinylation or an encounter complex of low stability. This

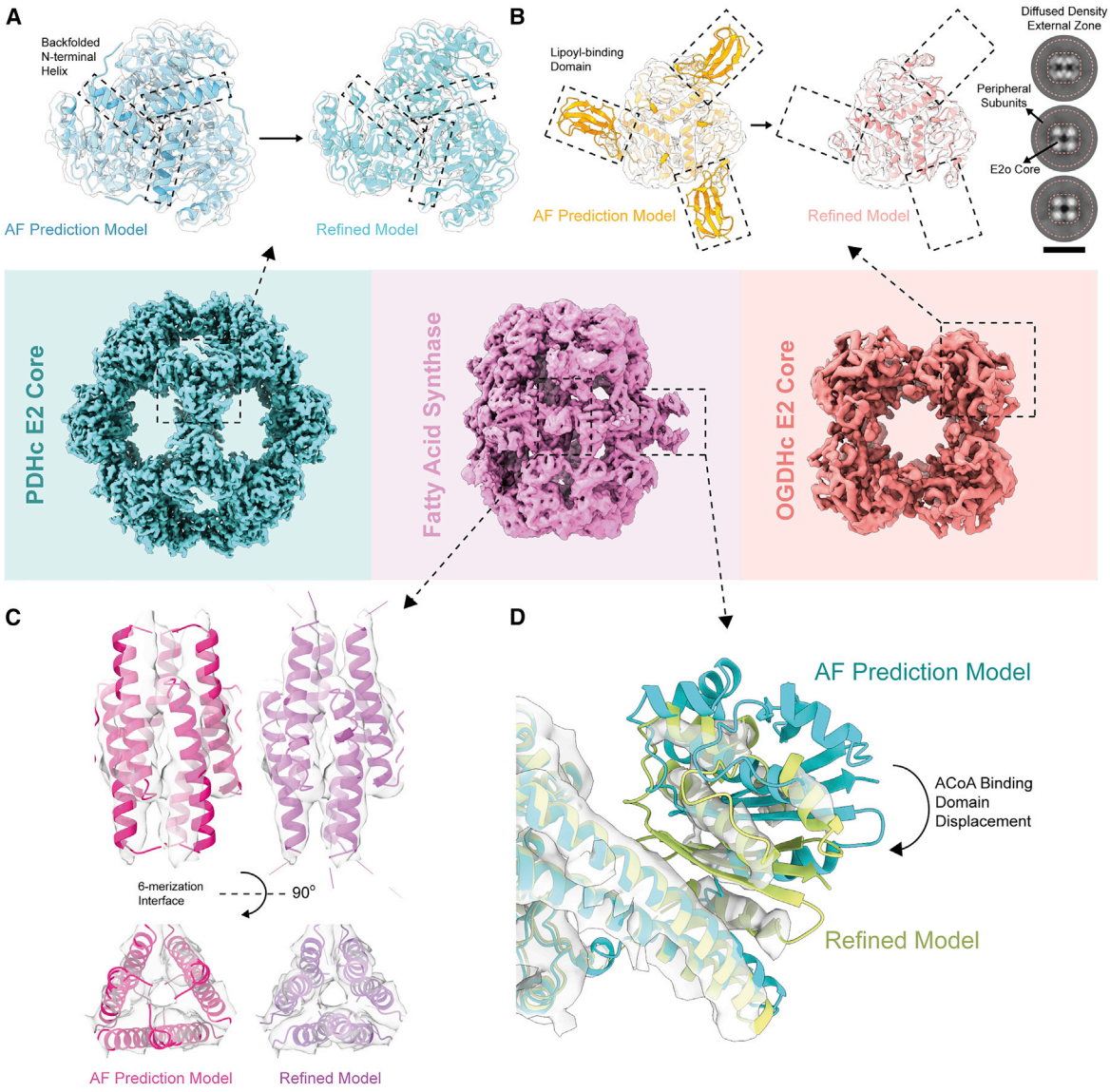


Figure 8. Artificial intelligence-predicted models present critical differences compared with experimental data

(A) AlphaFold2 predicts a backfolded N-ter helical structural element (left), which is absent from the model that is refined in the experimental map (right).

(B) A lipoyl-binding domain is predicted to be bound on the OGDHc E2 core trimer by AlphaFold2, where the same is clearly absent from the experimental map data. 2D class averages of OGDHc particles show the same, where there is a strong signal for the E2 core but a diffused and weak signal for the peripheral subunit densities. Scale bar: 20 nm.

(C) AlphaFold2 accurately predicts the overall fold of the hexameric sub-complex of AT domains of the *C. thermophilum* FAS, revealing possible interface information, even before refinement in the experimentally obtained map.

(D) AlphaFold2-predicted model of the acetyl-CoA-binding domain of PPT is once again shown to be displaced compared with the model that was refined in the experimental map.

conformation is also not observed in the cryo-EM 2D class averages by any means, showing an opposite effect via outer subunit diffusion (Figure 8B).

A basic component of FAS is a heterodimer, too large to model as a whole with AlphaFold2 due to GPU card memory limitations. Therefore, for this molecule, six overlapping fragments were predicted, independently fitted, refined into EM reconstruction, and merged. Although most inter-chain interfaces of the complex could not have been accurately predicted (e.g., hexameric sub-complex of AT domains [Figure 8C]), we were able to obtain a reliable model of MPT domain, formed by an intermingled, tight interaction of α- and β-subunits, which would be otherwise very difficult to model given limited resolution of the reconstruction. It is striking that when the D2 symmetry is applied for multiplication of the monomer, an interface is discovered within the FAS dome (Figure 8C). This derived interface superimposes almost perfectly with the experimentally measured homologous interfaces from yeast. Results show that, although the machine-learning algorithm was not previously trained with multimers (which was recently

communicated [Evans et al., 2021]), information regarding protein-protein complexation is still encoded.

Finally, we previously derived a homology model for the acetyl-CoA-binding domain of *C. thermophilum* FAS PPT (Figure 5B). This domain is also present in the AlphaFold2 model, albeit in the same conformation as the yeast homolog (Singh et al., 2020). This conformation as shown in Figure 5 is not the optimal for *C. thermophilum* FAS, as a displacement via rotation of ~25° occurs as revealed by the cryo-EM data (Figure 8D). Again, relative proximity of the PPT domains is predicted with low confidence for the AlphaFold2 model, and therefore, the cryo-EM map was really instrumental for the unambiguous placement of the acetyl-CoA-binding domain within the final FAS model.

DISCUSSION

Rigorous and systematic analysis of cell lysate fractions has previously shown recovery of protein complexes and high-resolution information (Ho et al., 2020; Su et al., 2021); however, in-depth structural investigation of protein community members was unexplored. Despite being combined with other integrative structural biology methods, such as mass spectrometry, protein community member identification was limited to ~4.7–7.0 Å (Kyrilis et al., 2021b). In addition, these community members were the single structural species resolved within a cell extract fraction, composed of various biomacromolecules. Here, taking advantage of modern cryo-EM schemes and recent AI-based algorithms for image processing and protein modeling, we were able to retrieve high-resolution structures for four different structural signatures reaching the sub-4.6-Å regime that corresponded to multiple protein community members. Note that final reconstructions were derived utilizing "blob picking," which can severely bias the centering and introduce inaccuracies when proteins stick together, which is reasonable to consider for intricate biomolecular structures such as protein communities. Therefore, a lot of proteins may be missed using our current picking strategy, and complementary or new picking strategies for metabolons or large protein communities should be considered in the future. Despite this limitation, four different structural signatures were derived from a single fraction, reaching the sub-4.6-Å regime.

