Cryo-electron microscopy studies of native cell extracts – Elucidating an active pyruvate dehydrogenase complex from *Chaetomium thermophilum*

Dissertation

for the degree of
Doctor of Natural Sciences (Dr. rer. Nat.)

Submitted to the
Faculty of Natural Sciences I – Biosciences –
of Martin Luther University,
Halle-Wittenberg

Presented by:
Fotios L. Kyrilis
Born on 01.06.1984 in Agrinio (Greece)

Assessors:
Jun. Prof. Dr. Panagiotis L. Kastritis
Prof. Dr. Jochen Balbach
Prof. Dr. Christos Gatsogiannis

Submitted on: 01.03.2022 Defended on: 29.06.2022

To my family

and my parents

Table of contents

Table of figures

List of tables

List of abbreviations

Ø	diameter
2D	two dimensional
3D	three dimensional
Å	Ångström
aa	amino acid(s)
A_{280}	Absorbance at 280 nm
Acetyl-CoA	acetyl coenzyme A
AP-MS	Affinity purification mass spectrometry
ATP	Adenosine triphosphate
BCKDHc	Branched-chain keto-acid dehydrogenase complex
BSA	Bovine serum albumin
C-ter	carboxyl-terminus, carboxy-terminus, C-terminal tail, C-terminal end, or COOH-terminus
CAPRI	Critical assessment of protein-protein interactions
CASP	Critical assessment of protein structure prediction
CCM	name of the medium used for *Chaetomium thermophilum* culture
CNNs	Convolutional neural networks
CoASH	Coenzyme A
cryo-CLEM	cryo correlative light and electron microscopy
cryo-EM	cryo electron microscopy
cryo-ET	Cryo-electron tomography, electron cryo-tomography
CTF	Contrast transfer function
C. thermophilum or *Ct* or *ct*	*Chaetomium thermophilum*
ddH2O	double-distilled water
DOL	Division-of-labour mechanism for oxo-acid dehydrogenase complexes
DTT	Dithiothreitol
E1p/o/b	E1 keto acid dehydrogenase corresponding to pyruvate (p) /2-oxoglutarate (o) /branched-chain (b), with ECs 1.2.4.1, 1.2.4.2, and 1.2.4.4, respectively
E2p/o/b	dihydrolipoyl (or dihydrolipoamide) transacetylase/transsuccinylase (or acetyl/succinyltransferase)) of pyruvate (p), 2-oxoglutarate (o) and branched-chain (b), with ECs 2.3.1.12, 2.3.1.61, and 2.3.1.168, respectively
E3	Dihydrolipoyl dehydrogenase
E3BP	E3 binding protein
EBI	European Bioinformatics Institute
E. coli	*Escherichia coli*
EM	Electron microscopy
Exp.	Experiment
FAD	Flavin adenine dinucleotide
FAS	Fatty acid synthase
FPLC	Fast protein liquid chromatography
Fr.	Fraction
FSC	Fourier shell correlation
G	Gauge, internationally-used scale for sizing needles
GANs	Generative adversarial networks
HADDOCK	Highly ambiguous data-driven docking
IEX	Ion exchange
LB	Luria-Bertani medium

LC	Liquid chromatography
LD	lipoyl domain
M. smegmatis	*Mycobacterium smegmatis*
M. tuberculosis	*Mycobacterium tuberculosis*
MD	Molecular dynamics
MDa	Megadalton (1,000,000 Dalton)
MS	Mass spectrometry
microED	Microcrystal electron diffraction
min.	minimum
min	minute(s)
MRC	Multiple random coupling mechanism for oxo-acid dehydrogenase complexes
MW	Molecular weight
N-ter	Amino-terminus, NH_2-terminus, N-terminal end or amine-terminus
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydride
N. crassa	*Neurospora crassa*
NMR	Nuclear magnetic resonance
NMWL	Nominal molecular weight limit or pore size
NPC	Nuclear Pore Complex
Nup107	Nucleoporin 107
OADHc	Family of α-keto acid/2-oxo-acid dehydrogenase complexes
OGDHc	2-oxoglutarate or α-ketoglutarate dehydrogenase complex
PBS	Phosphate-buffered saline
PC	Positive control
PDB	Protein data bank
PDHc	Pyruvate dehydrogenase complex
PPI	Protein-protein interactions
PSBD	Peripheral subunit binding domain
RMSD	Root-mean-square-deviation
rpm	Rounds per minute
RNN	Recurrent neural network
SAXS	Small-angle X-ray scattering
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
TAP	Tandem affinity purification
TCA or CAC	Tricarboxylic acid cycle or Krebs cycle or Citric acid cycle
TPP	Thiamine pyrophosphate
TEMED	Tetramethylethylenediamine
WB	Western blot
XL-MS	Crosslinking mass spectrometry

1. Introduction

In biotechnological and medical approaches, understanding biomolecular function through high-resolution characterisation provides chemical tools to study the inner workings of the studied system. In the light of atomic resolution, we come to know and understand the physical-chemical properties of life processes, like DNA replication, transcription, translation, enzymatic catalysis, folding, biomolecular recognition and aggregation. All of them comprise important works that occasionally receive Nobel Prizes (**Table S1, Appendix**). Nobel Prizes have also been awarded to several innovations in biophysical methods that yielded high-resolution structures (**Table S1, Appendix**).

Indeed, many high-resolution structures have been resolved using methods like X-ray diffraction (*e.g.*, crystallography [1] or X-ray free-electron lasers [2]), Nuclear Magnetic Resonance spectroscopy (solution [3-5] and solid-state NMR [6]), cryo-electron microscopy (cryo-EM) [7], micro-electron diffraction (MicroED) [8] and single-particle analysis [9], [10] and in extraordinary cases, cryo-electron tomography (cryo-ET) followed by subtomogram averaging [11-13].

The approaches mentioned above gain further impact when combined with the immensely advancing field of computational structural biology [14], pioneered by *Karplus* [15], *van Gusteren* [16], *Levitt* [17], *Warshel* [18], *Chothia* and *Janin* [19]. The benefits of this field provide the advantage of hypothesis and prediction at a molecular level, supplemented immensely with evolutionary- and physics-based calculations. A pinnacle of these developments is the machine learning software `Alphafold` [20] from Google Inc., which is now revolutionising structural biology via the rapid and, very often, accurate calculation of protein structures and, very recently, some of their stable interactions [21].

Each high-resolution structural method has a different concept for analysing biomolecules. X-ray crystallography, for example, can provide insights concerning the function and reports structures of highly purified and concentrated, crystallised biomolecules. After a century of structural studies, X-ray crystallography holds nine out of ten biomolecular structures at high-resolution, pointing out the power of this methodological approach (www.pdb.org). After the works of the groups of *Wültrich* [4] and *Kaptein* [3], solution NMR spectroscopy, although considered a high-resolution method, has focused on resolving biomolecular dynamics, kinetics and excited states. As in crystallography, in the case of in solution NMR spectroscopy biomolecules

usually have to be highly pure to be studied, but might also be examined at room temperature and not in a crystal-like state, bearing the advantage of being investigated in a more dynamic state. Within the same field and following a similar pattern, solid-state NMR attempts nowadays to overcome the difficulties that solution NMR had, concerning samples of increased complexity [22].

When discussing traditional structural methods to describe *in situ* structures, electron microscopy has a significant share, although until recently, high-resolution with such a method was not realistic. However, recent advances [23,24], including direct electron detectors to increase the signal-to-noise ratio in electron micrographs [25], automation imposed while acquiring data on the microscopes [26,27] and in image processing [28-30], contributed to the "*resolution revolution*" as portrayed by *Külhbrandt* [31]. The pinnacle of cryo-EM achievements was the award of the Nobel Prize (**Table S1, Appendix**) to *Henderson*, *Dubochet* and *Frank* "*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*". Currently, near-atomic resolution of biomolecular structures can be achieved with less effort and methodological limitations like radiation damage [32,33] as well as smaller size issues [34] are being tackled more efficiently. On the other hand, space for improvement still exists, as the resolution is often not uniform, or interpretation of the electron density maps (basically, Coulomb potential maps) is difficult due to decreased local resolution [35]. The origins of the fluctuations of local resolution observed in cryo-EM maps are pretty complex and often include biomolecular dynamics, biochemical/biophysical heterogeneity, or single-particle alignment inaccuracies. As it is quite a new field with actually not settled standard procedures, there can be a lot of discussion and improvement regarding, for example, the data collection practices, image processing strategies and resolution estimation methodologies based on different interpretations of theory [36,37]. Up to recently, cryo-EM shared the philosophy of X-ray crystallography, where in order to capture a low energy state and thus have better stability, samples are purified to homogeneity.

Finally, cryo-ET, often followed by subtomogram averaging integrates a more "in-cell" philosophy, currently addressing method-specific methodological limitations (*e.g.,* data acquisition tilt schemes [38], contrast transfer function correction of thick specimen [39], *etc.*). Such advances allowed a few laboratories to reach the previously unachievable atomic resolution [11-13]. Usually, cryo-ET is concerned with *in situ* processes at the highest possible resolution [40] with a combination of data of

supplementary biochemical and biophysical methods [41,42]. One characteristic example of this method comes from the resolved structure of the nuclear pore complex (NPC), which has also undergone its own resolution revolution [43] with major contributions from the *Beck* [44], *Hoelz* [45] and *Schwartz* [46] laboratories. The higher-resolution story starts in 2013 when *Bui et al.*[47] determined the structure of a significant scaffold motif of the human NPC, the Nup107 subcomplex, integrated into the NPC, providing a resolution at around 33 Å and evolves to those very close to identifying and localising all structural subunits and proteins of the NPC unambiguously at approximately 20 Å [45,47-50]. Even more recently, resolutions where a more detailed comparison is achieved, shed more light to the architecture of this huge (>200 nm in diameter) biomolecular assembly. For example, studies over the conformation changes of the NPC inner rings (I.R.s), the volume of its central channel and in general the cytoplasmic and nucleoplasmic rings organisation between previous and more recent models [51-53] are among the reported ones. All these achievements were possible due to the combination of state-of-the-art cryo-ET, proteomics, crystallography, computational modelling and unique integrative approaches to explain protein-protein interactions [50], sometimes directly from structures within cells [54].

Usually, samples studied with the structural methods mentioned above are highly purified biomolecules or even cells/organelles deposited on a grid in order to be structurally characterised. Of course, there are many advantages of studying samples purified to homogeneity. For example, they are more likely to reach near-atomic resolution and access subsequent chemical tools. On the other hand, these molecules/complexes are studied outside their native environment and cellular context, meaning that they might not represent exactly their in-cell status. *In situ* methods study the architecture of molecular species in cells and if the identity of the biomolecules is not known *a priori*, characterising unassigned electron-optical densities of biomolecules in any reconstruction from *in situ* maps becomes more difficult. Currently, protein identification with mass spectrometry (MS) is not yet at the level of categorising molecules *in situ* [41] with single-cell proteomics, although advances in this field are in progress [55]. Soon advances in analytical chemistry might allow protein identification from single-cell samples or even directly from the cryo-EM grids. Till then, scientists are called to bridge the gap in understanding the existing cellular architecture using the current structural tools. In general, molecules in purified samples can be unambiguously determined. However, they remain outside their cellular milieu,

conversely to *in situ* molecules, and although present in a closer-to-life biomolecular state, they can be identified only if they are already known, rarely reaching high-resolution (**Figure 1**). This gap is described as the "molecular identity gap" [56].

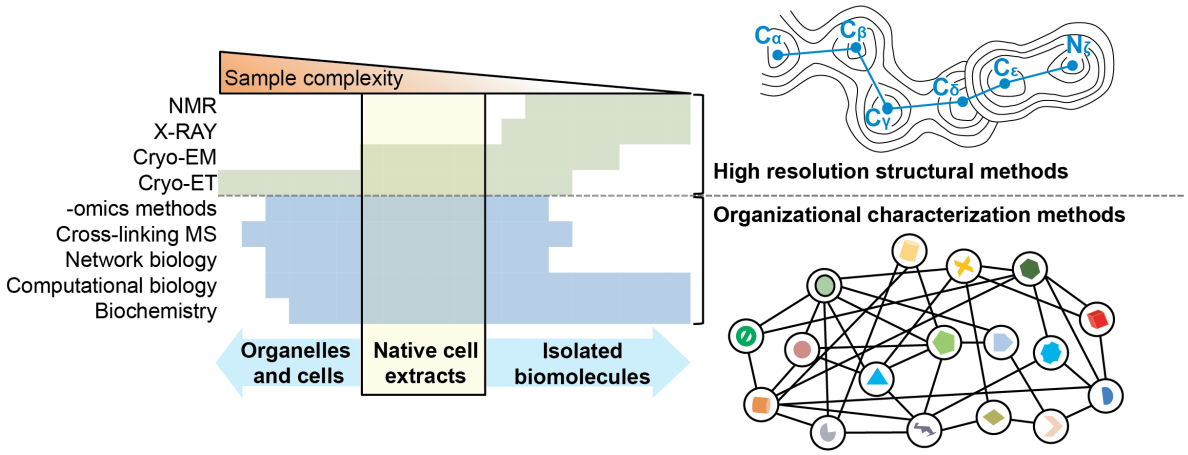


Figure 1. "The molecular identity gap".
Structural and protein characterisation methods have various advantages and disadvantages when studying biomolecular complexes of variable degree of complexity (see text). However, only molecules from cell extracts act as a bridge between *in vitro* and *in situ* studies, taking advantage of high-resolution structural characterisation, -omics technologies, biophysical modelling and network biology. The two images on the right side imply results from the different methods: High-resolution structural methods allow atoms to be placed in resolved electron densities; Organisational characterisation methods allow understanding interactions of proteins and their complexes within complex networks (connecting lines represent interactions, shapes represent different molecules). (Source: Figure reproduced from [56])

A way to bridge this "molecular identity gap" is by studying biomolecules in a quasi-purified state (**Figure 1**) *with structural biology*. These molecules are studied in highly heterogeneous mixtures, for example, cell extracts. This way, the advantage of being in a closer-to-life state compared to homogeneous biomolecules [57,58] is more apparent since interactions that otherwise would have been removed, to approach the desired high purity for structure elucidation using traditional methods, have higher chances of being retained. Additionally, analytical methods advanced together with the study of native cell extracts. For example, proteomic identification of protein complexes was extremely challenging [59], but is now evolving to a common task, with, *e.g.*, theories regarding the evolution of protein complexes [60]. Of course, the quantitative proteomics algorithmic development (*e.g.,* `OpenMS` [61]) combined with the increased sensitivity of mass spectrometers [62] both aimed to this direction. On the same page crosslinking mass spectrometry (XL-MS) methods have evolved from studying purified samples [63,64] to characterising samples of increased complexity [65,66]. Thousands of biomolecular interfaces in cell extracts [57], directly in purified organelles [67], or, even, *in vivo* [68] are recovered nowadays through XL-MS.

Furthermore, recent advances in XL-MS technology [69,70] are expected to increase the discovery of crosslinks in heterogeneous samples. The fact that cell extracts retain certain advantages for biochemical manipulation, which are not available when studying intact cells, cannot be neglected. Another advantage of studying cell extracts is that computational structural biology algorithms can be applied to understand the structure of protein-protein interactions discovered with proteomics [57] and integration of available computational tools. Additionally, network biology characterises the architecture of higher-order interaction networks that reside within those extracts [71]. Development of novel protocols that merge proteomic identification in cell extracts with direct protein quantification are available [71] and sophisticated network biology is applied to understand the interconnectivity of the proteome [57]. Such improvements drive the new era of understanding novel architectures of protein complexes and introduce a validation and expansion of interconnected metabolic pathways [60], initially identified at a global level using affinity purification coupled to mass spectrometry (AP-MS) by the pioneering work of *Gavin* [59,72].

1.1 Identifying protein communities in native cell extracts

Back in the 50's *Palade*, *Claude* and *de Duve* [73-75] received the Nobel prize in Physiology or Medicine (1974) for their work visualising organelle structures' architecture within cell extracts. Also, other scientists have reported imaging of protein complexes in cell extracts with electron microscopes [76-78]. However, only recently comprehensive near-atomic imaging of biomolecular complexes in native cell extracts was achieved with work in *Kastritis et al.* [57]. A single step biochemical fractionation was performed on cell extracts of a thermophilic fungus (*Chaetomium thermophilum*), followed by a large-scale electron microscopy imaging of all fractions (**Figure 2** A-C). Importantly, proteomic identification and electron microscopy were applicable due to the wise selection of the separation method [79] and the known advantages of using a thermophilic organism for structural biology [80] allowing the elution of large molecular weight complexes amenable to electron microscopy studies due to their superior visibility. Thus, large-scale identification and visualisation of protein complexes and their interconnectivity were achieved through investigating the proteinaceous material of native cell extracts.