Predicted AlphaFold2 models are accurate and can be used for two distinct steps in our pipeline. First, to characterize the sequence of a certain map *de novo*, based on the modeled Alphafold2 main chain and subsequent search in sequence databases, we have shown that sufficient resolution to optimally fit the full sequence is not a necessary prerequisite, expanding the field of *de novo* sequence identification to resolutions as low as ~4.6 Å. Previously, *de novo* polypeptide chains were modeled in overexpressed or highly purified endogenous species (Ho et al., 2020; Su et al., 2021; Chojnowski et al., 2022), while our work suggests that protein community members can be modeled *de novo* at near-atomic resolutions despite their substantial complexity and inherent flexibility. Second, initial models for refinement in cryo-EM maps belonging to community members from native cell extracts can be used and then optimized in the experimental densities of protein community members. We, however, observed that certain regions within the models are "overfolded" or limited in terms of domain proximity. Examples include (1) the additional LD predicted as bound in the AlphaFold2 model but not visible in the OGDHc E2 core experimental map; (2) the back-folded N-ter helical region that is absent in the experimental data for PDHc; and (3) the conformation change for the FAS acetyl-CoA-binding domain that is not observed in the experimentally measured conformation for either *C. thermophilum* or yeast FAS. Specifically for this conformation change observed after experimental refinement of both homology-based and AlphaFold2 models, this adaptation is highly intriguing. Besides the difference of the linker regions described in the Results that would lead to obvious differences in exploration of distinct conformations, another functional adaptation could well be present. This observed increased flexibility of the linker region of the acetyl-CoA-binding PPT domain of *C. thermophilum* can possibly be a community-embedded specific adaptation that facilitates substrate channeling to FAS from proximal community members. If this is the case, it is rational that is not observed in yeast, as already published models are derived from highly purified and homogeneous specimen.

These results underlie the fact that the algorithm was trained with structural models for which the vast majority is derived from crystallographic structures of overexpressed and purified to homogeneity, crystallized specimen, often isolated from heterologous systems, outside the cellular millieu. Although the folded state of individual domains is accurately predicted, their mutual orientation can substantially differ in solution, exemplified by the vitrified specimen from endogenous protein communities from native cell extract that we have studied. Both structural bioinformatics extensions of our protocol to study native cell extract architecture will offer unprecedented opportunities to identify, characterize, model, and understand protein communities and their members from complex endogenous mixtures.

Previous studies have already proposed that the idea (and application) of cell lysate fractionation can allow us to better understand community organization and function (De Duve and Berthet, 1954; Kastritis and Gavin, 2018) in a way that the native context is retained (Verbeke et al., 2018). Nevertheless, resulting samples show an increased level of complexity, making their analysis by a single technique quite prohibiting. The application of modern integrative structural biology methods is indispensable for obtaining the essential complementary insights and provide meaningful results. This is exemplified by the fact that protein communities are rarely in a linear assortment of biomolecular complexes and in the 2D micrograph are expected to overlap, complicating efficient and clear 2D class averaging. Such a limitation can be tackled with complementary cryo-EM methods such as cryo-electron tomography (or electron cryo-tomography, cryo-ET). On the other hand, tomography is slower in the output compared with single-particle cryo-EM (despite impressive recent advances in tilt series throughput [Bouvette et al., 2021]), and therefore, a fusion of both methods can be considered.

Despite the overall advances of our methodological implementations to understanding how protein communities are structured, limitations are quite apparent. To begin with, ice thickness can already introduce bias concerning the visibility and analysis of protein communities and their members. There is a possibility that thinner ice, which is a prerequisite for high-resolution

single-particle analysis, may prohibit a more complete view of the protein community organization due to preferential orientations or denaturation at the air-water interface. Alternative vitrification strategies should also be considered that are optimized not for protein community member analysis but rather for the visualization of the complete communities, such as thicker ice, one-side botting, sample spraying methods, etc. Such strategies can nicely be complemented with data collection strategies optimized for cryo-ET.

In addition, due to the extreme complexity of the sample, the number of single-particles corresponding to each unique signature is limited to a few per micrograph (or even less, e.g., in the case of OGDHc). This means that, for the final reconstructions, although achieved resolution is quite high, allowing atomic model building with AlphaFold2 and subsequent refinement, conformational flexibility and structural states are currently very challenging to be captured. For example, the movement of the acetyl-CoA-binding domain of FAS PPT could be further analyzed with 3D classification without imposing symmetry or new strategies, such as 3D variability analysis (3DVA) (Punjani and Fleet, 2021). With the lack of classification, good resolution is achieved, but functional interpretation might still be limited due to the dynamic nature of protein communities and their members. Despite that, acquisition schemes with 300-kV microscopes and newer electron detectors can increase data throughput by more than 10 times (Efremov and Stroobants, 2021). Such hardware, will, in principle, resolve the above-mentioned limitation.

This kind of complex elucidation of cellular fractions heavily depends on a semi-automatic approach, with the investigator having to rigorously shift through daunting amounts of data to identify, separate, and reconstruct members of the protein communities during image analysis. Such processes inherently pose the risk of sample misinterpretation and generally bias the analysis pipeline toward protein assemblies that are easier to identify in the sample. For this reason, it is paramount to stress the importance of complementing biochemical assays performed here, identifying active pyruvate and α-ketoglutarate oxidation activities. Novel activity assays have to be designed in the future to verify the integrity and function of other protein community members. This is critical for those community members that so far lack one and could also be used as a discovery-based approach for targeting functional extracts and, therefore, retrieved active communities. Furthermore, novel analysis pipelines employed at different steps of protein community studies that rely more on machine-learning algorithms (Kyrilis et al., 2021a) should always be considered, since they can minimize user intervention. Even though our work took advantage of the latest developments in neural-network technology for image analysis (Punjani et al., 2020), simultaneous *de novo* reconstructions of chemically heterogeneous mixtures are still out of reach.

Contemporary cryo-EM methods, notwithstanding their suitability for the study of higher-order assemblies, still pose a risk to retrieving structural data for dissociated, artificial, or inactive protein communities. This can be ameliorated with the choice of the correct model organism. Proteins of a thermophilic, eukaryotic organism, like the one used in this study, display higher stability, which may contribute to the preservation of higher-order organization throughout the lysis, fractionation, and vitrification processes. Thermophilicity-specific adaptations can be additionally underlying the observed structural variation of protein community members, which can confer thermostability advantages in addition to their involvement in the community (i.e., distinct species-specific flexibility of unstructured linker of the PPT acetyl-CoA-binding domain of FAS). Additionally, follow-up analysis of fractions with cross-linking mass spectrometry (XL-MS) as we previously performed (Kastritis et al., 2017; Kyrilis et al., 2021b), or even simple MS analysis (Ho et al., 2020), can further aid in the validation of protein community member identities and their interactions within. Another aspect in the structural analysis of protein communities should be toward membrane interactions and membrane-associated complexes. Other works have shown that various protein community members require either close proximity to a membrane structure (Sanders and Hutchison, 2018) or to be embedded into one in order to perform their function (Guo et al., 2020). For the aforementioned reasons, membrane mimetics can be added to a fractionated sample with the goal of recovering such interactions for subsequent cryo-EM analysis.

In this work, we demonstrate the feasibility and potential that is inherently held by the study of cell extract fractions as a means of elucidating the organization of native protein communities. We reconstruct and properly identify multiple structural signatures belonging to members of these communities by combining unobtrusive sample preparation with the goal of retaining intracellular principles with modern cryo-EM data collection and analysis. We extend our methodological approach with the incorporation of experimentally refined, AI-driven cryo-EM image processing routines, *de novo* sequence identification methods, and protein model prediction algorithms in order to provide deeper insights into protein community member structural adaptations. Our combined cryo-EM/AI pipeline is complemented by classic biochemical assays, mass spectrometry, independent statistical analysis, and traditional bioinformatic methods, thus adding another level of verification to our findings from cryo-EM. This type of pipeline can be considered a prime example of how an in-depth analysis of native protein community members can be achieved, providing us with critical information concerning their identity and general organizational principles as well as novel structural insights of community members that can aid in interpreting their function in this context. We are confident that this work can act as a guideline to broaden our understanding of native protein community structure, function, and interactions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Model organism
- METHOD DETAILS

- Mycelium harvesting, culture lysis and sample preparation
- Size exclusion chromatography
- Pyruvate dehydrogenase (PDC) and oxoglutarate dehydrogenase (OGHD) activity assays
- Immunoblotting experiments
- Cryo-EM sample preparation and data collection
- Image processing of cryo-EM data and signature identification
- Atomic model building
- Protein sequence identification

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Signature identification
  - FAS additional density detection and modelling
  - Signature 1 identification
  - Ribosomal subunit identification

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2022.01.001.

ACKNOWLEDGMENTS

We thank the members of the Kastritis laboratory for valuable discussions. This work was supported by the Federal Ministry of Education and Research (BMBF, ZIK program) (Grant nos. 03Z22HN23, 03Z22HI2 and 03COV04 to P.L.K.), the European Regional Development Funds (EFRE) for Saxony-Anhalt (Grant no. ZS/2016/04/78115 to P.L.K.), the Deutsche Forschungsgemeinschaft (project number 391498659, RTG 2467) and the Martin-Luther University of Halle-Wittenberg.

AUTHOR CONTRIBUTIONS

Conceptualization: P.L.K. Methodology: F.L.K., I.S., C.T., F.H., G.C., and PLK. Investigation: I.S., F.L.K., F.H., G.C., and PLK. Visualization: I.S., F.L.K., C.T., F.H., and PLK. Supervision: P.L.K. Writing—original draft: I.S., G.C., and P.L.K. Writing—review and editing: I.S., G.C., and PLK.

DECLARATION OF INTERESTS

Authors declare no competing interests.

Received: July 30, 2021
Revised: November 5, 2021
Accepted: January 3, 2022
Published: January 28, 2022

REFERENCES

Abbas, Y.M., Wu, D., Bueler, S.A., Robinson, C.V., and Rubinstein, J.L. (2020). Structure of V-ATPase from the mammalian brain. Science *367*, 1240–1246.

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

Ban, N., Beckmann, R., Cate, J.H., Dinman, J.D., Dragon, F., Ellis, S.R., Lafontaine, D.L., Lindahl, L., Liljas, A., Lipton, J.M., et al. (2014). A new system for naming ribosomal proteins. Curr. Opin. Struct. Biol. *24*, 165–169.

Baretic, D., Pollard, H.K., Fisher, D.I., Johnson, C.M., Santhanam, B., Truman, C.M., Kouba, T., Fersht, A.R., Phillips, C., and Williams, R.L. (2017). Structures of closed and open conformations of dimeric human ATM. Sci. Adv. *3*, e1700933.

Beaufay, H., Amar-Costesec, A., Thines-Sempoux, D., Wibo, M., Robbi, M., and Berthet, J. (1974). Analytical study of microsomes and isolated subcellular membranes from rat liver. 3. Subfractionation of the microsomal fraction by isopycnic and differential centrifugation in density gradients. J. Cell Biol. *61*, 213–231.

Beck, M., and Baumeister, W. (2016). Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? Trends Cell Biol *26*, 825–837, In press. https://doi.org/10.1016/j.tcb.2016.08.006.

Behrmann, E., Loerke, J., Budkevich, T.V., Yamamoto, K., Schmidt, A., Penczek, P.A., Vos, M.R., Burger, J., Mielke, T., Scheerer, P., and Spahn, C.M. (2015). Structural snapshots of actively translating human ribosomes. Cell *161*, 845–857.

Bepler, T., Kelley, K., Noble, A.J., and Berger, B. (2020). Topaz-Denoise: general deep denoising models for cryoEM and cryoET. Nat. Commun. *11*, 5208.

Bouvette, J., Liu, H.F., Du, X., Zhou, Y., Sikkema, A.P., da Fonseca Rezende, E.M.J., Klemm, B.P., Huang, R., Schaaper, R.M., Borgnia, M.J., and Bartesaghi, A. (2021). Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. Nat. Commun. *12*, 1957.

Braunger, K., Pfeffer, S., Shrimal, S., Gilmore, R., Berninghausen, O., Mandon, E.C., Becker, T., Forster, F., and Beckmann, R. (2018). Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. Science *360*, 215–219.

Cavadini, S., Fischer, E.S., Bunker, R.D., Potenza, A., Lingaraju, G.M., Goldie, K.N., Mohamed, W.I., Faty, M., Petzold, G., Beckwith, R.E., et al. (2016). Cullin-RING ubiquitin E3 ligase regulation by the COP9 signalosome. Nature *531*, 598–603.

Chojnowski, G., Pereira, J., and Lamzin, V.S. (2019). Sequence assignment for low-resolution modelling of protein crystal structures. Acta Crystallogr. D Struct. Biol. *75*, 753–763.

Chojnowski, G., Sobolev, E., Heuser, P., and Lamzin, V.S. (2021). The accuracy of protein models automatically built into cryo-EM maps with ARP/wARP. Acta Crystallogr. D Struct. Biol. *77*, 142–150.

Chojnowski, G., Simpkin, A.J., Leonardo, D.A., Seifert-Davila, W., Vivas-Ruiz, D.E., Keegan, R.M., and Rigden, D.J. (2022). findMySequence: a neural-network-based approach for identification of unknown proteins in X-ray crystallography and cryo-EM. IUCrJ *9*. https://doi.org/10.1107/S2052252521011088.

Ciccarelli, L., Connell, S.R., Enderle, M., Mills, D.J., Vonck, J., and Grininger, M. (2013). Structure and conformational variability of the mycobacterium tuberculosis fatty acid synthase multienzyme complex. Structure *21*, 1251–1257.

Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530.

Danev, R., Yanagisawa, H., and Kikkawa, M. (2019). Cryo-electron microscopy methodology: current aspects and future directions. Trends Biochem. Sci. *44*, 837–848.

Davis, J.H., Tan, Y.Z., Carragher, B., Potter, C.S., Lyumkis, D., and Williamson, J.R. (2016). Modular assembly of the bacterial large ribosomal subunit. Cell *167*, 1610–1622.e15.

De Duve, C., and Berthet, J. (1954). In The Use of Differential Centrifugation in the Study of Tissue Enzymes. International Review of Cytology, *3*, G.H. Bourne and J.F. Danielli, eds. (Academic Press), pp. 225–275.

de la Rosa-Trevin, J.M., Quintana, A., Del Cano, L., Zaldivar, A., Foche, I., Gutierrez, J., Gomez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., et al. (2016). Scipion: a software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. *195*, 93–99.

Dick, R.A., Xu, C., Morado, D.R., Kravchuk, V., Ricana, C.L., Lyddon, T.D., Broad, A.M., Feathers, J.R., Johnson, M.C., Vogt, V.M., et al. (2020). Structures of immature EIAV Gag lattices reveal a conserved role for IP6 in lentivirus assembly. PLoS Pathog. *16*, e1008277.

Efremov, R.G., and Stroobants, A. (2021). Coma-corrected rapid single-particle cryo-EM data collection on the CRYO ARM 300. Acta Crystallogr. D Struct. Biol. *77*, 555–564.