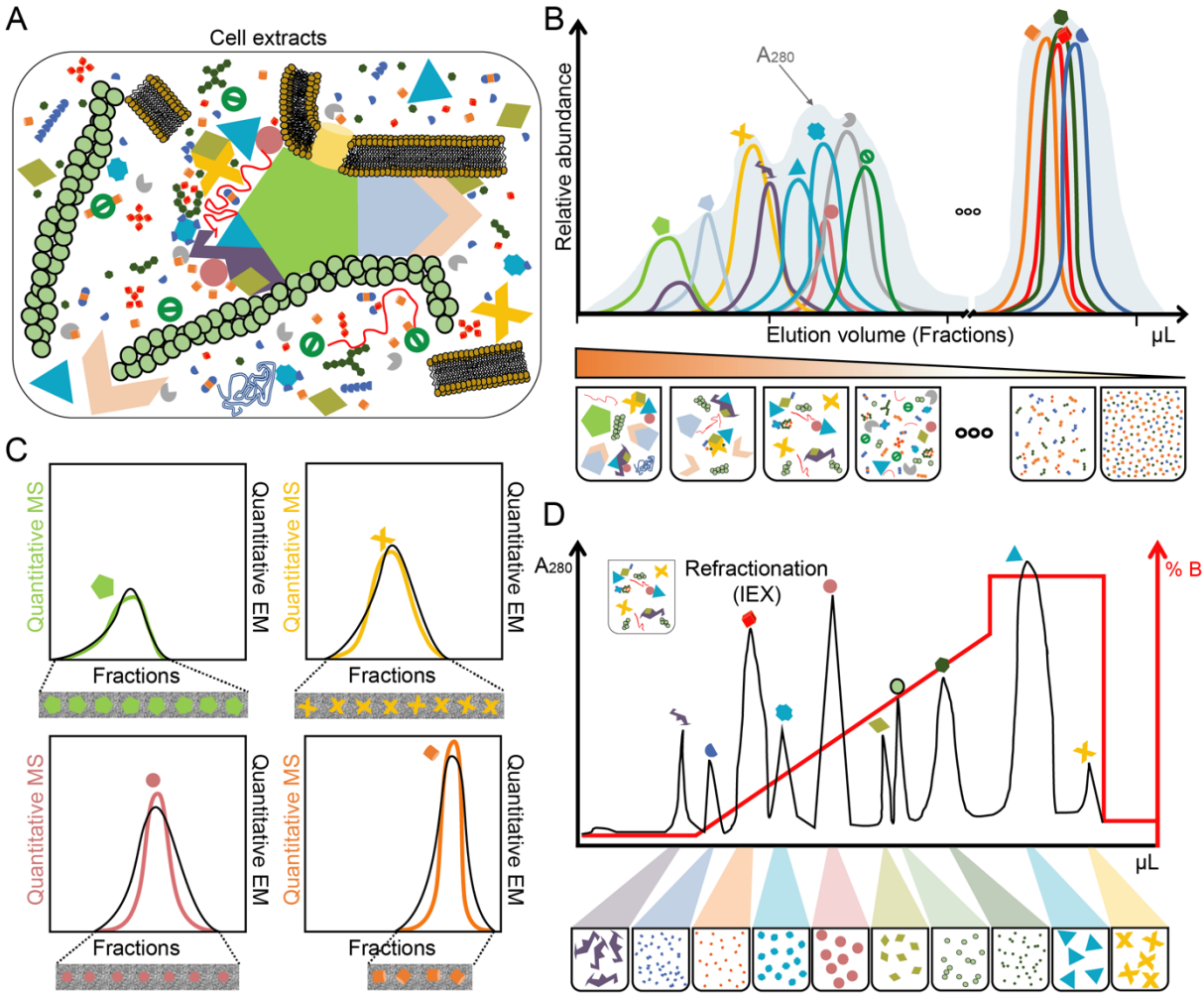


Figure 2. Cell extracts for structural characterisation and identification of molecular species.
A. Illustration of a cell extract, with biomolecules illustrated as a very concentrated suspension of particles. Different sizes, shapes and complexes are apparent with different abundances. **B.** Fractionation of cell extracts retains interactions of the homogenate and is advantageous in a way that molecular abundances can be inferred using absorbance and quantitative mass spectrometry measurements (shown in the plot). In addition, those fractions of native cell extracts are accessible to electron microscopy, therefore, identifying the shapes and abundances of proteins, protein complexes and their higher-order assemblies (shown below the plot). As the size of the biomolecules increase, their concentration decreases. **C.** Method to annotate structural signatures from native cell extracts and therefore addresses the challenge of the "molecular identity gap". Each fraction is analysed, *e.g.,* with proteomics methods to identify the abundance of the molecules (coloured lines). As shown in (B), protein complexes elute in different fractions according to their molecular weight. By performing quantitative mass spectrometry (Quantitative MS, left y-axis) in each fraction (x-axis), an elution profile can be derived, which illustrates their abundance. Each profile is shown in different colour for each of the identified molecules. Additionally, large-scale electron microscopic imaging is applied to the same fractions, therefore, quantitatively describing observed structural signatures (Quantitative EM, right y-axis; black lines and coloured shapes). This means that, in black solid lines, all molecules within each fraction are counted after acquiring hundreds of images (large-scale electron microscopic imaging) using the electron microscope. The solid black line corresponds to the abundance profile of each molecule. Then, abundances are correlated and the molecular signature observed in Electron Microscopy is identified using the abundance measurements derived from the mass spectrometry data. Details of the method can be found in [57]. **D.** Another method to identify protein complexes; extracts are subsequently fractionated in a second dimension, *e.g.*, with ion exchange.

Fractions are greatly simplified and therefore, only by proteomically identifying the molecules within the fractions and selectively imaging the fractions with electron microscopy the biomolecule of interest can be identified and determined and subsequently identified in the coarser fractionation step. Abbreviations: A_{280}, absorbance at 280 nm; IEX, ion exchange; %B, elution of buffers. (Source: Figure reproduced from [56])

In the lack of proteomic identification, electron microscopy of the cell extract cannot characterise the identity of the visualised biomolecular assembly, a problem also encountered in cryo-ET of organelles or cells. To overcome this, a method to proteomically identify the shapes of the biomolecules (structural signatures) observed under the electron microscope [57] was developed. Briefly, protein abundance quantification in each fraction performed by MS is correlated to the frequency of appearance of each structural signature in a large-scale electron microscopic imaging (**Figure 2** C). With this type of correlation, highly abundant structural signatures that elute over four or more native fractions can be annotated. An alternative or extra identification step of structural signatures within native cell extracts can be a further orthogonal purification of each fraction (**Figure 2** D). In this way, the content of the fraction can be further simplified. If there is no aggregation or change of conformation introduced, then a complex can be visualised in the subsequent purification step and therefore, identified easier in a fraction using again MS. Following this methodology which is by default relying less on statistical calculations would automatically lead to a less ambiguous identification as well. However, the possibility that scaffolds or binders might dissociate and therefore, the identification of binders will not be possible in downstream purification protocols, should not be ignored.

In 2006 *Gavin et al.* introduced the idea of "protein communities" [72] in native cell extracts. The idea of protein communities is mainly described as flexibly associated molecules of several macromolecular complexes arranged in close proximity and encode functionally synchronised biomolecular entities. Among other functions, they may transduce signals and regulate protein synthesis on local cellular demand or even transfer substrates along enzymatic pathways (dubbed *metabolons*, reviewed in [58]). However, up to now and due to their complexity, their intrinsic structural characterisation was limited to a few abundant biomolecular complexes, but machine learning applications can massively improve this approach and lead to deeper comprehension of a higher number of protein communities captured in the cell extracts.

1.2 Low-resolution 3D reconstructions of proteins and protein complexes in native cell extracts

Nowadays, improvements in image processing bring to life the heterogeneous specimen image analysis, even to the level of 3D reconstruction of biomolecules within these samples [81] and native cell extracts pose as a case amenable to image processing both at low- and high-resolution ([57] and **Figure 3** A-B). An increasing number of works validate and expand image processing methods of cell extracts, either by calculating multiple reconstructions from a negatively stained single fraction [82] or by using the negatively stained cell extract of single cells to produce low-resolution 3D reconstructions of ribosomes [83]. Identifying structural states of native protein complexes is within reach and similar protocols used to derive class averages from cellular fractions will lead to an overview of cellular organisation. The biggest challenge for the image processing algorithms is the simultaneous existence of a variety of protein complexes within the same fraction/sample (**Figure 3** A). However, the user-friendly Bayesian approaches [84], applied by *Scheres* regarding the semi-automated cryo-EM data analysis [29] can outstandingly distinguish complexes of different states. These approaches may reconstruct independently structural signatures in three dimensions, therefore, partially solving the heterogeneity problem, not only at the single-molecule level but also at the level of discriminating and reconstructing simultaneously distinct protein complexes of high abundance (**Figure 3** A and [57]). An example demonstrating such a case is the simultaneous reconstruction of the human ribosome and proteasome, present in the same cellular fraction after applying negative stain [82].

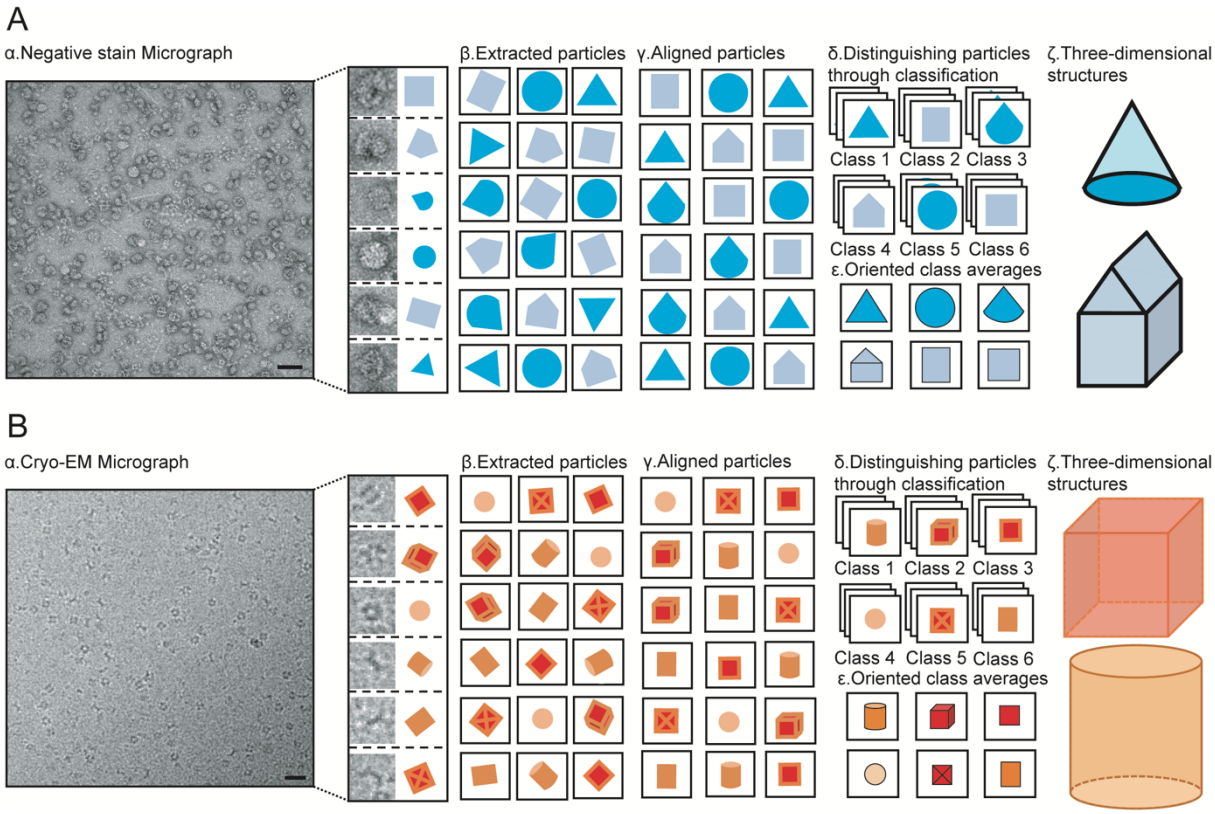


Figure 3. Cell extracts are amenable to image processing and simultaneous characterisation of 3D shapes residing within, both at low- (A) and at high- (B) resolution.
On the left, images of cellular fractions acquired with a JEOL JEM-3200FSC in Halle (upper image, negative staining with uranyl acetate 2 % in water) and the FEI Tecnai Polara in EMBL-Heidelberg (lower image, cryo-electron microscopy). Molecules are apparent in both micrographs, showing single particles from native cell extracts from high-molecular-weight fractions (upper image) to low molecular weight fractions, coming from *C. thermophilum* native cell extracts. Subsequent image processing can acquire 3D reconstructions for all abundant biomolecules (see text for details). Scale bars: 60 nm (for negative stain image); 20 nm (for cryo-EM image). (Source: Figure reproduced from [56])

1.3 High-resolution structure determination from cell extracts: Fatty acid synthase, a breakthrough story

High-resolution data acquisition of cells and cell extracts [85], came within reach by *Selenko, P. et al.* [86] and *Theilet, F. X. et al.* [87] and brought one step further by *Inomata, K. et al.* [88] for in-cell high-resolution multi-dimensional NMR spectroscopy of proteins in 2009. However, even in the last case, ubiquitin was mutated to preserve a monomeric state, ideal for NMR. Additionally, proteins are expressed in micromolar or millimolar concentrations, rendering the cell in a state of single-protein overexpression.

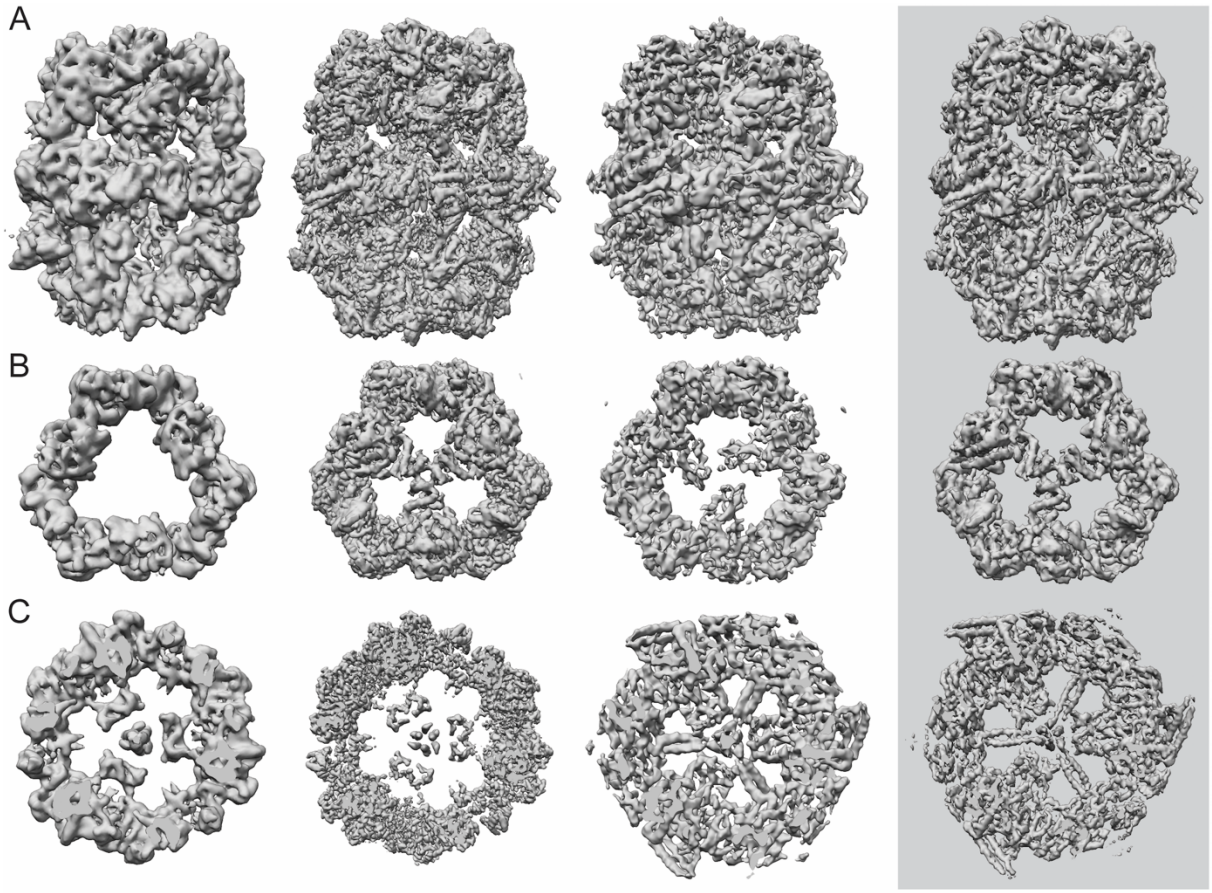


Figure 4. High-resolution structures of fatty acid synthase (FAS) from different organisms, resolved at a resolution below 7.5 Å.
C. thermophilum FAS, resolved from native cell extracts, exhibits much stronger statistics as compared to the others from different organisms previously resolved, highlighting the advantages of (a) thermostability for structure determination and (b) native cell extracts for structural stability and subsequent high-resolution. Density map contour levels are shown as suggested by the authors. FAS in various views (**A.** side, **B.** top and **C.** central wheel), stressing different aspects of the coulomb potential. (Source: Figure reproduced from [56])

On the other hand, the alternative for high-resolution analysis of native cell extracts provided by cryo-EM is quite appealing because: (a) molecules are not overexpressed but are visualised in their native abundances; (b) cells are not grown in method-specific media; and (c) biomolecules can be visualised without *a priori* knowledge of what is being imaged. Some of the most appealing examples include a thermophilic eukaryotic fatty acid synthase (FAS) from cell extracts [57] and purified Tau filaments from human Alzheimer's disease brain and subsequent near-atomic description of their structural heterogeneity [89]. Especially for the example of FAS, the

enrichment within the studied native cellular fraction was less than 50 %. After EM acquisition and image processing, 3,933 single particles of FAS were selected and a reconstruction at near-atomic resolution was achieved ([57] and **Figure 4**). Compared to previously resolved cryo-EM maps of FAS from other species [90-92] the reconstruction, in this case, revealed superior features as the cap of the molecule, completely unknown and absent in previous cryo-EM studies. Another striking point of this study is the low number of particles that contribute to the final 3D reconstruction, illustrating the advantages of a thermophilic organism (*C. thermophilum*) as a model system for high-resolution structure determination, further corroborating the usage of thermophilic eukaryotes for structural biology.