Ehrenreich, J.H., Bergeron, J.J., Siekevitz, P., and Palade, G.E. (1973). Golgi fractions prepared from rat liver homogenates. I. Isolation procedure and morphological characterization. J. Cell Biol. *59*, 45–72.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-multimer. bioRxiv. https://doi.org/10.1101/2021.2010.2004.463034.

Forsberg, B.O., Aibara, S., Howard, R.J., Mortezaei, N., and Lindahl, E. (2020). Arrangement and symmetry of the fungal E3BP-containing core of the pyruvate dehydrogenase complex. Nat. Commun. *11*, 4667.

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature *440*, 631–636.

Groothuizen, F.S., Winkler, I., Cristovao, M., Fish, A., Winterwerp, H.H., Reumer, A., Marx, A.D., Hermans, N., Nicholls, R.A., Murshudov, G.N., et al. (2015). MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA. Elife *4*, e06744.

Guo, R., Cang, Z., Yao, J., Kim, M., Deans, E., Wei, G., Kang, S.-g., and Hong, H. (2020). Structural cavities are critical to balancing stability and activity of a membrane-integral enzyme. Proc. Natl. Acad. Sci. U S A *117*, 22146–22156.

Han, B.G., Dong, M., Liu, H., Camp, L., Geller, J., Singer, M., Hazen, T.C., Choi, M., Witkowska, H.E., Ball, D.A., et al. (2009). Survey of large protein complexes in *D. vulgaris* reveals great structural diversity. Proc Natl Acad Sci U S A *29*, 16580–16585. https://doi.org/10.1073/pnas.0813068106.

Han, Y., Reyes, A.A., Malik, S., and He, Y. (2020). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. Nature *579*, 452–455.

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

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.

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., et al. (2017). Capturing protein communities by structural proteomics in a thermophilic eukaryote. Mol. Syst. Biol. *13*, 936.

Kastritis, P.L., and Gavin, A.C. (2018). Enzymatic complexes across scales. Essays Biochem. *62*, 501–514.

Kater, L., Thoms, M., Barrio-Garcia, C., Cheng, J., Ismail, S., Ahmed, Y.L., Bange, G., Kressler, D., Berninghausen, O., Sinning, I., et al. (2017). Visualizing the assembly pathway of nucleolar pre-60S ribosomes. Cell *171*, 1599–1610.e14.

Kato, M., Wynn, R.M., Chuang, J.L., Brautigam, C.A., Custorio, M., and Chuang, D.T. (2006). A synchronized substrate-gating mechanism revealed by cubic-core structure of the bovine branched-chain alpha-ketoacid dehydrogenase complex. EMBO J. *25*, 5983–5994.

Kühlbrandt, W. (2014a). Cryo-EM enters a new era. Elife *3*, e03678.

Kühlbrandt, W. (2014b). The resolution revolution. Science *343*, 1443–1444.

Kuhner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., et al. (2009). Proteome organization in a genome-reduced bacterium. Science *326*, 1235–1240.

Kyrilis, F.L., Meister, A., and Kastritis, P.L. (2019). Integrative biology of native cell extracts: a new era for structural characterization of life processes. Biol. Chem. *400*, 831–846.

Kyrilis, F.L., Belapure, J., and Kastritis, P.L. (2021a). Detecting protein communities in native cell extracts by machine learning: a structural biologist's perspective. Front. Mol. Biosci. *8*, 660542.

Kyrilis, F.L., Semchonok, D.A., Skalidis, I., Tuting, C., Hamdi, F., O'Reilly, F.J., Rappsilber, J., and Kastritis, P.L. (2021b). Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. Cell Rep. *34*, 108727.

Leighton, F., Poole, B., Beaufay, H., Baudhuin, P., Coffey, J.W., Fowler, S., and De Duve, C. (1968). The large-scale separation of peroxisomes, mitochondria, and lysosomes from the livers of rats injected with triton WR-1339. Improved isolation procedures, automated analysis, biochemical and morphological properties of fractions. J. Cell Biol. *37*, 482–513.

Liang, X., Zuo, M.-Q., Zhang, Y., Li, N., Ma, C., Dong, M.-Q., and Gao, N. (2020). Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. Nat. Commun. *11*, 3542.

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkoczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr. D Struct. Biol. *75*, 861–877.

Lou, J.W., Iyer, K.R., Hasan, S.M.N., Cowen, L.E., and Mazhab-Jafari, M.T. (2019). Electron cryomicroscopy observation of acyl carrier protein translocation in type I fungal fatty acid synthase. Sci. Rep. *9*, 12987.

Maco, B., Ross, I.L., Landsberg, M.J., Mouradov, D., Saunders, N.F., Hankamer, B., and Kobe, B. (2011). Proteomic and electron microscopy survey of large assemblies in macrophage cytoplasm. Mol Cell Proteomics *10*. https://doi.org/10.1074/mcp.M111.008763.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636–W641.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2021). ColabFold - making protein folding accessible to all. bioRxiv. https://doi.org/10.1101/2021.2008.2015.456425.

Nagy, B., Polak, M., Ozohanics, O., Zambo, Z., Szabo, E., Hubert, A., Jordan, F., Novaček, J., Adam-Vizi, V., and Ambrus, A. (2021). Structure of the dihydrolipoamide succinyltransferase (E2) component of the human alpha-ketoglutarate dehydrogenase complex (hKGDHc) revealed by cryo-EM and cross-linking mass spectrometry: implications for the overall hKGDHc structure. Biochim. Biophys. Acta Gen. Subj. *1865*, 129889.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehling, J., et al. (2020). Single-particle cryo-EM at atomic resolution. Nature *587*, 152–156.

O'Reilly, F.J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., Blötz, C., Singh, N., Hagen, W.J.H., Cramer, P., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. Science *369*, 554–557.

Palade, G.E., and Siekevitz, P. (1956). Liver microsomes; an integrated morphological and biochemical study. J. Biophys. Biochem. Cytol. *2*, 171–200.

Passmore, L.A., and Russo, C.J. (2016). Specimen preparation for high-resolution cryo-EM. Methods Enzymol. *579*, 51–86.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat. Methods *17*, 1214–1221.

Punjani, A., and Fleet, D.J. (2021). 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. J. Struct. Biol. *213*, 107702.

Saibil, H.R. (2000). Conformational changes studied by cryo-electron microscopy. Nat. Struct. Biol. *7*, 711–714.

Sanders, C.R., and Hutchison, J.M. (2018). Membrane properties that shape the evolution of membrane enzymes. Curr. Opin. Struct. Biol. *51*, 80–91.

Shen, P.S., Park, J., Qin, Y., Li, X., Parsawar, K., Larson, M.H., Cox, J., Cheng, Y., Lambowitz, A.M., Weissman, J.S., et al. (2015). Protein synthesis. Rqc2p

and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. Science *347*, 75–78.

Singh, K., Graf, B., Linden, A., Sautner, V., Urlaub, H., Tittmann, K., Stark, H., and Chari, A. (2020). Discovery of a regulatory subunit of the yeast fatty acid synthase. Cell *180*, 1130–1143.e20.

Skalidis, I., Tuting, C., and Kastritis, P.L. (2020). Unstructured regions of large enzymatic complexes control the availability of metabolites with signaling functions. Cell Commun. Signal. *18*, 136.

Spitzer, M., Wildenhain, J., Rappsilber, J., and Tyers, M. (2014). BoxPlotR: a web tool for generation of box plots. Nat. Methods *11*, 121–122.

Srere, P.A. (1985). The metabolon. Trends Biochem. Sci. *10*, 109–110.

Su, C.C., Lyu, M., Morgan, C.E., Bolla, J.R., Robinson, C.V., and Yu, E.W. (2021). A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. Nat. Methods *18*, 69–75.

Suzuki, H., Kawabata, T., and Nakamura, H. (2015). Omokage search: shape similarity search service for biomolecular structures in both the PDB and EMDB. Bioinformatics *32*, 619–620.

Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods *16*, 1146–1152.

Tüting, C., Kyrilis, F.L., Muller, J., Sorokina, M., Skalidis, I., Hamdi, F., Sadian, Y., and Kastritis, P.L. (2021). Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. Nat. Commun. *12*, 6933.

UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489.

Verbeke, E.J., Mallam, A.L., Drew, K., Marcotte, E.M., and Taylor, D.W. (2018). Classification of single particles from human cell extract reveals distinct structures. Cell Rep. *24*, 259–268.e3.

Wang, X., Chu, H., Lv, M., Zhang, Z., Qiu, S., Liu, H., Shen, X., Wang, W., and Cai, G. (2016). Structure of the intact ATM/Tel1 kinase. Nat. Commun. *7*, 11655.

Webb, B., and Sali, A. (2016). Comparative protein structure modeling using MODELLER. Curr. Protoc. Bioinformatics *54*, 5.6.1–5.6.37.

Wieczorek, M., Urnavicius, L., Ti, S.C., Molloy, K.R., Chait, B.T., and Kapoor, T.M. (2020). Asymmetric molecular architecture of the human gamma-tubulin ring complex. Cell *180*, 165–175.e16.

Wu, S., Tutuncuoglu, B., Yan, K., Brown, H., Zhang, Y., Tan, D., Gamalinda, M., Yuan, Y., Li, Z., Jakovljevic, J., et al. (2016). Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. Nature *534*, 133–137.

Yip, K.M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. Nature *587*, 157–161.

Zhang, W., Tarutani, A., Newell, K.L., Murzin, A.G., Matsubara, T., Falcon, B., Vidal, R., Garringer, H.J., Shi, Y., Ikeuchi, T., et al. (2020). Novel tau filament fold in corticobasal degeneration. Nature *580*, 283–287.

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife *7*, e42166.

Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253–267.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal antibody a-E2o against *C. thermophilum* E2o-His-Tag (39-420)	(Kyrilis et al., 2021b) (custom-made, Genscript)	N/A
Rabbit polyclonal antibody a-E2p against *C. thermophilum* E2p-His-Tag (29-459)	(Kyrilis et al., 2021b) (custom-made, Genscript)	RRID: AB_2888985
Chemicals, peptides, and recombinant proteins		
C. thermophilum E2o-His-Tag (39-420)	(Kyrilis et al., 2021b) (custom-made, Genscript)	N/A
C. thermophilum E2p-His-Tag (29-459)	(Kyrilis et al., 2021b) (custom-made, Genscript)	N/A
1,4-Dithiothreit, min. 99%, p.a.	Carl Roth	Cat#6908.4
Acrylamide/Bis solution, 37.5:1	Serva	Cat#10688.01
Agar-Agar, bacteriological highly pure	Carl Roth	Cat#2266.3
Ammonium acetate, ≥97%, p.a., ACS	Carl Roth	Cat#7869.2
Ammonium persulfate	Serva	Cat#13376.02
Aprotinin from bovine lung	Merck	Cat#A1153-1MG
Bestatin, 10MG	Merck	Cat#10874515001
Clarity Western ECL substrate	BIO-RAD	Cat#170-5060
D-Sucrose, ≥99,5%, p.a.	Carl Roth	Cat#4621.1
D(+)-Glucose p. a., ACS, anhydrous	Carl Roth	Cat#X997.2
Dextrin for microbiology (from potato starch)	Carl Roth	Cat#3488.1
di-Potassium hydrogen phosphate trihydrate	Carl Roth	Cat#6878.1
di-Potassium hydrogen phosphate, ≥99%, p.a., anhydrous	Carl Roth	Cat#P749.1
DNAse I	Merck	Cat#10104159001
E-64	Merck	Cat#E3132-1MG
EDTA disodium salt dihydrate, min. 99%, p.a., ACS	Carl Roth	Cat#8043.2
Glycine	Serva	Cat#23391.02
HEPES PUFFERAN®, min. 99.5%, p.-1 kg	Carl Roth	Cat#9105.3
Iron (III) sulphate hydrate, 80%, pure	Carl Roth	Cat#0492.1
Isopropanol	Carl Roth	Cat#CP41.1
Leupeptin	Merck	Cat#L2884-1MG
Magnesium chloride hexahydrate, min. 99%, p.a., ACS	Carl Roth	Cat#2189.1
Magnesium sulphate heptahydrate, ≥99%, p.a., ACS	Carl Roth	Cat#P027.1
Methanol	Carl Roth	Cat#4627.6
Milk powder	Carl Roth	Cat#T145.3
Pefabloc	Merck	Cat#11585916001
Pepstatin A	Merck	Cat#77170-5MG
Peptone ex casein	Carl Roth	Cat#8986.1
Phosphate buffered saline tablets (PBS)	Merck	Cat#P4417
Potassium chloride min. 99.5%, -1 kg	Carl Roth	Cat#6781.1
Potassium dihydrogen phosphate, ≥99%, p.a., ACS	Carl Roth	Cat#3904.2
Precision plus protein all blue standards (marker)	BIO-RAD	Cat#161-0373
Sodium chloride 99,5%, p.a., ACS, ISO	Carl Roth	Cat#3957.2
Sodium dodecyl sulfate (SDS)	Carl Roth	Cat#0183.2
Sodium nitrate, ≥99%, p.a., ACS, ISO	Carl Roth	Cat#A136.1
TEMED	Carl Roth	Cat#2367.3
Tris	Carl Roth	Cat#AE15.2

(*Continued on next page*)