Compared to other recently resolved structures of FAS [93,94], resolution might be lower, but in those studies, the number of single particles used is much higher (**Figure 4**). Of course, *C. thermophilum* FAS has inherent stability and rigidity and has been acquired with direct electron detection (FEI Falcon II camera). In contrast, other FAS structures do not have similar features or/and have been acquired with older detection technology. For this reason, compared cryo-EM maps of FAS should be carefully interpreted. The advantages of studying *C. thermophilum* FAS in cell extracts should be ultimately compared to *C. thermophilum* FAS purified to homogeneity.

C. thermophilum stands among the ideal model organisms for structural studies [95] with striking examples like the recombinant expression of nucleoporins and the subsequent structural determination by X-ray crystallography [96]. *C. thermophilum* is already regarded as a system of extensive study in the NPC field [95] and is nowadays expanding as a model system for structure determination due to the above-mentioned encouraging results concerning cell extracts [57]. The genetic toolkit can be enriched for structural biotechnology using a thermophilic source like the recently described genetic modification of the fungus pioneered by the *Hurt* laboratory [97]. For all the above-mentioned reasons, this organism was selected as the model organism for this work.

1.4 Cell lysis: a prominent limiting step shading over the many advantages of studying cell extracts

Since the first experiments of cell extracts and subsequent application of purification protocols, including centrifugation steps for biological samples [98], there has been a debate on how representative the results of such approaches are. Among the

main arguments is the distortion of the natural cellular state since lysis steps might lead to an irreversible thermodynamic state that is highly unlikely to be the one that biomolecules attained before lysis. Another unavoidable effect is that aggregates must be removed by applying, for example, an extra ultracentrifugation step, therefore, losing high amounts of cellular material and inevitably important information. A striking example where gentle manipulation of the sample is required is illustrated by the pyruvate dehydrogenase complex that should be gently sedimented; otherwise, subunits might dissociate, resulting in diminished activity in homogeneous fractions compared to native cell extracts [99]. Also, when the cell is lysed, the excluded volume effect encountered within the cell's environment is not any more present, reducing in such a way, biomolecular interactions otherwise existent in the cell or even inducing folding effects driven by dilutions. Further dilutions that might occur during the homogenate handling could lead to additional dissociation events. Another aspect that should be considered is that the membrane-embedded or bound material is usually discarded, thus excluded from further investigation. Since only the soluble material is further processed for analysis, all the recovered structural states of biomolecules in the cell extracts are affected because molecules that function in a dynamic exchange between membrane-bound and membrane-free environments are under-represented in the studied cell extract, as well as, of course, membrane-embedded proteins.

Cellular material after lysis can be structurally affected and this was also shown with a deformation observed in yeast mitochondria where *in situ,* they have an elongated shape [100]. At the same time, when isolated and purified after sucrose gradient centrifugation, they appear circular. Based on that, no one can guarantee that dramatic changes in shape could not also occur at a simpler molecular level, such as in the case of biomolecular assemblies implying that a lysis step could introduce alterations at the level of the structure. Another effect of cell lysis is the possible association of non-specific binders that would otherwise just have been in vicinity inside the cell. Therefore, higher-order assemblies of biomolecular complexes usually observed in cell extracts should be carefully interpreted as they might stem from the binding of non-specific interactors that happened to coexist in the same cellular micro-environment.

These reasonable concerns cannot be easily avoided when applying any biochemical method. The study of the cellular organisation employing cell extracts is the closer-to-native biochemical approach in our research toolkit up to now, until a

combination of structural and molecular biology methods will unquestionably identify and structurally characterise protein complexes directly from living cells. This has not happened so far, but cryo-EM resolved a vitrified state of the (extremely abundant) in-cell expressosome [101].

Of course many arguments come from the opposite direction and against the described above phenomena, claiming that they are reduced since: (a) various separation methods are being applied leading to same results and crosslinking experiments have shown recovery of the same in-cell and in-extract interactions (b) thermostable organisms are employed therefore minimising dissociation of biomolecular assemblies due to their superior stability; (c) the cellular material is directly observed with electron microscopy, confirming that aggregates or non-specific interactions are minimised; (d) measuring reproducible protein abundance in cell extracts in different experiments to partially avoid non-specific interactions are performed by monitoring for example reproducible protein abundance in cell extracts; and, most importantly, (e) cell extracts and especially the complexes studied can be tested as catalytically active (*e.g.*, by activity assays) strengthening in such a way the observed interactions as being closer to reality.

Overall, technical and biological reproducibility is essential and mandatory, especially when tackling the concerns above. Equally important could be the validation with complementary in-cell visualisation/characterisation techniques/methods (*e.g.*, super-resolution microscopy, cellular tomography, in-cell NMR) that give another view of the same in cell phenomena.

1.5 Native cell extracts enable functional studies by retaining activity

Cell extracts have conquered the world of functional studies utilising model systems as they have also frequently addressed many of the afore-mentioned concerns. Numerous examples build up a list of studies that strengthen this role of cell extracts, including the investigations of *Nirenberg and Matthaei* [102] who used cell extracts to translate poly-Uracyl RNA sequences, leading to the discovery of polypeptide synthesis. In this way, cell extracts served a pivotal role in deciphering the genetic code. Recent examples of the employment of various types of cell extracts, such as rabbit reticulocytes, *E. coli*, wheat germ and insect cells for massive protein

production, are impressive biotechnological application improvements, achieving yields that usually exceed a few grams of protein per litre of reaction volume [103]. In another direction, the one of understanding DNA damage and repair [104], homogenates of extracts from *Xenopus* eggs were used since they retain all molecular factors required to perform DNA repair outside the cellular context, by mechanisms also conserved in humans [105]. The list further enriched with many more examples of cell extracts applications: metabolic manipulation [106], optimal ATP production[107], unnatural amino acid incorporation [108] and in particular, manipulations for structural biology studies [109].

The advantages of using cell-free systems like cell extracts are not only limited to the studies of cellular function but expand to applied research as compared to their cellular counterparts when regarded from the biotechnological point of view [110]. Even though they are vulnerable to degradation as they are outside their host environment, *i.e.,* the cell, in comparison to their cellular counterparts, they have: (a) higher product yields as by-products are reduced; (b) higher tolerance to toxic compounds; (c) ability to implement biological reactions that living systems or chemical catalysts cannot implement, or even if they can, cell-free systems can do it more efficiently; (d) possibly, higher reaction rates, *e.g.,* higher power outputs are observed in enzymatic fuel cells [111]; and (e) a broader range of reaction conditions, *e.g.,* higher temperatures, lower p*H*s, tolerance to ionic liquids and organic solvents, thus giving them a broader spectrum of biochemical conditioning.

1.6 Higher-order assemblies revealed utilising cryo-EM of native cell extracts

In previous studies, the FAS metabolon was described from native cell extracts and observed with various unknown MDa-complexes bound [57]. Employing proteomics and XL-MS methods, a fungal-specific carboxylase forming an interaction interface with FAS was identified as one of these binders. The binding interface was specific and the site localised at an entrance/exit tunnel of the full complex, coming in agreement with XL-MS data and cryo-EM results pointing to the same location [57]. Connections of biomolecules with electron-dense material representing higher-order states are present in cryo-EM micrographs of cell extracts (**Figure 5**). Of course, this material is of unknown origin and a hypothesis that it maintains the structural integrity

of this higher-order organisation is made. This observation has never been performed before at such a resolution and the combination of proteomics with cryo-EM can provide additional insights into the identity of scaffolds organising biomolecular assemblies. An assumption that such higher-order complexes are organised not only by cytoskeletal elements or membranes but also on flexible biomacromolecules like disordered proteins and/or RNA is quite appealing, but the characterisation of such flexible interactions remains for the moment out of reach. Such cases can only be studied currently by applying NMR, as shown in the example of *Kalodimos* and his group revealing a complex of a chaperone with a disordered protein [112]. It remains a significant challenge to model such systems. This flexibility will obviously limit and hinder insights into dynamic higher-order assemblies by cryo-EM of cell extracts.

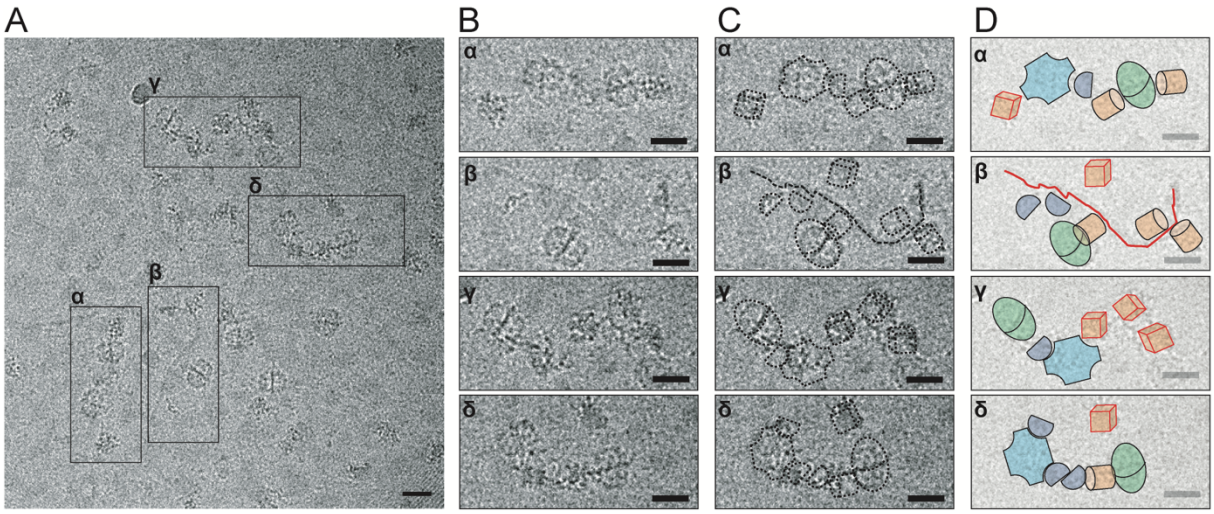


Figure 5. Higher-order organisation within cell extracts.
Fatty acid synthase is shown in a typical cryo-electron micrograph **A.**, also shown in [57], used to reconstruct the 4.7 Å resolution structure from *C. thermophilum*. It is apparent that FAS participates in higher-order assemblies (**B-D**) and various structural organisation principles are seen (α-δ). Higher-order complexes are frequently bound to electron-dense highly flexible material (β), connecting more stable biomolecular structures; Other times the protein complexes are directly bound (α,γ-δ) showing a characteristic higher-order organisation of a metabolon ([58]). Scale bar: 30 nm. (Source: Figure reproduced from [56])

1.7 Studying cell extracts at high-resolution

1.7.1 Structural characterisation of less abundant molecular signatures is currently limited

Due to the nature of the biochemical sample of a cell extract, it is expected that abundant protein complexes, such as FAS, the ribosome, proteasome, glycolytic enzymes *etc,* will be amenable to high-resolution structure determination. This is because high-resolution structure determination of highly abundant proteins has higher

probabilities than the rarer ones. Still, hardware and software automation advances will soon improve and allow resolving structures of less abundant biomolecular complexes. Recent advances, like the Volta phase plate [113] which is a hardware upgrade of the electron microscope, can dramatically enhance the contrast of electron micrographs to the extent that even smaller biomolecules will be reliably identified at the single-particle level. This will aid in understanding higher-order structures in cell extracts with fewer images and might also simplify computational sorting of highly heterogeneous specimens. Another advance is towards software used for image processing, such as deep learning approaches for single-particle picking [114]. The application of those sophisticated image recognition algorithms to highly heterogeneous samples will further support particle assignment and will aid in the structural characterisation of many different biomolecules from a single sample. From the biochemical point of view, the fractionation of different organelles could also enrich for organelle-specific pathways in the extracts. This can simplify the cryo-EM study by minimising non-specific interactions with biomolecules from other compartments.

Even if the toolbox mentioned above can improve the study of the abundant complexes, the rare or very low abundant ones remain more unlikely to be resolved and properly annotated. Considering that, molecular biology methods and genetic manipulation of the model organisms can be a solution to monitor assembly principles of biomolecules using gold tags to aid visualisation in electron microscopy or fluorescent tags to visualise using light microscopy. Theoretically, such rare events can also be imaged by cryo-correlative light-electron microscopy (cryo-CLEM) on cellular fractions. A medium resolution electron density map might be subsequently provided by cryo-ET and, in combination with computational modelling of rare biomolecular assemblies, resolve the higher-order organisation of these low abundance higher-order assemblies. In addition, advances in the cryo-microscope level, such as cameras with wider field of views, higher sensitivity and, overall, more efficient imaging capabilities can further allow the structural characterisation of increasingly rare biomacromolecules and their interactions.

1.7.2 Software for the interpretation of molecular complexity

XL-MS coupled with flexible protein docking using as restraints cryo-EM maps is a common combination used nowadays to resolve biomolecular interactions. Several approaches can be used for such modelling procedures like HADDOCK from the *Bonvin* laboratory, for atomic modelling [115] and the integrative modeling platform (IMP)

[116] for a coarse-grained description of biomolecular assemblies. In a recent development of HADDOCK, a docking program that utilises the crystallographic and NMR System (CNS) [117] as its computational engine, one can find significant signs of progress in software to handle heterogeneous data. HADDOCK was at the very beginning conceived for data integration [118] and included NMR and mutagenesis data to guide the modelling procedure. Since in the last decade, a lot of improvement has been incorporated in the development of HADDOCK, it has turned into an integrative modelling suite [115] utilising data derived from cryo-EM [119], co-evolution [120], chemical crosslinking (XL-MS) [57], ion mobility Spectrometry/Mass Spectrometry (IMS-MS) and small-angle X-ray scattering (SAXS) [121]. Nowadays, the software is in a status that can handle various types of biomolecules (small molecules, proteins and peptides, nucleic acids or lipids) and model large macromolecular complexes [122] or membrane complexes [123] at atomic resolution. The user can interpret the data as distance restraints [124] and combine HADDOCK with other high-end modelling software [125]. Studying the architecture of cell extracts is a demanding process and computational algorithms that can deal with a variety of data are essential. For example, the scoring function of HADDOCK [126], which is a force field used to rank the plausibility of predicted biomolecular interactions and is based on physical-chemical principles, is even efficient in providing qualitative predictions of macromolecular binding affinity [127], *i.e.,* qualitatively correlates with the dissociation constant (K_d) which denotes the concentration of protein A at which half of the binding sites of protein B are occupied. Another option of the scoring process of HADDOCK is that it can also describe within experimental uncertainty the binding strength of protein-protein complex inhibitors [128].