Continued

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Tris Hydrochloride	Carl Roth	Cat#9090.2
Tween 20	Carl Roth	Cat#9127.1
Tryptone	Merck	Cat#T7293
Yeast Extract, micro-granulated	Carl Roth	Cat#2904.3
Critical commercial assays		
A-KG Dehydrogenase activity assay kit	Merck	Cat#MAK189-1KT
PDH Activity assay kit	Merck	Cat#MAK183-1KT
Deposited data		
Cryo-EM map of *C. thermophilum* Oxoglutarate Dehydrogenase Complex E2 core	This Study	EMDB: EMD-13844
Refined model of *C. thermophilum* Oxoglutarate Dehydrogenase Complex E2 core	This Study	PDB: 7Q5Q
Cryo-EM map of *C. thermophilum* Pyruvate Dehydrogenase Complex E2 core	This Study	EMDB: EMD-13845
Refined model of *C. thermophilum* Pyruvate Dehydrogenase Complex E2 core	This Study	PDB: 7Q5R
Cryo-EM map of *C. thermophilum* Fatty Acid Synthase	This Study	EMDB: EMD-13846
Refined model of *C. thermophilum* Fatty Acid Synthase	This Study	PDB: 7Q5S
Cryo-EM map of *C. thermophilum* Immature 60S Ribosomal Subunit	This Study	EMDB: EMD-13093
Cryo-EM dataset of *C. thermophilum* MDa-range protein communities	This Study	EMPIAR: EMPIAR-10892
C. thermophilum Pyruvate Dehydrogenase Complex Core	(Kyrilis et al., 2021b)	PDB: 7BGJ
Structure of fatty acid synthase complex with bound gamma subunit from *Saccharomyces cerevisiae* at 2.8 angstrom	(Singh et al., 2020)	PDB: 6QL5
Structure of fatty acid synthase complex with bound gamma subunit from *Saccharomyces cerevisiae* at 2.8 angstrom	(Singh et al., 2020)	EMDB: EMDB-4577
Cryo-electron microscopic structure of the dihydrolipoamide succinyltransferase (E2) component of the human alpha-ketoglutarate (2-oxoglutarate) dehydrogenase complex [residues 218-453]	(Nagy et al., 2021)	EMDB: EMDB-0108
Crystal structure of a cubic core of the dihydrolipoamide acyltransferase (E2b) component in the branched-chain alpha-ketoacid dehydrogenase complex (BCKDC), apo form	(Kato et al., 2006)	PDB: 2IHW
Cryo-EM structure of a pre-60S ribosomal subunit - state preA	(Liang et al., 2020)	PDB: 6LSS
Cryo-EM structure of eukaryotic pre-60S ribosomal subunits	(Wu et al., 2016)	PDB: 3JCT
Electron cryomicroscopy Structure of *Candida albicans* FAS in the Apo state	(Lou et al., 2019)	PDB: 6U5V
MutS in complex with the N-terminal domain of MutL - crystal form 2	(Groothuizen et al., 2015)	PDB: 5AKC
Human dimeric ATM kinase	Unpublished	EMDB: EMD-9523
A structure of the EIAV CA-SP hexamer (C6) from Gag-deltaMA spheres assembled at pH6	(Dick et al., 2020)	EMDB: EMD-10384
Structure of the intact ATM/Tel1 kinase	(Wang et al., 2016)	EMDB: EMD-6399
SWI/SNF nucleosome complex with ADP-BeFx	(Han et al., 2020)	EMDB: EMD-20934
Density map of lumenal bridge in reconstruction of the native human gamma-tubulin ring complex	(Wieczorek et al., 2020)	EMDB: EMD-21069
Structure of bL17-depleted large ribosomal subunit assembly intermediate - Class D2	(Davis et al., 2016)	EMDB: EMD-8447
Density map surrounding actin-like protein in reconstruction of the native gamma-tubulin ring complex	(Wieczorek et al., 2020)	EMDB: EMD-21063

(*Continued on next page*)

Continued

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Closed dimer (C2) of human ATM (Ataxia telangiectasia mutated)	(Baretic et al., 2017)	EMDB: EMD-3668
Cryo-EM structure of CSN-N8-CRL4ADDB2 at 8.3 A resolution	(Cavadini et al., 2016)	EMDB: EMD-3316
Experimental models: Organisms/strains		
Chaetomium thermophilum var. thermophilum (Freeze Dried)	DSMZ	DSM 1495
Software and algorithms		
Image Lab Software 6.1	BIO-RAD	https://www.bio-rad.com/de-de/product/image-lab-software
UNICORN 7 Workstation for ÄKTA pure, pilot, process, Ready to Process WAVE 25	GE Healthcare Europe GmbH	https://www.gelifesciences.com/en/us/shop/chromatography/software/unicorn-7-p-05649
UCSF ChimeraX	(Pettersen et al., 2021)	https://www.rbvi.ucsf.edu/chimerax/
Omokage Search	(Suzuki et al., 2015)	https://pdbj.org/emnavi/omo-search.php
RELION	(Zivanov et al., 2018)	https://github.com/3dem/relion
PHENIX	(Liebschner et al., 2019)	https://www.phenix-online.org
Pymol	Schrödinger, inc	https://pymol.org/
cryoSPARC	Structura Biotechnology Inc.	https://cryosparc.com/
Topaz	(Bepler et al., 2020)	https://emgweb.nysbc.org/topaz.html
MODELLER	(Webb and Sali, 2016)	https://salilab.org/modeller/
COOT	(Emsley and Cowtan, 2004)	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
AlphaFold2	(Jumper et al., 2021)	https://github.com/deepmind/alphafold
ColabFold	(Mirdita et al., 2021)	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
FindMySequence	(Chojnowski et al., 2022)	https://gitlab.com/gchojnowski/findmysequence
ARP/wARP	(Chojnowski et al., 2021)	https://www.embl-hamburg.de/ARP/
BoxPlotR	(Spitzer et al., 2014)	http://shiny.chemgrid.org/boxplotr/
Data Analysis ToolPak/MS Excel	Microsoft Corporation	https://office.microsoft.com/excel

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jun. Prof. Dr. Panagiotis L. Kastritis (panagiotis.kastritis@bct.uni-halle.de).

Materials availability

All unique/stable reagents generated in this study will be made available on request. A payment and/or a completed Materials Transfer Agreement may be required if there is potential for commercial application.

Data and code availability

- The cryo-EM map of the pre-60S ribosomal subunit is deposited in EMDB. Reconstructed cryo-EM maps and models corresponding to the *C. thermophilum* metabolon-embedded oxoglutarate dehydrogenase complex core, the pyruvate dehydrogenase complex core, and the fatty acid synthase complex are deposited in EMDB and PDB. Primary movies and their motion-corrected, averaged micrographs used for analysis are deposited in EMPIAR. All deposited data are publicly available as of the date of publication and all corresponding accession numbers are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Model organism

Model organism *Chaetomium thermophilum* var. *thermophilum* La Touche 1950 (DSM No.: 1495, Type strain) was obtained from DSMZ (Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH), preserved in freeze-dried ampoules, then cultivated according to company guidelines (Medium 188 DSMZ Media List, 45°C). Mycelium propagation took place in CCM media that included 0.50 g NaCl, 3.00 g sucrose, 0.50 g $MgSO_4$ x 7 H_2O, 0.01 g $Fe_2(SO_4)_3$ x H_2O, 0.65 g K_2HPO_4 x 3 H_2O, 1.00 g peptone, 5.00 g tryptone, 1.00 g yeast extract, 15.00 g dextrin per 1000 mL of deionized water, adjusting the pH to 7.1. For solid cultures, 15.00 g of Agar per 1000 mL of deionized water was added to liquid CCM, and freshly prepared plates were inoculated with mycelium and subsequently grown at 45°C supplied with 10% CO_2. For liquid CCM cultures, 2000 mL Erlenmeyer flasks containing 800 mL of media were inoculated with small pieces of mycelium from the freshly grown Agar plates and then incubated for 20 hours under shaking at 110 rpm.

METHOD DETAILS

Mycelium harvesting, culture lysis and sample preparation

Briefly, propagated mycelium was harvested by straining cultures through a 180 μm pore size metal sieve, washed with PBS and then freeze-ground with a pre-chilled mortar and stored at -80°C. For final sample preparation, 8 g of ground mycelium were lysed in 20 mL of Lysis Buffer containing 100 mM HEPES pH 7.4, 95 mM NaCl, 5mM KCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 5% Glycerol, 1 mM DTT, 10 μg mL^{-1} DNAse, Pefabloc 2.5 mM, E-64 40 μM, Bestatin 130 μM, Aprotinin 0.5 μM, Leupeptin 1 μM, Pepstatin A 60 μM, with a Fastprep cell homgenizer at 4°C, 3 repeats of 6.5 mps shaking speed for 25 seconds with in between 3 mins rest on ice. Large cell debris is pelleted with a 4000 *g* centrifugation step. Finally, a high-speed centrifugation of 100,000 *g* is performed, the supernatant is filtered with a 100,000 Da cut-off centrifugal filter and then concentrated to a final protein concentration of 30 mg mL^{-1}.

Size exclusion chromatography

The final sample, with an end concentration of 30 mg mL^{-1} was injected with a 500 μL loop, in a Biosep SEC-S4000 size exclusion column, mounted on an ÄKTA Pure 25M (FPLC) system. A filtered, degassed buffer (200 mM, $CH_3COO^-NH_4$, pH 7.4) was used to equilibrate the column before applying the sample. Flow rate was regulated at 0.15 mL.min^{-1} and fraction volume to 250 μL. After fractionation, fraction 6 was selected for further analysis.