Of course, progress in biomolecular modelling is needed until programs that model biomolecular architectures at atomic resolution can efficiently perform analytical calculations of higher-order assemblies that include a high number of atoms. Even AlphaFold struggles to model larger proteins. In the same direction, combined coarse-grained and fine-grained force fields can tackle such complicated refinement challenges.

Another advantage of cell extracts is that physical-chemical changes in their buffer environment can be altered according to desired conditions. Furthermore, culturing in different growth media can reveal whether specific changes in biomolecular assemblies are correlated to the growth conditions of the studied organism. Treatment of the cell extract with enzymes or metabolites that possibly could affect higher-order

structural organisation or function can examine the molecular organisation hypothesis. Finally, probing the interaction of cell extracts with additional scaffolds or crowding agents, such as membranes or chemicals, could be another approach from the synthetic biology point of view.

1.8 Studying metabolic pathways utilising cell extracts: Pyruvate oxidation

This thesis focuses on advancing our current knowledge in studying metabolic pathways within cell extracts. In particular, one of the most critical metabolic reactions during pyruvate catabolism is the oxidative decarboxylation of pyruvate. A massive enzyme, the pyruvate dehydrogenase complex (PDHc), performs this vital step of aerobic metabolism, forming acetyl-CoA, CO_2 and NADH (H^+) [129-133]. PDHc, together with 2-oxoglutarate dehydrogenase complex (OGDHc, also known as α-ketoglutarate dehydrogenase complex) and branched-chain keto-acid dehydrogenase complexes (BCKDHc), all belong to the α-keto acid/2-oxo-acid dehydrogenase complex (OADH complexes) family. All these complexes are localised in the inner mitochondrial membrane-matrix interface in eukaryotes [134] and are major metabolic checkpoints that finely tune sugar and amino acid degradation [134] (**Figure 6**). Other additional roles assigned to these complexes and making them more attractive as research, biotechnological and medically relevant targets, are that they have been shown to play roles as mitochondrial autoantigens [135] and their implication in several pathologies such as viral infection [136], neurodegeneration [137,138] and inflammation [139]. All of those complexes range from 4 to 10 MDa, being among the most significant and more sophisticated enzymatic systems of the cell and thus having a very complex architecture which is, currently, unknown in its native, endogenous state.

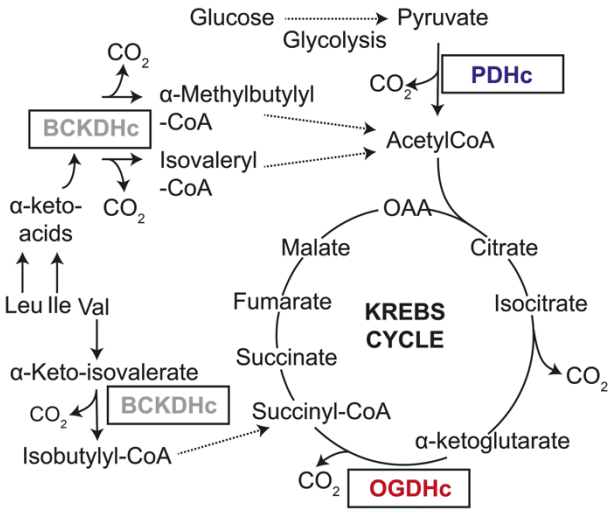


Figure 6. Pathways involving α-keto acid dehydrogenase complexes.
Biochemical pathway with PDHc, OGDHc, and BCKDHc as major participants. (Source: Figure reproduced from [140])

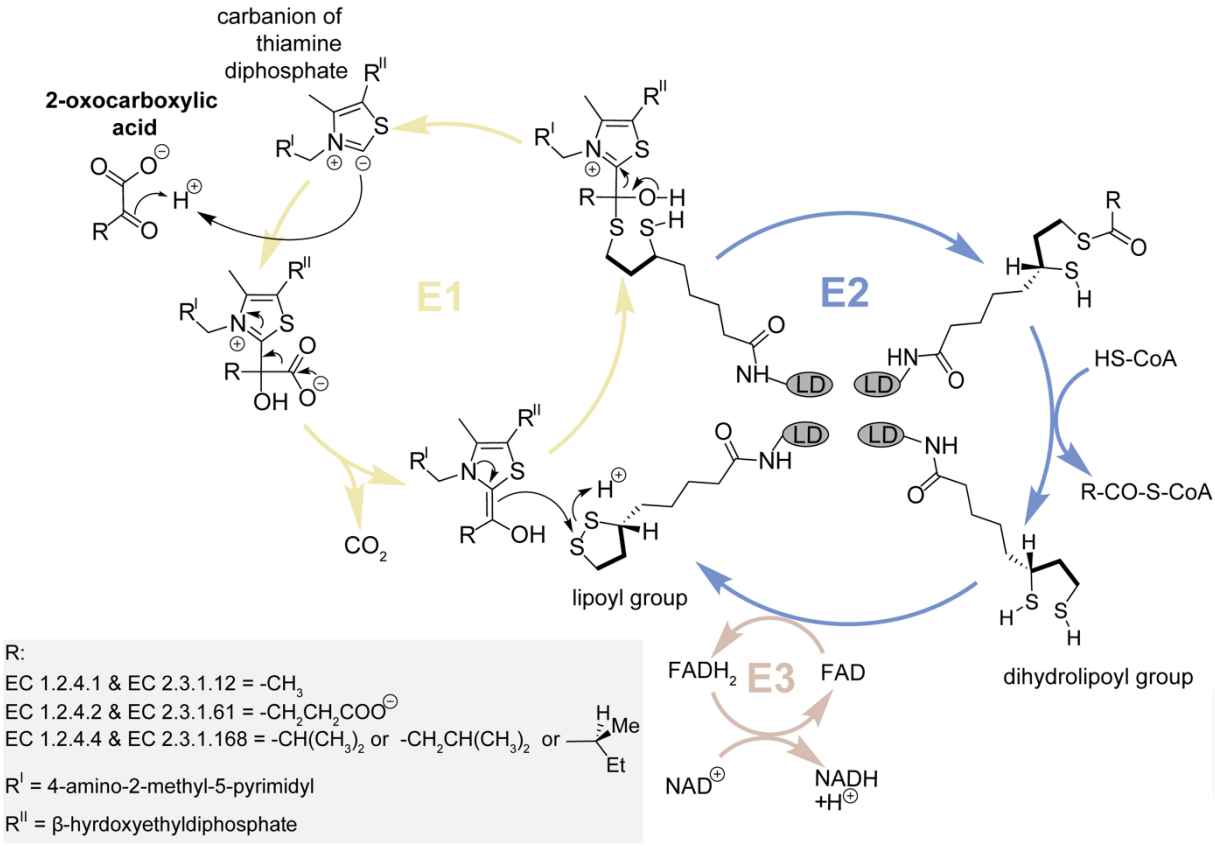


Figure 7. Reaction mechanism performed by the E1, E2 and E3 enzymes of the α-keto acid dehydrogenase complexes.
(Source: Figure reproduced from [140])

The OADH complexes mainly consist of three major components which are present in multiple copy numbers [131]: E1 (pyruvate/2-oxoglutarate/branched-chain keto acid dehydrogenase; EC 1.2.4.1/1.2.4.2/1.2.4.4, E1p, E1o and E1b, respectively), E2 (dihydrolipoyl (or dihydrolipoamide) transacetylase/transsuccinylase (or acetyl/succinyltransferase); EC 2.3.1.12/2.3.1.61/2.3.1.168, E2p, E2o and E2b, respectively) and E3 (dihydrolipoyl (or (dihydro)lipoamide) dehydrogenase; EC 1.8.1.4) (**Figure 7**). All the family members share the following coenzymes: TPP, NAD^+, FAD, lipoic acid and Coenzyme A (CoASH). The E3 component is common to all OADH complexes and is shared with the glycine cleavage system. In sequential reactions performed by the OADH family members α-keto acids are being decarboxylated and the remaining acyl group is covalently attached to coenzyme A, as shown in the **Figure 7**. PDHc converts pyruvate to acetyl-CoA, which then enters the citric acid cycle (TCA, Krebs cycle). OGDHc acting within TCA converts 2-oxoglutarate to succinyl-CoA. BCKDHc converts branched, short-chain α-ketoacids with broad specificity and provide various acyl-CoA precursors for TCA. A full reaction

cycle is the following: An E1 captures and decarboxylates an α-keto acid transferring it to the mobile lipoyl domain (LD) of an E2. Then the activated E2 LD transfers the acyl group (the nature of the acyl group depends on the particular E2) to the catalytic domain of E2, which afterwards transfers it to coenzyme A, forming the acyl-CoA. After this, the acyl-CoA is released and the LD of E2 is regenerated (reoxidised) by the E3 for a new cycle, with E3 transferring protons to NAD^+ and forming NADH. The complexes differ between prokaryotic and eukaryotic organisms and between PDHc, OGDHc and BCKDHc. Complexes might have additional components, variability in the number of LDs and stoichiometries of their subunits, distinct higher-order architectures and variation in the overall abundance in the cell [130,141-143].

Among the three OADH complexes, PDHc is the most studied. Its ordered domains of E1p, E2p and E3 proteins and their respective protein complexes have been characterised structurally in isolation (*e.g.*, in human, the heterotetramer of E1p [144], the core module of E2p [145] and homodimer of E3 [146]). Employing electron microscopy (EM) studies, the eukaryotic E2p core of PDHc has been structurally characterised, providing insights into its function [147-150]. PDHc also includes the E3-binding protein (E3BP) tethering E3. E3BP localises at the PDHc core, in mammals replacing some E2p [151] and in some fungi inside the core as four separate complexes [150]. Due to the absence of high-resolution data and the underlying structural intricacy [148] the number of E3BPs present remained debatable [152]. This is accompanied by the lack of data concerning the stoichiometry and the endogenous structure of all PDHc components (E2p, E3BP, E1p and E3 subunits). E1p and E3 attach to the flexible peripheral-subunit binding domains (PSBD) of E2p and E3BP, respectively. Molecular understanding further complicates because the catalytic mechanism depends on the highly flexible lipoyl domains (LDs) and on flexible linkers that tether E1p and E3 subunits proximal to the core [153]. All these intricacies, which are not known in detail to be structurally unambiguously characterised, limit our knowledge of the overall architectures of PDHc and other OADH complexes, the latter being far less studied.

A workflow is revealed in the current PhD thesis that studies OADH complexes from native cell extracts of *C. thermophilum*, a eukaryotic thermophilic filamentous fungus. Due to their thermal adaptation, *C. thermophilum* proteins are highly stable, making this mould an ideal model organism and advantageous for structural studies as mentioned before; Some additional statistics on the model organism relevance for structural studies are provided in the **Appendix** where the increase in cryo-EM maps

and atomic models released per year is shown (**Figure S1** A, **Appendix**) [154]. With a single fractionation step and limited purification, the native structure of protein complexes is maintained [56,82,155-157]. In this thesis the careful optimisation of growth conditions for *C. thermophilum* to retrieve active, enriched and complete OADHc complexes is described. Also, the combination of cryo-EM of OADHc-enriched native cell extracts with kinetic assays, mass spectrometry, chemical crosslinking and computational, biophysical modelling that led to the proposal of the first eukaryotic architectural model of the complete, endogenous PDHc metabolon is being reported. This thesis sheds light over the stoichiometries and proximity principles for E2p, E3BP, E1p and E3, localised in PDHc and unveils novel aspects of the endogenous PDHc structure and function from native cell extracts.

1.9 Aims of the work

Native cell extracts hold great potential for understanding biochemical processes in a level of complexity and organisation which is higher than that of highly purified, isolated and/or recombinantly expressed constructs. This is exemplified in the case of large biomolecular assemblies, *i.e.,* metabolons, such as the eukaryotic pyruvate dehydrogenase complex (PDHc), a 10-megadalton macromolecular machine: Knowledge about PDHc structure and function was majorly derived from overexpressing the component polypeptides or resolving the full structure at very low resolution from the smaller, bacterial counterpart. To, therefore, ultimately, structurally map the endogenous process of pyruvate oxidation, the main motivation of this work, three aims are pursued in this thesis:

- *Optimise* the reproducible recovery of eukaryotic cellular material for subsequent biochemical studies utilising the inherent advantages of the thermophilic fungus *Chaetomium thermophilum*. This organism was selected as model organism not only due to the high thermostability of its biomolecules, but also its eukaryotic nature because findings about its biology would better relate to those from other eukaryotes. This is a direct advantage over results obtained from *e.g.,* thermophilic bacteria or archaea.
- *Derive* an active, native cell extract in sufficient quantity to be analysed by high-resolution cryo-EM, an achievement that will provide unprecedented insights into the structure of distinct and multiple eukaryotic protein communities. Such achievement will be of great use for other researchers studying cellular biomolecular assemblies. This aim underlines the notion that the combination of architectural insights of protein complexes together with activity assays can lead to vital functional observations, which cannot be derived in the absence of either of the two analyses; In addition, cryo-EM was selected as the method of choice because it is the only method that can structurally characterise those highly heterogeneous native cell extracts.
- *Produce*, utilising these cell extracts, an endogenous assembly of an active metabolon by combining the cryo-EM results with complementary biochemical and biophysical data – In this thesis, the metabolon of the active pyruvate dehydrogenase complex from eukaryotes is studied, which is of key, fundamental importance for various metabolic and signalling functions, *e.g.,* cellular respiration.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals and enzymes

Table 1. Key resources table with chemicals and enzymes used in the current study.

Chemicals and enzymes	Source	Identifier
1,4-Dithiothreit, min. 99 %, p.a.	Carl Roth	6908.4
Acrylamide/Bis solution, 37.5:1	Serva	10688.01
Agar-Agar, bacteriological highly pure	Carl Roth	2266.3
Ammonium acetate, ≥97 %, p.a., ACS	Carl Roth	7869.2
Ammonium persulfate	Serva	13376.02
Aprotinin from bovine lung	Sigma-Aldrich	A1153-1MG
Bestatin, 10 mg	Sigma-Aldrich	10874515001
Roti®-Quant 5*X*	Carl Roth	K015.1
Clarity Western ECL substrate	BIO-RAD	170-5060
D-Sucrose, ≥99,5 %, p.a.	Carl Roth	4621.1
D (+)-Glucose p. a., ACS, anhydrous	Carl Roth	X997.2
Dextrin for microbiology (from potato starch)	Carl Roth	3488.1
di-Potassium hydrogen phosphate trihydrate	Carl Roth	6878.1
di-Potassium hydrogen phosphate, ≥99 %, p.a., anhydrous	Carl Roth	P749.1
DNAse I	Sigma-Aldrich	10104159001
E-64	Sigma-Aldrich	E3132-1MG
EDTA disodium salt dihydrate, min. 99 %, p.a., ACS	Carl Roth	8043.2
Glycine	Serva	23391.02
HEPES PUFFERAN®, min. 99.5 %, p.-1 kg	Carl Roth	9105.3
Iron (III) sulphate hydrate, 80 %, pure	Carl Roth	0492.1
Isopropanol	Carl Roth	CP41.1
Leupeptin	Sigma-Aldrich	L2884-1MG
Magnesium chloride hexahydrate, min. 99 %, p.a., ACS	Carl Roth	2189.1
Magnesium sulphate heptahydrate, ≥99 %, p.a., ACS	Carl Roth	P027.1
Methanol	Carl Roth	4627.6
Milk powder	Carl Roth	T145.3
Pefabloc	Sigma-Aldrich	11585916001
Pepstatin A	Sigma-Aldrich	77170-5MG
Peptone ex casein	Carl Roth	8986.1
Phosphate buffered saline tablets (PBS)	Sigma-Aldrich	P4417
Potassium chloride min. 99.5 %, -1 kg	Carl Roth	6781.1
Potassium dihydrogen phosphate, ≥99 %, p.a., ACS	Carl Roth	3904.2
Precision plus protein all blue standards (marker)	BIO-RAD	161-0373
Sodium chloride 99,5 %, p.a., ACS, ISO	Carl Roth	3957.2

Sodium dodecyl sulfate (SDS)	Carl Roth	0183.2
Sodium nitrate, ≥99 %, p.a., ACS, ISO	Carl Roth	A136.1
TEMED	Carl Roth	2367.3
Tris	Carl Roth	AE15.2
Tris hydrochloride	Carl Roth	9090.2
Tween 20	Carl Roth	9127.1
Tryptone	Sigma-Aldrich	T7293
Yeast extract, micro-granulated	Carl Roth	2904.3

2.1.2 Equipment and instruments

Table 2. Main equipment and instruments used in the current study.