Pyruvate dehydrogenase (PDC) and oxoglutarate dehydrogenase (OGHD) activity assays

Pyruvate Dehydrogenase (Catalog Number MAK183) and α-ketoglutarate Dehydrogenase (oxoglutarate dehydrogenase, OGDH) (Catalog Number MAK189) kinetic activity assay kits were obtained from Merck® and applied to fraction 6. Included instructions from the provider were followed and a coupled enzyme reaction which results to a measurable colorimetric product (450 nm) equivalent to enzyme activity present was applied in a technical triplicate for the detection and quantification of in-fraction activity of PDH and OGDH respectively. In more detail, 2 μL of fraction 6 were used for the PDH and OGDH activity assays. A lower molecular weight fraction (fraction 22) served as a negative control and included in kit positive controls were used in both assays. Standards were calculated by adding 0, 2, 4, 6, 8 and 10 μL of the 1,25 mM NADH Standard Solution (provided in kit) in duplicates in a 96-well plate, resulting in 0, 2.5, 5, 7.5, 10 and 12.5 nM per well standards. Activity calculations and background subtraction were performed, and the final corrected measurements were used to calculate the generated NADH in μmole/min/mL according to the included instructions.

Immunoblotting experiments

For the Western Blot experiments, in-house gels were prepared. For the separating phase of the gel, 3.34 mL Acrylamide 30% $^{w}/_{v}$, 2.46 mL 1.5 M Tris-HCl pH 8.8, 50 μL 20% sodium dodecyl sulfate (SDS), 40 μL 10% APS, 20 μL TEMED and 4.1 mL of ddH_2O and for the stacking phase 850 μL Acrylamide 30% $^{w}/_{v}$, 1.25 mL 0.5 M Tris-HCl pH 6.8, 25 μL 20% SDS, 20 μL 10% APS, 10 μL TEMED and 2.86 mL of ddH_2O were used. Final gel thickness of the freshly prepared gels was 1 mm. After mixing with a 4X loading dye (250 mM Tris–HCl (pH 6.8), 8% $^{w}/_{v}$ SDS, 0.2% $^{w}/_{v}$ bromophenol blue, 40% $^{v}/_{v}$ glycerol, 20% $^{v}/_{v}$ β-mercaptoethanol), samples were incubated for 5 mins at 100°C. In each lane, 5 μL of final volume were loaded for every sample along with 5 μL of Precision Plus Protein™ All Blue Prestained Protein Standards (Biorad #1610373). The loaded polyacrylamide gels were subjected to a 100V electrical field, partly submerged in 1X electrophoresis buffer (10X stock: 30.3 g Tris-base, 144 g Glycine in 1 L of deionized water) for 2 hours. A Trans-Blot® Turbo™ Transfer System (Biorad) was used to transfer gel contents to a nitrocellulose membrane with a preset protocol (25 V, 30 mins) and the membranes were blocked with 5% $^{w}/_{v}$ milk/TBST solution for one hour and constant stirring. Directly after blocking, membranes were incubated for 16 h with the primary antibody diluted in 2% $^{w}/_{v}$ milk/TBST solution, and then washed 3 times with 2% $^{w}/_{v}$ milk/TBST solution. Finally, membranes were incubated with the secondary antibody (in 2% $^{w}/_{v}$ milk/TBST solution) for 1 h, washed again 3 times, ECL fluorescent mixture was applied and imaged in a ChemiDoc MP Imaging system (Biorad), using optimized exposure times.

Cryo-EM sample preparation and data collection

Quantifoil® type R2/1 holey carbon-coated support films on copper 200 mesh copper grids were used in order to analyze samples with cryo-EM. A PELCO easiGlow™ was used to glow discharge the grids at 0.4 mbar, grid negative, 15 mA and a glowing time of 25 sec. Sample volume that was applied on each grid was 3.5 μL with a protein concentration of 0.3 mg mL^{-1}. Grids were subsequently plunge-frozen with a Vitrobot® Mark IV (Thermo Fisher Scientific), after being blotted with Vitrobot Filter Paper (ø55/20 mm and ash-free Grade 595 filter paper), with constant vitrification chamber conditions of 95% humidity at 4°C temperature. Vitrification parameters were adjusted to blot time of 6 s and blot force 2. After clipping, the grids were loaded under cryo, low humidity conditions on TFS (Thermo Fisher Scientific) Glacios Cryo-TEM (cryo-transmission electron microscope), equipped with a Falcon 3EC direct electron detector for data collection. A low dose regime was applied for data acquisition, in linear mode and 30 $e^{-}/Å^2$ total dose. Samples were illuminated with a parallel beam, perpendicular to the sample with a diameter of 2.5 μm diameter and a 100 μm objective aperture inserted to restrict the objective angle. Table S1 lists the acquisition parameters.

Image processing of cryo-EM data and signature identification

The cryoSPARC 3.1 high-performance computing software (Punjani et al., 2017) was used for all steps of image processing using the software's incorporated methods. A dataset of 2808 raw movies was imported and then corrected for motion with the Patch motion correction (multi) auto-tuning correction method. CTF parameters for the motion-corrected micrographs were then estimated with the Patch CTF estimation (multi) method. A deep neural network Blob picker was employed to perform reference-free particle picking, resulting in an initial set of 276399 single particles. These were subsequently used for reference-free 2D classification in 128 classes. Obvious junk classes were discarded and the rest were reclassified. By repeating the process for two more rounds with the same parameters, 4 prominent 2D structural signatures emerged and were then used as references for Template picking. After template picking, all signatures were reconstructed *ab-initio* in 3D, using 2 classes for Signatures 1 and 4, 5 classes for Signature 2 and 3 classes for Signature 3. The best ab-initio 3D reconstructions contained 1819 particles for Signature 1, 2582 particles for Signature 2, 3331 particles for Signature 3 and 20279 particles for Signature 4. These four structural signatures were used as queries for the Omokage search webserver (Suzuki et al., 2015), identifying them as the oxoglutarate dehydrogenase complex core (OGDHc, Signature 1), pyruvate dehydrogenase complex core (PDHc, Signature 2), fatty acid synthase complex (FAS, Signature 3) and the pre-60S ribosomal subunit (pre-60S, Signature 4) respectively. The 4 *ab-initio* models were projected in 2D and used for further template picking rounds to improve particle abundance. Final particle sets were cleaned after multiple 2D classifications, choosing optimal classes manually, guided by the 2D projections of the *ab-initio* maps, and discarding classes containing junk particles, ending up with 1819 particles for OGDHc, 7825 particles for PDHc, 5231 particles for FAS and 35773 particles for the pre-60S. OGDHc, PDHc and FAS maps were reconstructed and refined with the Homogeneous refinement (new) method, using the dynamic masking option. The pre-60S map was calculated and refined using the Local refinement method, employing the non-uniform refinement (Punjani et al., 2020) and dynamic masking options. Final maps reached a resolution of 4.38 Å (FSC=0.143) for the OGDHc, 3.84 Å (FSC=0.143) for the PDHc, 4.47 Å (FSC=0.143) for the FAS and 4.52 Å (FSC=0.143) for the pre-60S maps respectively. Local resolution estimations for all maps were performed with the Local resolution estimation method.