Instrument	Type	Company
Incubator	Heracell 150i	Thermo Fisher Scientific
Tabletop Centrifuge	Heraeus Megafuge 40 R	Thermo Fisher Scientific
Bead beater	FastPrep-24™ 5G	MP Biomedicals™
Ultracentrifuge	OPTIMA™ MAX-XP (TLA110)	Beckman Coulter
FPLC System	ÄKTA pure 25 M	Cytiva (GE Healthcare)
Plate reader	Epoch 2 Microplate Spectrophotometer	Agilent (BioTek)
Gel Imaging System	ChemiDoc™ MP Imaging Systems	Bio-Rad
Thermomixer	ThermoMixer C	Eppendorf
Glow Discharge Cleaning System	PELCO easiGlow™	Ted Pella, Inc.
Vitrification instrument	Vitrobot Mark IV System	Thermo Fisher Scientific
Microscope 1 (300 kV)	JEM-3200FS Field Emission Energy Filter Electron Microscope	JEOL
Camera (Microscope 1)	K2 IS Direct Detection Camera for low dose imaging	Ametek (Gatan)
Microscope (200 kV)	Thermo Fisher Scientific Glacios Cryo Transmission Electron Microscope (Cryo-TEM)	Thermo Fisher Scientific
Camera (Microscope 2)	Falcon 3EC Direct Electron Detector	Thermo Fisher Scientific

2.1.3 Model organism

Spores of *Chaetomium thermophilum var. thermophilum* La Touche 1950 (DSM No.: 1495, Type strain) were acquired from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH (DSMZ) in a freeze-dried ampoule and initially cultivated at the conditions proposed by the company and media (Medium 188 DSMZ Media List, 45 °C). As a backup, a pre-inoculated and grown organism petri dish was ordered from DSMZ.

Three different types of media were tested for optimal growth of the model organism those being: (a) DSM 188 using 2.00 g of $NaNO_3$, 0.01 g of $Fe_2(SO_4)_3$ x H_2O, 0.50 g of $MgSO_4$ x 7 H_2O, 0.14 g KH_2PO_4, 1.2 g K_2HPO_4, 0.02 g yeast extract and

15.00 g agar, (b) LB with 5.00 g yeast extract, 10.00 g tryptone, 10.00 g NaCl and 10.00 g agar and (c) CCM with 3.00 g sucrose, 0.50 g NaCl, 0.50 g $MgSO_4$ x 7 H_2O, 0.65 g K_2HPO_4 x 3 H_2O, 0.01 g $Fe_2(SO_4)_3$ x H_2O, 5.00 g tryptone, 1.00 g peptone, 15.00 g dextrin, 1.00 g yeast extract and 15.00 g agar. (Chemicals and enzymes used in the current thesis are listed in **Table 1**)

2.1.4 Antibodies

Production of specific antibodies against PDHc E1α-His-Tag (aa 37-411) and E2p-His-Tag (aa 29-459) were commissioned by GenScript (GenScript USA Inc., NJ) (**Table 3**). In short, the codon optimised sequence of the protein-region was cloned into a pET- 30a(+) vector in frame with a His-Tag. Protein was expressed in 1 L TB media. Proteins were obtained from the cell lysate supernatant with a two-step purification (Ni-NTA followed by Superdex 200 size-exclusion chromatography). Proteins were stored in PBS, 10 % Glycerol p*H* 7.4 (E2p-His-Tag) or PBS, 10 % Glycerol, 0.2 % SDS, p*H* 7.4 (E1α-His-Tag). Each rabbit (New Zealand rabbit) was immunised and an affinity-purified antibody was obtained from serum. Antibodies and recombinant proteins were subsequently shipped to *Kastritis* laboratory. Similar to those two, another batch of three antibodies against the other subunits of the PDHc was produced, namely being E1β, E3 and E3BP and described in **Table 3** as well. The storage buffer, in this case, was PBS, 1% BSA, 30% Glycerol. As a secondary antibody against all of the above-mentioned primary antibodies, Goat Anti-Rabbit IgG H&L (HRP) was acquired from Abcam and used according to the provider's instructions.

Table 3. Antibodies used in the current study.

Antibody	Source	Identifier
Rabbit polyclonal antibody a-E1α against *C. thermophilum* E1α-His-Tag (37-411)	Custom-made by Genscript	RRID: AB_2888984
Rabbit polyclonal antibody a-E2p against *C. thermophilum* E2p-His-Tag (29-459)	Custom-made by Genscript	RRID: AB_2888985
Rabbit polyclonal antibody a-E1β against *C. thermophilum* E1β-His-Tag (7-382)	Custom-made by Genscript	RRID: AB_2893234
Rabbit polyclonal antibody a-E3 against *C. thermophilum* E3-His-Tag (35-504)	Custom-made by Genscript	RRID: AB_2893235
Rabbit polyclonal antibody a-E3BP against *C. thermophilum* E3BP-His-Tag (34-442)	Custom-made by Genscript	RRID: AB_2893236
Goat Anti-Rabbit IgG H&L (HRP)	Abcam	ab205718

2.1.5 Kits

Table 4. Kits used in the current study.

Kit name	Source	Identifier
Gel Filtration Calibration Kit HMW	(Cytiva) GE Healthcare	28-4038-42
Pyruvate dehydrogenase activity assay kit	Sigma-Aldrich®	MAK183
α-ketoglutarate dehydrogenase activity assay kit	Sigma-Aldrich®	MAK189

2.1.6 Software and algorithms

Table 5. Table with software and algorithms used in the current study.

Software	Source	Identifier
3DFSC	[158]	https://3dfsc.salk.edu/
EMAN 2.0	[159]	https://blake.bcm.edu/emanwiki/EMAN2
Fiji	[160]	https://imagej.net/Fiji
Gctf	[161]	https://www.mrc-lmb.cam.ac.uk/kzhang/
Gen5™	BioTek Instruments	https://www.biotek.com/products/software-robotics-software/gen5-microplate-reader-and-imager-software/
HADDOCK	[115]	http://haddock.science.uu.nl/services/HADDOCK2.2
Image Lab Software 6.1	BIO-RAD	https://www.bio-rad.com/de-de/product/image-lab-software
MODELLER	[162]	https://salilab.org/modeller/
MotionCor2	[163]	https://emcore.ucsf.edu/ucsf-motioncor2
Needle	[164]	https://www.ebi.ac.uk/Tools/psa/emboss_needle/
Phenix	[165]	https://www.phenix-online.org
PyMOL	Schrödinger, inc	https://pymol.org/
Relion 3.0	[10]	https://github.com/3dem/relion
Scipion 2.0	[166]	https://scipion-em.github.io/docs/
SEGGER	[167]	https://www.cgl.ucsf.edu/chimera/
SPHIRE-crYOLO	[168]	http://sphire.mpg.de/
UCSF Chimera	[169]	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX	[170]	https://www.rbvi.ucsf.edu/chimerax/
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
Xmipp	[171]	https://github.com/I2PC/xmipp

2.2 Methods

The structural and biochemical characterisation of native cell extracts turned into an attractive but complex field, revealing challenges overshadowed by the many insights into endogenous interactions at unprecedented resolution. Here, a complete workflow for organism growth, harvesting, sample preparation, loading on a cryo-microscope for high-resolution data collection and subsequent data analysis steps until structure determination is described (**Figure 8**). The biochemistry part of this protocol was further optimised to decrease waiting times and substantially increase protein concentration without introducing aggregation while maintaining active macromolecular complexes intact for structural characterisation. Parts of the protocols and figures described in this thesis were adapted from [57,140].

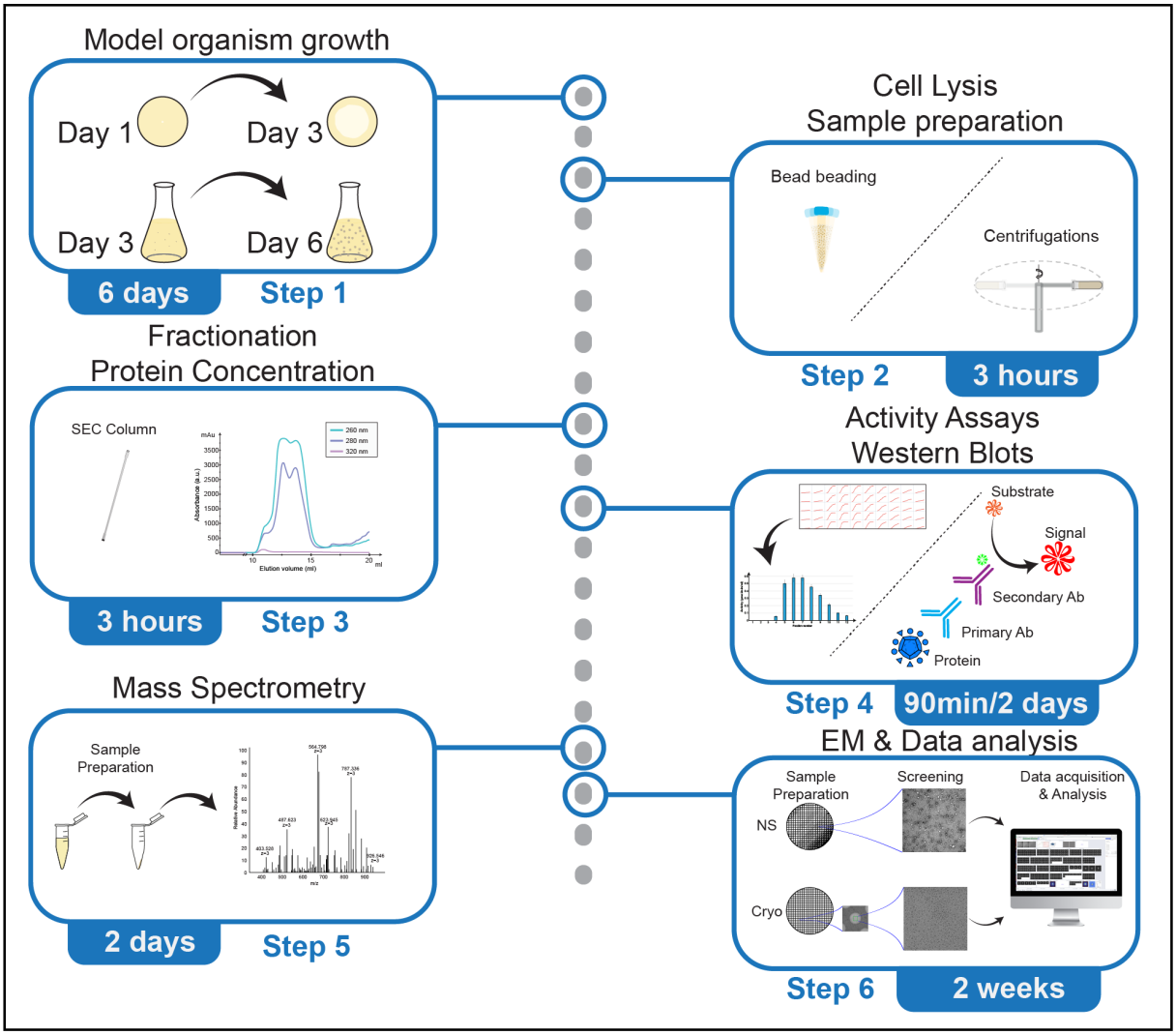


Figure 8. Graphical representation of the main steps of the workflow from model organism growth until data acquisition and analysis.

2.2.1 Model organism growth

2.2.1.1 Initial culturing

Spores of *Chaetomium thermophilum var. thermophilum* were initially cultivated at the conditions and media (Medium 188 DSMZ Media List, 45 °C) proposed by the company. As a backup, a pre-inoculated and grown organism petri dish was ordered from DSMZ. All inoculation steps were performed on a clean bench, under sterile conditions, as well as Bunsen light, to avoid contamination. Briefly, a double vial preparation (sealed under vacuum with silica gel and blue or red humidity indicator) was received from DSMZ. The tip of the ampoule was heated using a Bunsen burner flame and two drops of autoclaved water were placed onto the hot tip to crack the glass. The glass tip was stroked off using sterile forceps and insulation material was removed to approach the inner vial. The cotton plug was removed under sterile conditions, the top of the inner vial was flamed and 0.5 mL of the proposed medium was added and the pellet was rehydrated for 30 minutes. The content was gently mixed and approximately 200 µL were used for inoculating 188 DSMZ Media Agar plates. All remains of the original ampoule were autoclaved before discarding. (Detailed description of the whole process, including pictures, can be found in the CD attached to this thesis or through the following link:
https://www.dsmz.de/fileadmin/Bereiche/Microbiology/Dateien/Kultivierungshinweise/Kultivierungshinweise_neu_CD/Opening_17new.pdf).

2.2.1.2 Growth optimisation

Three different media types of agar-plates: DSM 188, LB and CCM (recipe described in Chapter 2.1.3), all made up of 1 L double distilled water and a p*H* adjusted to 7.1, were used to propagate mycelium from the initial culture. This was a step performed to further optimise the growth by titrating for different conditions: as of three different growth media for solid cultures (DSM 188, LB and CCM) and two different media for liquid cultures (LB and CCM) all grown at two different temperatures (45 °C and 50 °C), (**Figure S1**, **Appendix**). Mycelium was inoculated on freshly prepared plates of the three different media and grown in two temperature conditions (45 °C and 50 °C) for eight days (**Figure S1, Appendix**). Agar was excluded from the above recipes for LB and CCM liquid media. 500 mL Erlenmeyer flasks with 200 mL of each medium were inoculated using small pieces of freshly grown mycelium and incubated

for five days in two different temperature conditions (**Figure S1, Appendix**) under constant shaking at 100 rpm.

2.2.1.3 Routine agar plate culturing method and pre-culture inoculation

After determining optimal growth conditions, *Chaetomium thermophilum* was grown routinely at 45 °C in a rotary shaker at 100 rpm and 10 % CO_2. Petri dishes with ventilation and suitable for high-temperature growth conditions, including a layer of CCM-agar medium, were prepared regularly and kept at 4 °C until use. Before inoculation (~10 minutes), plates with CCM agar medium were warmed up in the incubator to equilibrate at 45 °C. A square-shaped mycelium of approximately 1 x 1 cm was cut using a sterile scalpel blade from the surface of a pre-grown plate and placed upside down in the middle of a new equilibrated plate. After inoculation, the pre-warmed (~45 °C plate) was placed in a plastic bag containing a wet towel (high humidity required), sealed and incubated until the CCM-agar surface was 95 % covered with mycelium (**Figure 9**). The growth temperature was gradually increased until the model organism adapted to thermophilic conditions and a temperature close to 50 °C.

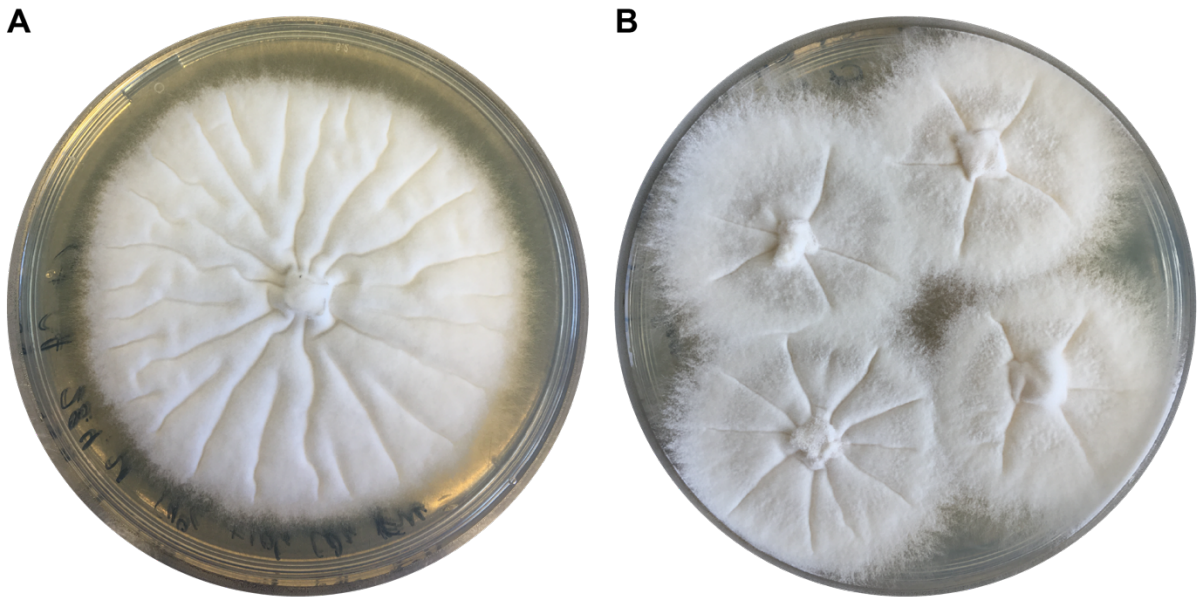


Figure 9. Representative solid cultures of *Chaetomium thermophilum* **on a CCM media agar plate.**
On the left **A.** one square piece of *Chaetomium thermophilum* grown mycelium was used for the inoculation of the plate, whereas on the right, **B.** four pieces of mycelium.