Atomic model building

Atomic models of pyruvate dehydrogenase (PDHc), oxoglutarate dehydrogenase (OGDHc), and Fatty Acid Synthase (FAS) complexes were built based on structures predicted using AlphaFold2 (Jumper et al., 2021) within ColabFold (Mirdita et al., 2021) advanced notebook that allows modelling multimeric structures. Default parameters and the final-model AMBER force-field relaxation option were used. Resulting models were fitted into corresponding EM reconstructions using "Jiggle-Fit this molecule with Fourier Filter" tool and refined in real space with all-molecule self-restraints at 5 Å distance cut-off using COOT version 0.9.2-pre (Emsley and Cowtan, 2004). Representative monomers were further extracted from the models and expanded into complete complexes using symmetry operators extracted directly from EM reconstructions using tools implemented in PHENIX (phenix.map_symmetry and phenix.apply_ncs) (Liebschner et al., 2019). The final models were refined using phenix.real_space_refine using starting model and NCS restraints.

For the PDHc and OGDHc structures, AlphaFold2 predictions of the complete trimeric sub-complexes were obtained. Steric conflicts between side-chain moieties identified after symmetry-expansion of the complex were refined manually using ChimeraX (Pettersen et al., 2021) and ISOLDE version 1.2.5 (Croll, 2018) for a representative monomer that was further symmetry-expanded to a complete complex and refined as described above.

The Fatty Acid Synthase (FAS) heterodimer was split into 6 overlapping fragments that were then independently fitted, refined into the EM reconstruction, and merged using COOT as described above. During model assembly, a related FAS structure from *C. albicans* (PDB ID: 6U5V (Lou et al., 2019)) was used for guidance. The resulting complex model's steric clashes were resolved in ISOLDE in a representative heterodimer, which was further symmetry-expanded into a complete complex and refined as described above.

Protein sequence identification

To resolve ambiguity between E2o and E2b for Signature 1 model reconstruction (sequence identity of 22%), both homotrimeric structures were predicted with AlphaFold2 using the ColabFold advanced notebook. Resulting models were fitted into EM reconstruction using "Jiggle-Fit this molecule with Fourier Filter" tool and refined in real space with all-molecule self-restraints at 5Å

distance cut-off using COOT version 0.9.2-pre. Finally, the model's backbones were used as input to the findMySequence (Chojnowski et al., 2022) software to identify sequences in *C. thermophilum* proteome (taxonomic identifier 759272). The E2o sequence was selected based on calculated E-value. A similar analysis for FAS was performed and a sequence of both chains built into a map using findMySequence was unambiguously confirmed (E-values 6.1e-180 and 3.4e-93 for α- and β-subunits respectively). Sequence identification of the third reconstruction of the pyruvate dehydrogenase complex was also pursued. In this case, a relatively high resolution of the reconstruction of 3.8 Å allowed *de novo* interpretation of the map and an optimized modelling approach. For a manually selected map region corresponding to a trimeric subcomplex, a mainchain-only model was automatically built using ARP/wARP (Chojnowski et al., 2021) with a sequence-independent loop building algorithm (Chojnowski et al., 2019).

QUANTIFICATION AND STATISTICAL ANALYSIS

Signature identification

The top-10 hits for all signatures were compared to the bottom-10 hits (N=20, number of cross-correlation scores) from the first 100 hits returned by the Omokage search (Suzuki et al., 2015) with single factor analysis of variance (ANOVA) (P=0.05) with the Analysis ToolPak in Microsoft Excel and then visually inspected. All resulting *P*-values for Signatures 1,2, and 3 that were compared are listed in Figure 2 legend. Means and variance for the top-10 and bottom-10 groups for each Signature are as follows: Signature 1: 0.79026/0.69987 (top-10/bottom-10, means), 0.000172607/6.47789E-06 (top-10/bottom-10, variance); Signature 2: 0.74119/0.59715 (top-10/bottom-10, means), 0.00074327/4.48056E-06 (top-10/bottom-10, variance); Signature 3: 0.84787/0.70323 (top-10/bottom-10, means), 0.000215289/9.49E-07 (top-10/bottom-10, variance); As each time only 2 groups were compared, no *post hoc* test was performed. All boxplots were generated with BoxPlotR (Spitzer et al., 2014). In the case of Signature 4, due to the heterogeneity of the returned hits, top-10 hits were also projected in 2D in RELION 3.0 (Zivanov et al., 2018), comprising of: *E. coli* MutS (Groothuizen et al., 2015), *H. sapiens* ATM kinase ((EMD-9523) and (Baretic et al., 2017)), equine infectious anemia virus EIAV CA-SP hexamer (Dick et al., 2020), *S. pombe* ATM/Tel1 (Wang et al., 2016), *S. cerevisiae* SWI/SNF complex (Han et al., 2020), *H. sapiens* γ-tubulin ring complex (Wieczorek et al., 2020), *E. coli* pre-60S (Davis et al., 2016), *H. sapiens* CSN-N8-CRL4ADDB2 complex (Cavadini et al., 2016), along with the *ab-initio* reconstructed Signature 4 map and then visually compared. All model fits were performed in ChimeraX 1.1 (Pettersen et al., 2021).

FAS additional density detection and modelling

For detection of additional densities, PDB ID: 7BGJ (Kyrilis et al., 2021b) was downloaded and using the molmap command, an electron density map was created at 4.5 Å, which was then fitted into the PDHc E2 core map (at contour level 0.33 and 1.15 respectively). Additionally, EMDB-3757 (Kastritis et al., 2017) was downloaded (contour level 0.06) and fitted in the reconstructed FAS map (contour level 0.69). Modelling of the PPT domain of the *C.thermophilum* FAS was performed as follows: Residues 1000-1900 of Uniprot ID: G0S867 (UniProt, 2021) were aligned with Clustal Omega (Madeira et al., 2019) to the *S. cerevisiae* model PDB ID: 6QL5 (Singh et al., 2020), a homology model was generated with MODELLER 10.1 (Webb and Sali, 2016) and was then real-space refined in Phenix (Liebschner et al., 2019) and fitted in the reconstructed FAS map (contour level 0.85). EMDB-4577 (Singh et al., 2020) (contour level 0.01) was also fitted in the reconstructed FAS map along with the PPT domain of PDB ID: 6QL5 and then compared.

Signature 1 identification

EMDB-0108 (Nagy et al., 2021) was lowpass filtered to 4.38 Å (contour level 0.71) and an electron density map was simulated from PDB ID: 2IHW (Kato et al., 2006) at the same resolution (contour level 0.36). Both maps were fitted in the reconstructed OGDHc E2 core map (contour level 0.825) 100 times (N=161, total number of unique cross-correlation scores of fits), checking for statistically significant difference between the 2 fits, by comparing the two groups with single-factor ANOVA (P=0.05), with the Analysis ToolPak in Microsoft Excel. Resulting P-value for the comparison is listed in Figure 6 legend. Means and variance for each group were as follows: 0.676125/0.642242353 (vs EMD-0108/vs 2IHW means) and 0.067526/0.053296562 (vs EMD-0108/vs 2IHW variance). As only 2 groups were compared, no *post hoc* test was performed. All boxplots were generated with BoxPlotR (Spitzer et al., 2014).

Ribosomal subunit identification

For the identification of the ribosomal subunits included in the reconstructed pre-60S map (contour level 0.34), models PDB ID: 6LSS (Liang et al., 2020) and PDB ID: 3JCT (Wu et al., 2016) were fitted in the map. For visualizing the pre-60S, a superposition in ChimeraX via comparison to a bacterial pre-60S (Davis et al., 2016) and a complete yeast 60S subunit (Shen et al., 2015) was performed. All ribosomal subunits that were included in the pre-60S map, along with the rRNA, were isolated, combined in a new model and then a simulated electron density map was created at a resolution of 4.52 Å.