For small-scale liquid cultures, 400 mL of freshly prepared and autoclaved CCM media were poured in a 1 L Baffled Flask and inoculated with 5-6 small pieces (0.5 cm) of mycelium scraped off from a freshly grown plate. The residual agar part supporting the mycelium was removed and only the mycelium was used as inoculation

material. After 20 hours of growth, the grown mycelium was blended in 3 x 20 seconds rounds with 20 seconds of rest time in between. Using a standard kitchen blender is possible here, but cautious washing steps should be followed to avoid contamination.

2.2.1.4 Large scale liquid culturing and harvesting

For large scale mycelium growth, 20 mL of the blended pre-culture were used for the inoculation of a 2 L Baffled Flask containing 800 mL of CCM media and incubated for 20 hours (or until the mycelium balls formed in the liquid are spread out evenly and reach a size of around 1 cm) at 45 °C in a rotary shaker at speed 100 rpm and 10 % CO_2 supply. After 20 hours of growth, cultures were visually inspected for possible contamination as well as using a light microscope in bright field mode. Turbid media in between the spherically shaped colonies or low growth percentage could mean possible contamination and, in such cases, cultures were autoclaved before being discarded. Representative (1 L) flask (bottom view) of efficiently grown *Chaetomium thermophilum* is shown in **Figure 10**.

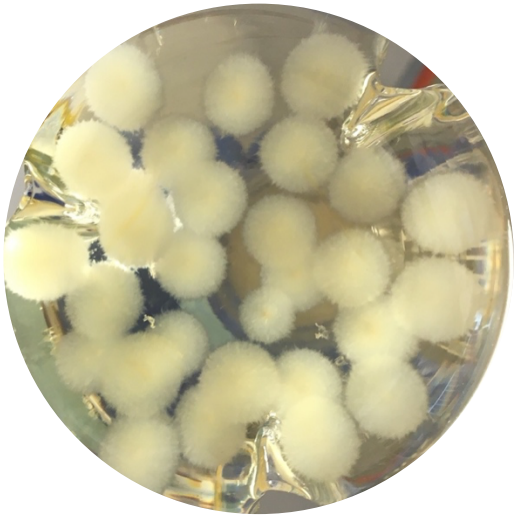


Figure 10. Representative "healthy" liquid culture growth of *Chaetomium thermophilum* **in CCM media 1 L baffled flask (picture is from the bottom side of the Erlenmeyer flask).**

A metal sieve (180 µm pore size) was used to strain the grown mycelium that was washed three times in ice-cold phosphate-buffered saline (PBS). Washing steps were performed using a precooled (at 4 °C) centrifuge and centrifugation speed of 2,200 *g*. Residual moisture was removed using a cell strainer. A prechilled mortar and liquid nitrogen (LN_2) were used to freeze ground mycelium producing pieces of approx. 0.5 cm by beading the frozen mycelium in the mortar. The material was either used fresh for subsequent experiments or stored in 50 mL falcon tubes at −80 °C until use. Again, it is very critical to point out that the CCM media at the point before harvesting the mycelium must look clear and transparent.

2.2.2 Cell lysis and protein purification/fractionation

The day before the experiment, Lysis and Running buffer (**Tables 6** & **7**) were freshly prepared and the column was washed with filtered and sonicated water overnight. The recipes for the buffers mentioned above are described in the tables below:

Table 6. Lysis buffer ingredients (1*X*).

Reagent	**Final concentration**	**Amount**
HEPES (2-[4-(2-Hydroxyethyl) piperazin-1-yl] ethane-1-sulfonic acid)	100 mM	5.96 g
NaCl	95 mM	1.39 g
Glycerol	5 % $^{w}/_{v}$	12.5 mL
KCl	5 mM	1.25 mL (1 M Stock)
$MgCl_2$	1 mM	1.25 mL (0.2 M Stock)
dd H_2O	n/a	n/a
Total	**n/a**	**250 mL**

Lysis buffer was adjusted at p*H* of 7.4 using 2 M NaOH, passed through 0.22 µm pore size filter and stored at 4 °C.

Table 7. Running buffer ingredients (1*X*).

Reagent	**Final concentration**	**Amount**
NH_4CH_2COOH	200 mM	3.85 g
dd H_2O	n/a	Up to 250 mL
Total	200 mM	**250 mL**

Running buffer was adjusted at a p*H* of 7.4 using 2 M NaOH, passed through 0.22 µm pore size filter, sonicated/degassed and kept at 4 °C overnight.

To save time between the steps of the experiment, the equipment and instruments required for at least the biochemistry part of the protocol were established and set up at the minimum distance possible. The goal of this step was to lyse the cells, get rid of the cell debris and membrane material and keep the high molecular weight complexes as intact as possible for further fractionation.

The morning before starting the experiment, the following preparations were performed:

- Equipment to be used like the benchtop centrifuge, the ultracentrifuge and the rotors were cooled down.
- The column (BioSep™ 5 µm SEC-s4000 500 Å, LC Column 600 x 7.8 mm) was equilibrated with the degassed and prechilled running buffer while the p*H* of the buffer that should be 7.4 was also double-checked, following the manufacturer's instructions (p*H* range: 2.5 - 7.5 and maximum backpressure: 1,500 psi).

- Protease inhibitors cocktail aliquots (described in **Table 8**) were thawed and freshly added to the lysis buffer.

Table 8. Protease inhibitors cocktail used in a total volume of 30 mL lysis buffer of this protocol.

Reagent	Final concentration	Amount
Pefabloc (AESBF)	2 mM	600 µL (100 mM Stock)
E-64	2 µM	60 µL (1 mM Stock)
Leupeptin	1 µM	3 µL (10 mM Stock)
Pepstatin A	1.5 µM	30 µL (1.45 mM Stock)
Bestatin	9.6 µM	100 µL (2.89 mM Stock)
Aprotinin	0.3 µM	30 µL (0.3 mM Stock)
DTT	1 mM	30 µL (1 M Stock)
DNase	0.01 $mg \cdot mL^{-1}$	30 µL (10 $mg \cdot mL^{-1}$ Stock)
EDTA	0.5 M	30 µL (0.5 M Stock)

(Stocks stored in single-use aliquots (to avoid freeze-thawing cycles) at –20 °C except EDTA)

When all the above steps have been prepared, exact volumes of the protease inhibitors cocktail described in **Table 8** were added in 30 mL of prechilled lysis buffer. Five 15 mL falcon tubes were filled with 10 mL zirconia beads and 5 mL buffer per tube and vortexed to fully hydrate the beads. In the protocol described here, 8 g of frozen mycelium were used as starting material, which was sufficient for achieving in the end concentrations in the fractions of interest optimal for cryo-EM experiments. 2 g of frozen material were added per tube (trying not to overpack the mycelium) and parafilm strips were used to seal the tubes. Falcon tubes were placed in a bead beater (step was performed in the cold room, 4 °C) and contents were bead-beated for 6 cycles of 20 seconds with 3 min in between cooling down intervals on the ice to avoid sample heating up. The lysate was cleaned from the beads through a 60 mL syringe with a Ø 1.1 x 50 mm, 19 G x 2" needle attached. A low-speed centrifugation step followed for 5 min at 4,000 *g* to precipitate large aggregates, non-lysed material and cell debris. The supernatant was carefully loaded in ultracentrifugation tubes, balanced and centrifuged for 45 min at 100,000 *g* using a TLA110 110K rpm (S/N 20U1718) fixed angle rotor. After the ultracentrifugation step, the supernatant was carefully passed through a 0.22 µm filter and the concentration of cell lysate was determined by Bradford assay. The cell lysate was spin filtrated and concentrated using (pre-equilibrated with lysis buffer) Amicon® Ultra-15 Regenerated cellulose filters of 100,000 NMWL cut-off and centrifugation rounds of 3,000 *g* for 15 min by constantly monitoring the protein concentration to a maximum final concentration of 30 $mg \cdot mL^{-1}$ or a volume of 600 µL. A small proportion (approx. 40 µL) of each step of the purification (from a to

h), as shown in **Figure 11** was kept for later kinetic validation and enrichment tracking of the molecules of interest.

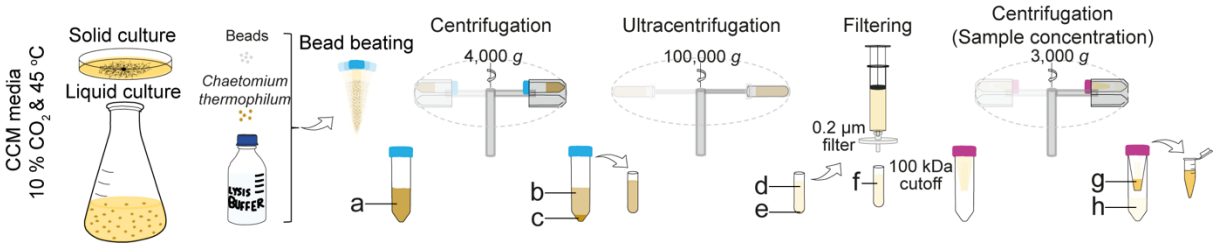
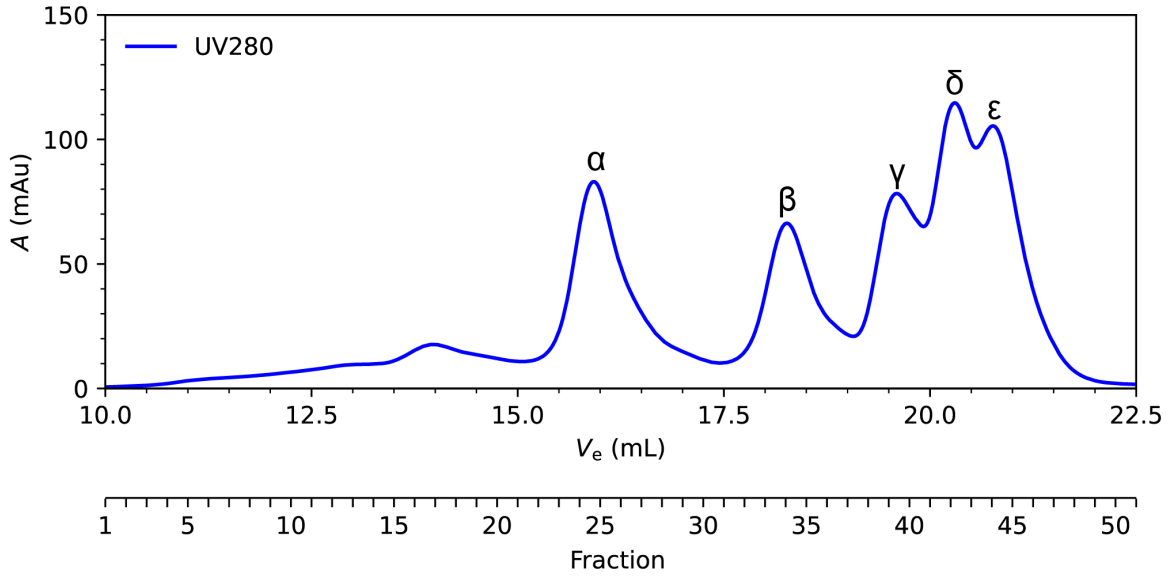


Figure 11. The experiment until the injection of the cell extract is shown with different steps of purification.
(**a**) Lysate after bead homogenisation; (**b**) and (**c**), lysate supernatant and pellet, respectively after the 4,000 *g* centrifugation; (**d**) and (**e**), lysate supernatant and pellet, respectively, after the ultracentrifugation step; (**f**) sample after filtration and (**g**) and (**h**), sample after final concentration and filtering, (**g**) being the material to be injected for size exclusion chromatography and (**h**) complexes of <100 kDa. (Source: Figure reproduced from [140])

The concentrated cell lysate was applied on the pre-equilibrated SEC S4000 Phenomenex column using a 500 µL loop attached to an ÄKTA pure 25M FPLC system. Size exclusion chromatography running parameters were as follows: Flow 0.15 mg·mL^{-1}, Pre-column pressure alarm 10 MPa, Detection absorbances 280 nm (proteins), 320 nm (aggregates) and 260 nm (nucleic acids). The run provided a relatively reproducible high-resolution separation of the cell extract in a range between 100 kDa to approximately 12 MDa high molecular weight complexes, as shown in **Figure 12** B.

A



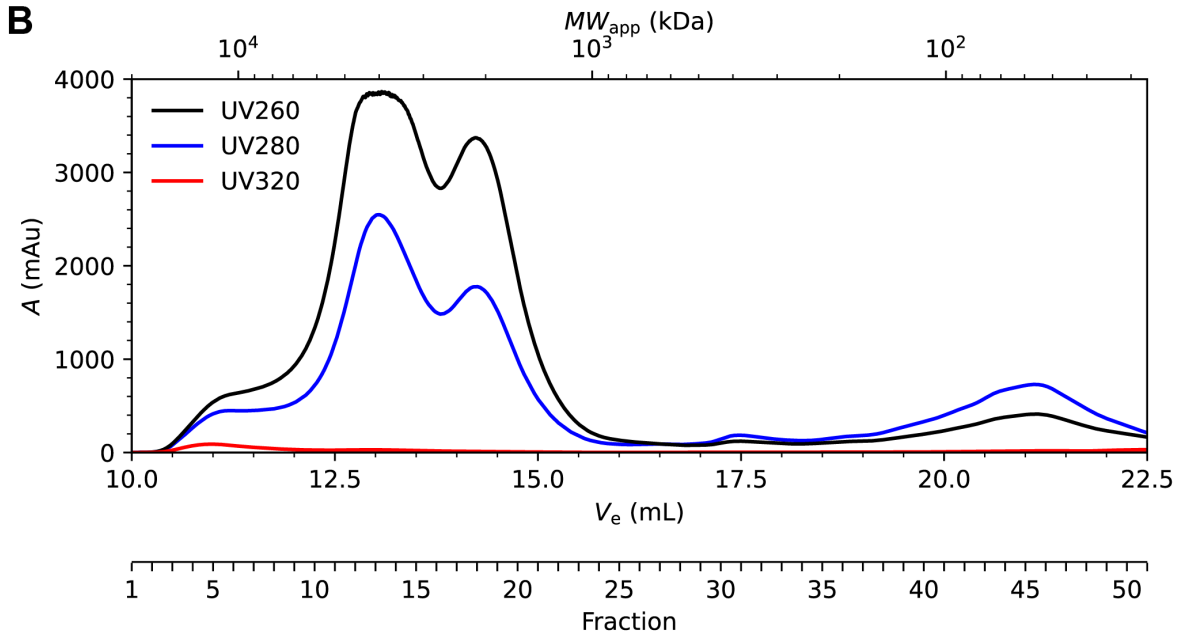


Figure 12. Representative gel filtration profiles of A. high MW calibration standards and B. *Chaetomium thermophilum* **native cell extracts.**
In **A.** the elution volume of the standards of Gel Filtration Calibration Kit HMW (high molecular weight calibration kit shown in **Table 4**), correlated with the fraction number for the column used (SEC S4000) is illustrated, where: (α) stands for the peak of thyroglobulin (669 kDa), (β) for ferritin (440 kDa), (γ) for aldolase (158 kDa), (δ) for conalbumin (75 kDa) and (ε) for ovalbumin (44 kDa), whereas in **B.** a representative elution profile of *Chaetomium thermophilum* native cell extracts using the same SEC column is shown. Absorbance detections at 280 nm (proteins, blue line), 260 nm (nucleic acids, black line) and 320 nm (aggregates, red line) are shown and a correlation between molecular weights (MW), elution volumes (V_e) and fraction numbers (Fraction) are illustrated.

In case the samples would be further processed for crosslinking mass spectrometry experiments (using for example, BS_3 as a crosslinker), the lysis buffer was used as an alternative running buffer instead of NH_4CH_2COOH to avoid lowering the efficiency of the crosslinking reaction.

2.2.3 Protein concentration determination

Protein concentration within the fractions was determined using 4 µL of each step of the experiment and each fraction (after the SEC column run) in combination with 240 µL of 1*X* Bradford solution (Stock solution Roti®-Quant 5*X*) to determine the protein concentration by measuring the absorbance at 595 nm using the BioTek Epoch 2 microplate spectrophotometer. As a guide for protein concentration determination, a standard curve with known concentrations of Bovine Serum Albumin (BSA) standards was freshly produced as a standard reference for total protein quantitation by this

colorimetric assay. A representative concentration estimation of the fractions collected during such a run is shown in **Figure 13**.

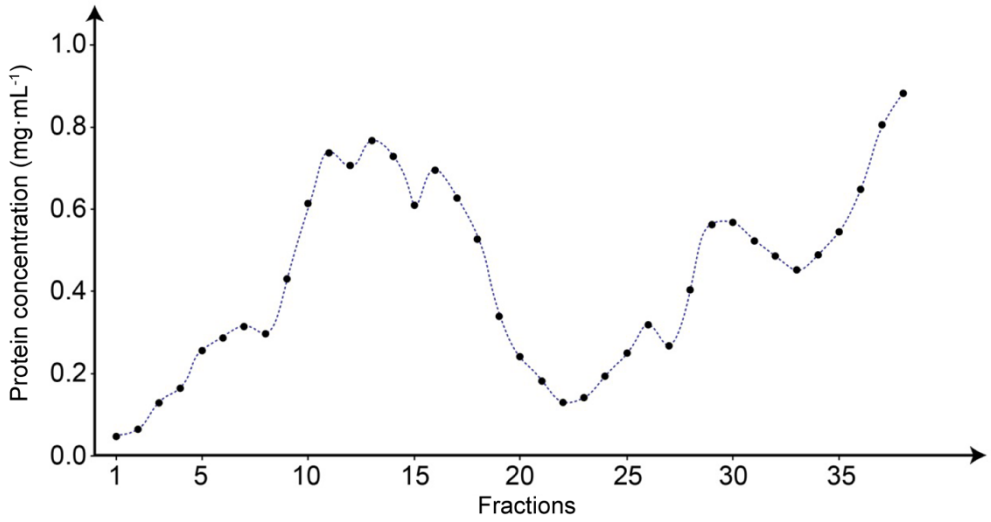


Figure 13. Representative Bradford measurement after fractionation.

At this step, the experiment could be paused. Fractions should be snap-frozen in liquid nitrogen (LN_2) and kept at −80 °C for further investigation. In the case of electron microscopy and mainly for the high molecular weight fractions, samples should preferably be prepared fresh (within the same day) to avoid the freeze-thaw step that might introduce aggregation.

2.2.4 Activity assays against PDHc and OGDHc

Activity assay kits to monitor kinetics for both pyruvate dehydrogenase and α-ketoglutarate dehydrogenase (**Table 4**) were executed for different steps of the experiment (**Figure 11**) as well as for the fractions of interest (**Figure 14**). Both experiments were performed after the provider's instructions. The activity was determined using a coupled enzyme reaction, which, as described in the protocol, results in a colorimetric (450nm) product relative to the enzymatic activity present. In detail for both, Pyruvate Dehydrogenase and α-ketoglutarate Dehydrogenase activity assays of the different steps of the experiment (**Figure 14**) were performed. For each step, 2 µL were used and tested for activity. The pellets of steps (c) and (e) of **Figure 11** were resuspended in 800 µL of lysis buffer and 2 µL were used per well. 0, 2, 4, 6, 8 and 10 µL, respectively, of the 1.25 mM (1.25 nmole·µL^{-1}) NADH standard solution (provided within the kit) were added (in duplicate) in a 96-well plate to generate 0 (blank), 2.5, 5, 7.5, 10 and 12.5 nM, respectively, per well standards.

As already described, high molecular weight complexes involved in the citric acid cycle such as PDHc and OGDHc was the main focus and according to the elution volumes of the standards shown in **Figure 12** A, they were expected to elute in the early fractions (1-12) of the gel filtration profile shown in **Figure 12** B. Thus, it was decided to focus on those fractions for subsequent screening experiments (*e.g.,* activity assays, immunoblotting experiments *etc.*) to enrich for the abundance of the complexes of interest. Also, the steps of the experiment (a-h) described in **Figure 11** were screened in terms of activity and as a way to optimise the protocol and enrich for the complexes of interest (**Figure S2**). For the high molecular weight fractions 1-12, the activity per fraction was tested in technical duplicate for each biological triplicate experiment (I, II and III) shown in **Figure 14**. For each fraction, 2 µL (sample volume) out of the 250 µL of the fraction were used. One of the low molecular weight fractions (fraction 29 mentioned as Fr29) was used as negative control in both assays. PC I and PC II contained a sample volume of 5 µL and 2 µL of the positive control provided within the kit. Calculations were performed following the protocol's instructions and only values obtained within the linear range of the NADH standard curves (**Figure 14**, bottom right) were used for the overall activity calculations. The highest average obtained value of NADH standard duplicates (both for PDHc and OGDHc) were used to plot a standard curve out of this curve, the equation for the NADH generated according to the absorbance was excluded. Correction for the background by subtracting the final measurement [$(A_{450})_{final}$] of the standards and samples. The alteration in absorbance measurement from, $T_{initial}$ to T_{final} for the samples was calculated ($\Delta A_{450}=(A_{450})_{final}-(A_{450})_{initial}$). Using the corrected measurements, the produced NADH ($nM \cdot min^{-1} \cdot mL^{-1}$) was calculated using the following equation:

$$nM \cdot min^{-1} \cdot mL^{-1}(milliunits \cdot mL^{-1}) = \frac{S_a}{(T_{final}-T_{initial})SV}$$

where S_a corresponds to the difference in NADH amount generated based on the equation among the initial and final time point per well and SV stands for the sample volume in mL.

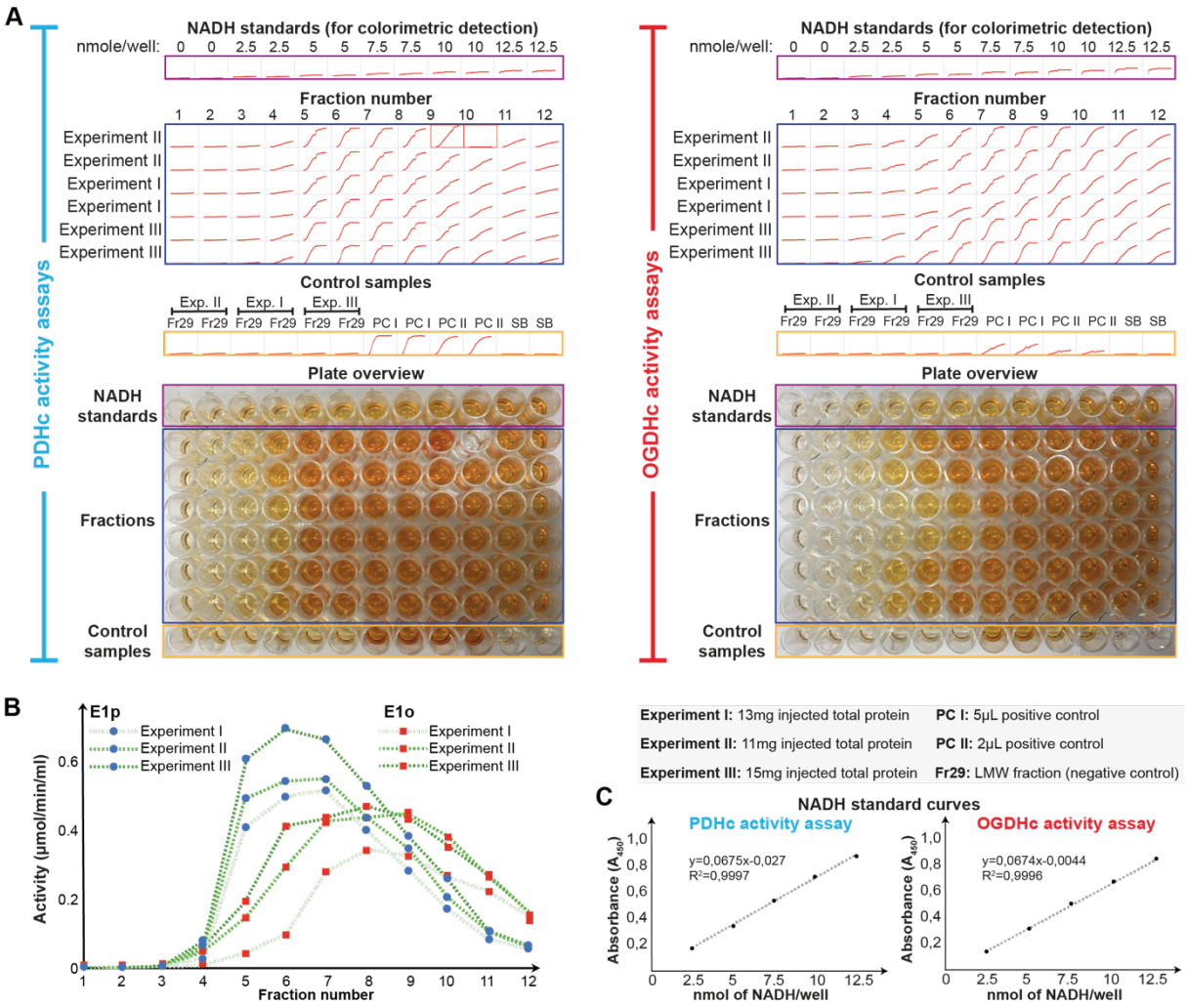


Figure 14. Representative kinetic assays for PDHc E1p and OGDHc E1o subunits. **A.** On the left part of the image, activity assays for PDHc are shown. On the top NADH standards are demonstrated for the triplicate assays within the fractions (fractions 1-12). Original spectrophotometric curves are shown in technical duplicates of biological triplicate assays for the high molecular weight fractions (fractions 1-12); control samples are also shown and below controls, a picture of the plate with the colouration of the corresponding wells, showing successful reactions. A similar illustration is depicted on the right for the kinetic assays performed in the fractions for the E1o. SB stands for sample blank. **B.** On the bottom, the activity of PDHc (blue) and OGDHc (red) is shown, calculated in milliunits·mL^{-1} for every replicate. **C.** The linear NADH standard curves are shown, used in the PDHc and OGDHc activity analysis. (Source: Figure reproduced from [140])

2.2.5 Immunoblotting experiments

For the immunoblotting experiments in-house, casted gels were prepared following the recipes described in **Tables 9** and **10**. All glassware and equipment used were thoroughly washed and the thickness of all gels was 1 mm.

Table 9. Stacking phase gel ingredients.

Reagent	Final concentration	Amount
Acrylamide/Bis solution, 37.5:1	10 %	3.34 mL (Stock 30 % $^{w}/_{v}$)
Tris-HCl p*H* 8.8	0.37 M	2.46 mL (Stock 1.5 M)
Sodium dodecyl sulfate (SDS)	0.1 %	50 µL (Stock 20 %)
APS	0.04 %	40 µL (Stock 10 %)
TEMED	13.4 mM	20 µL (Stock 6.71 M)
dd H_2O	n/a	4.1 mL
Total	**n/a**	**10.01 mL**

Table 10. Separating phase gel ingredients.

Reagent	Final concentration	Amount
Acrylamide/Bis solution, 37.5:1	5.08 %	0.85 mL (Stock 30 % $^{w}/_{v}$)
Tris-HCl p*H* 6.8	0.5 M	1.25 mL (Stock 0.5 M)
Sodium dodecyl sulfate (SDS)	0.1 %	25 µL (Stock 20 %)
APS	0.04 %	20 µL (Stock 10 %)
TEMED	13.4 mM	10 µL (Stock 6.71 M)
dd H_2O	n/a	2.86 mL
Total	**n/a**	**5.02 mL**

Samples were incubated for 5 minutes at 100 °C under shaking after being mixed with a 4*X* loading dye (250 mM Tris-HCl (p*H* 6.8), 8 % $^{w}/_{v}$ SDS, 0.2 % $^{w}/_{v}$ bromophenol blue, 40 % $^{v}/_{v}$ glycerol and 20 % $^{v}/_{v}$ β-mercaptoethanol in 50 µL aliquots kept at –20 °C). Samples were then centrifuged on a tabletop centrifuge for one minute at full speed and 5 µL per sample were loaded in each lane accompanied with 5 µL of Precision Plus Protein™ All Blue Prestained Protein Standards (Biorad #1610373) for each set of samples. Polyacrylamide gels were electrophoresed at an electrical field of 100 V for approximately 2 hours, in a 1*X* electrophoresis buffer (100 mL of 10*X* electrophoresis buffer stock solution, 5 mL of 20 % SDS stock solution and 895 mL dd H_2O) freshly prepared from a 10*X* electrophoresis buffer stock solution (0.25 M Tris, 1.92 M Glycine). Gels were then transferred (in pairs) to a nitrocellulose membrane in a setup similar to the one shown in **Figure 15** in a pre-set protocol of 25 V (1A) applied field, for 30 minutes using a Trans-Blot® Turbo™ Transfer System of BioRad. Whatman paper stacks were hydrated in 1*X* transfer buffer (100 mL of 10*X* electrophoresis buffer stock solution, 200 mL methanol, 2 mL of 20 % SDS stock solution and 698 mL dd H_2O) and possible air bubbles were removed from the in-between of the layers of the transfer pack.

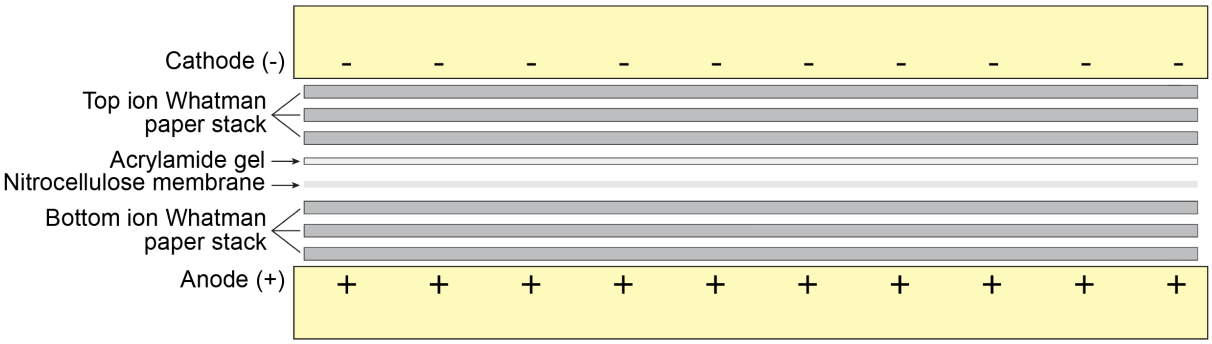


Figure 15. Layering of the assembled transfer pack.

Membranes were then blocked in 5 % $^{w}/_{v}$ milk/Tris-buffered Saline Tween-20 (TBST) solution for one hour under mild shaking and then incubated under gentle shaking for 16 hours at 4 °C with the primary antibody (Ab) that was diluted and freshly prepared in 2 % $^{w}/_{v}$ milk/TBST solution. 1*X* TBS buffer was freshly prepared using a 10*X* stock solution (200 mM Tris, 1.5 M NaCl, p*H* 7.6 with 12 N HCl and stored at 4 °C for up to 3 months) and the addition of 0.1 % Tween 20. Concentrations of applied Ab per case were titrated and optimised for all the antibodies (primary and secondary) used. The next day the primary Ab was removed and three washing steps of 10 min each with 2 % $^{w}/_{v}$ milk/TBST solution were performed. The secondary Ab (in 2 % $^{w}/_{v}$ milk/TBST solution) was applied and incubated for 1 hour. All membranes were washed two times (10 min each wash) with 2 % $^{w}/_{v}$ milk/TBST and one time (5 min) with TBST buffer without milk. Finally, the membranes were imaged using a ChemiDoC MP Imaging system with a freshly prepared ECL fluorescent mixture under optimal exposure time conditions.

2.2.6 Preparation for Mass Spectrometry Analysis

Identification of the protein of interest was necessary for the next steps of this project. Fractions were precipitated after protein concentration determination to a specific protein concentration per tube using an acetone precipitation protocol described below:

Four volumes of cold (−20 °C) acetone was added to one volume of each sample, vortexed and kept for 60 minutes (or overnight) at −20 °C. Centrifugation at 15,000 × *g* was performed and the supernatant was disposed of by not interrupting the protein pellet. Acetone was allowed to evaporate from the tubes at room temperature for approximately 20 minutes but not over-dried as it would not properly dissolve. Samples were shipped to the collaborator's laboratory under dry ice for further analysis.

When crosslinking mass spectrometry experiments were performed, the crosslinker used was (bis(sulfosuccinimidyl)suberate) suberate (BS_3) and the elution buffer during gel filtration was modified as described above (**Chapter 2.2.2**). Also, the crosslinker was titrated in concentrations from 0 up to 5.0 mM (representative SDS Coomassie-stained gels shown in **Figure 16**) before determining the final concentration for the efficiency of the reaction.

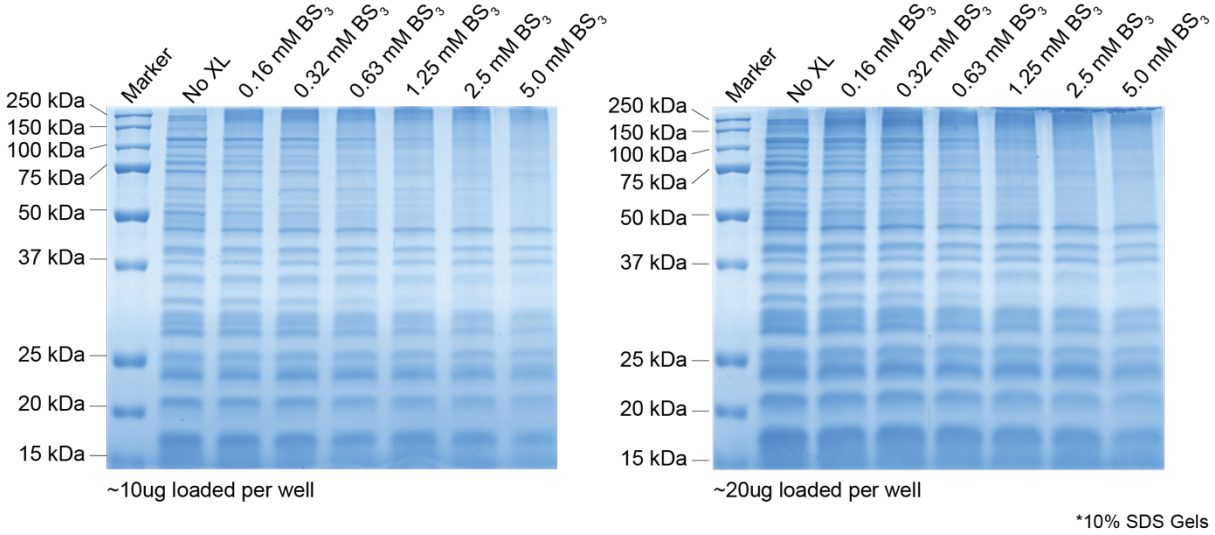


Figure 16. Representative Coomassie-stained gels of titrated crosslinker concentration.

After deciding on the optimal crosslinker concentration, samples were incubated with the freshly diluted crosslinker on ice for 2 hours in the dark. The reaction was then quenched using 50 mM ammonium bicarbonate for 30 min on ice. Acetone precipitation as described above was performed and samples were shipped to the collaborator's laboratory under dry ice for further analysis.

2.2.7 Sample preparation for electron microscopy

2.2.7.1 Negative stain grid preparation

One screening method that guides the high-resolution preparation methods is the experiment of negative stain. This experiment aims to screen the sample at low resolution, determine the concentrations used in the high-resolution method preparation of cryo-EM and, in general, have a first view of the complexes abundant in the sample. Because protein complexes are stained with a heavy salt, their observation in the electron microscope is easier due to the increased contrast; However, the protein complexes within the fractionated extract are stained and therefore, internal features of the complexes are lost.

Grids were applied to a glass side with their carbon-coated side facing to the top and then glow discharged just before starting the experiment (carbon support side on the top) in a PELCO easiGlow™, at 0.4 mbar and 25 seconds glowing time using 15 mA. Reverse tweezers were properly labelled and used to hold the glow-discharged grids conveniently. Each sample, in a volume of 3.5 µL and optimal concentration (~0.003 mg·mL^{-1}), was applied on the carbon side (top) of the grid for 45 seconds before being blotted out by using pre-cut pieces of blotting paper. Blotting was always applied from the same position/corner of the grid. Two times washing step with 15 µL of dd H_2O was performed. After blotting the water (the same way and direction as the sample), 15 µL of 2 % uranyl acetate (the staining solution) was applied and incubated for 1 min on the carbon side of the grid. Finally, the staining agent was blotted away and grids were let to dry overnight, covered with a lid (to avoid contaminations). The complete process is graphically represented in **Figure 17**.

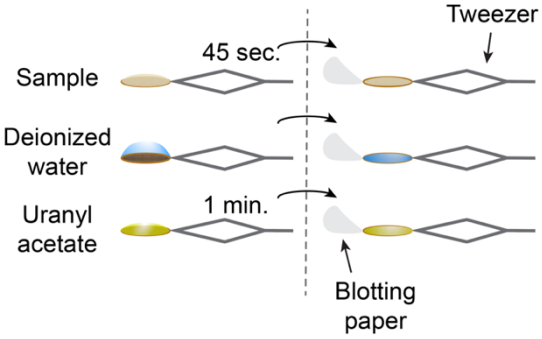


Figure 17. Step description of negative stain experiment

2.2.7.2 Cryo-electron microscopy grid preparation (vitrification)

For the cryo-electron microscopy grid preparation, carbon-coated holey support film type R2/1 on 200 mesh copper grids from Quantifoil® were used. Using a PELCO easiGlow™, grids were glow discharged at 15 mA, 0.4 mbar and 25 seconds glowing time. Then 3.5 µL of the sample/fraction at a total protein concentration of 0.3 mg·mL^{-1} was applied on each grid. Grids were plunge-frozen with a Vitrobot® Mark IV System (Thermo Fisher Scientific) using standard Vitrobot Filter Paper (Grade 595 ash-free filter paper ø55/20 mm) that was loaded in the Vitrobot just before starting the process. Conditions in the chamber were kept stable at 4 °C temperature and 95 % humidity during the whole process. For grid plunging, blot force of 2 and blotting time of 6 seconds were applied. This blot force and blot time were benchmarked in similar fractions and adjusted to optimal conditions for the ice thickness required for the protein complexes investigated in the fractions of interest. A representation of the entire process of plunge freezing of the samples with the Vitrobot is shown in **Figure 18**. The vitrified grids were carefully clipped according to the standard procedure of Thermo Fisher Scientific and loaded onto the Thermo Fisher Scientific Glacios 200 kV

Cryo-transmission electron microscope in the *Kastritis* laboratory under cryo and low humidity conditions.

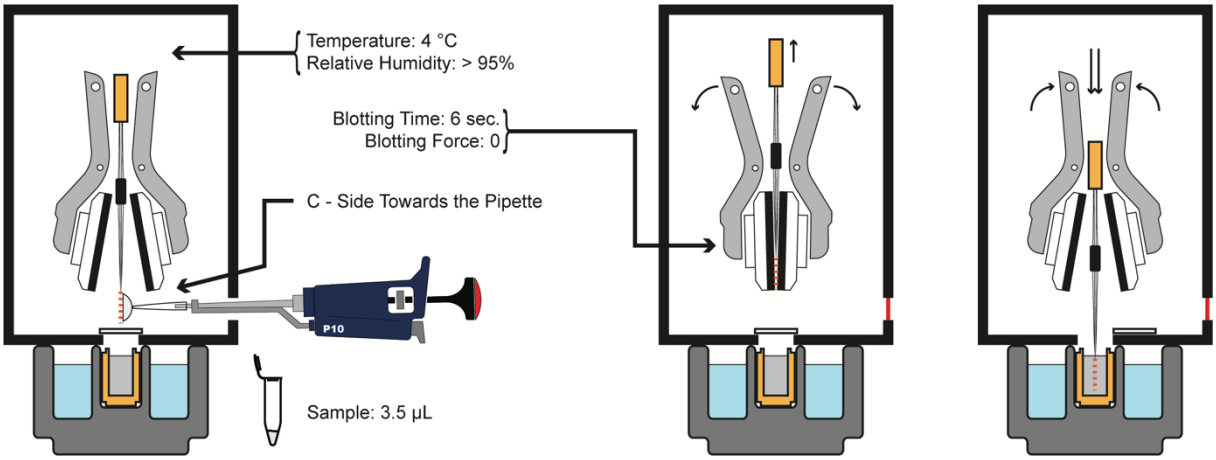


Figure 18. Graphical representation of the main steps of plunge freezing carried out using the Vitrobot® Mark IV System.
(Thermo Fisher Scientific) (Courtesy of Farzad Hamdi, MLU Halle-Wittenberg)

2.2.8 Electron microscopy

2.2.8.1 Grid screening

After loading samples on the electron microscope, an overview at low magnification was acquired to judge the quality of the sample and find the candidate grid-squares for further investigation (**Figure 19**). After the quality of grids in low magnification (510*X*) was determined, the screening process was continued in steps from lower to higher magnifications to evaluate for the distribution of the particles, concentration, contamination level (*e.g.*, crystalline ice, ethane contamination), enrichment levels for particles of interest, the thickness of the vitreous ice in the vitrified samples and other parameters of interest in many different locations of the grid.

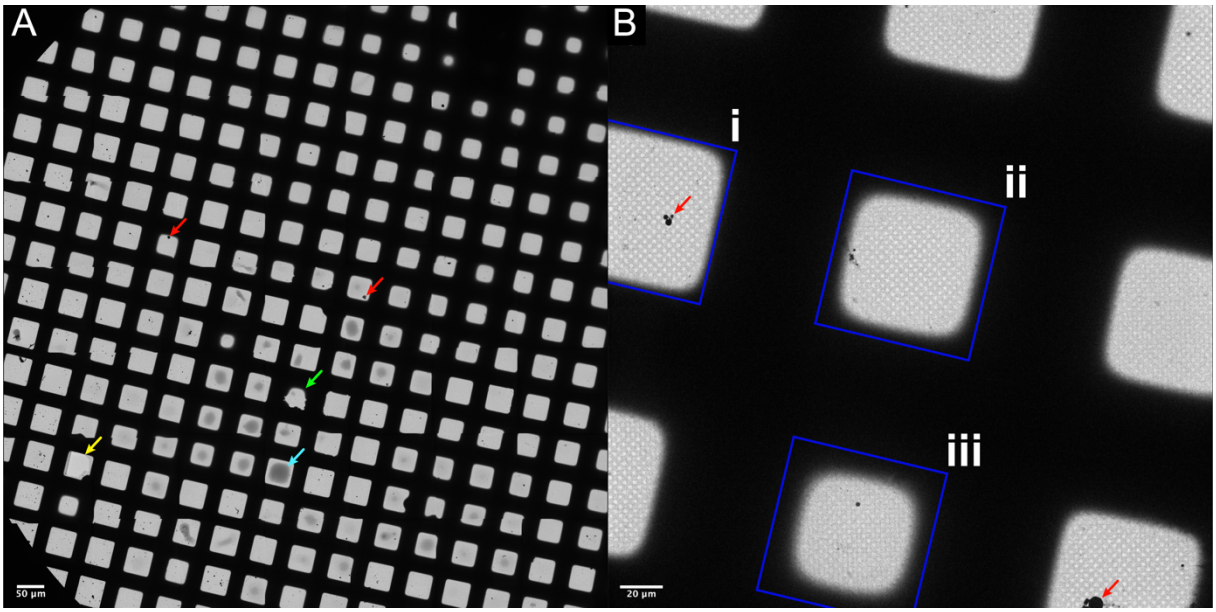


Figure 19. View of a representative carbon-coated holey support film type R2/1 (200 mesh) copper grid from Quantifoil® after the Atlas has been acquired.
A. View of the full Atlas of the grid in lower magnification with a gradient of ice thickness starting from the top right (thicker area) till the bottom left side (thinner area). In red arrows, ice contamination is indicated, cyan arrow points to areas where the buffer was not blotted away efficiently, resulting in locally thicker ice, yellow arrow indicates a broken grid square where the carbon film was mechanically damaged during manipulations of the grid and green arrow a grid square with non-canonical shape maybe due to manufacturing reasons. **B.** View of a single tile out of the full Atlas in higher magnification with the grid holes inside the grid boxes being visible. Red arrows, same as in (A) indicate ice contamination, whereas i, ii and iii represent three grid squares (highlighted in blue frames) with different levels of ice thickness.

2.2.8.2 Negative staining data acquisition

After acquiring the atlas view and screening, a protocol initially made for lacey cryo-EM grids (Quantifoil®) was used to acquire negative staining samples. In this protocol, the microscope controlling software focuses on a selected spot, then moves to the following neighbouring location, takes an image and focuses on the place where it acquired the last image and repeats the cycle. The software for implementing this protocol was Thermo Fisher Scientific EPU Version: 2.11.1.11REL[172].

A set of suitable grid-squares were selected for running this protocol and a local map was acquired after rough eucentric height adjustment. Then, inside the grid square, a set of acquisition spot grids were generated and then the protocol mentioned above was performed on each point. The acquisition parameters were selected based on the nature of the sample and compromised based on the abundance, particle size, desired resolution, *etc*.

2.2.8.3 Cryo-EM data acquisition

Like negatively stained samples, after acquiring the atlas view and screening, a protocol designed to acquire low dose Cryo-EM images from Quantifoil® cryo-EM grids was used. In this protocol, the software finds and centers a hole (**Figure 20** B) within the desired region of interest, focuses on the surrounding carbon area and finally takes a picture in the adjusted location within the hole. The software used for implementing this protocol was Thermo Fisher Scientific EPU Version: 2.11.1.11REL[172].

Similar to the negative staining protocol, a set of suitable grid-squares were selected and a local map was acquired after rough eucentric height adjustment. In the acquired local map, the locations of the holes were found using an automated hole finding function in the software. After selecting all holes, a template for defining the focusing and data acquiring regions according to the holes' center was defined (see **Figure 20** C. Like negatively stained samples, the acquisition parameters were adjusted in each data set based on the nature of samples, *e.g.,* the abundance of particles of interest, particle size, desired resolution, *etc*.

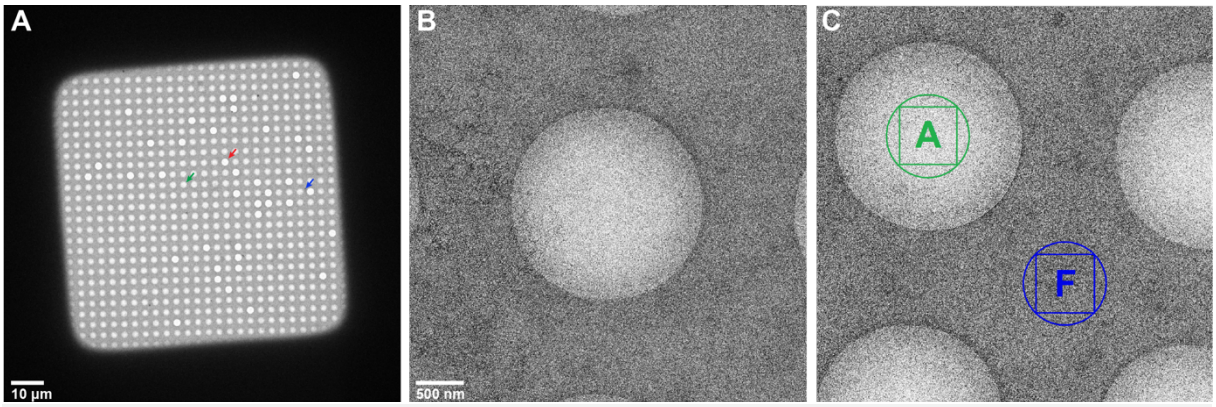


Figure 20. View of a representative carbon-coated holey support film type R2/1 (200 mesh) copper grid from Quantifoil® after the Atlas has been acquired.
A. View of an entire grid square. A "popped" hole (no ice) is indicated with the red arrow, with a green arrow showing a typical hole with vitreous ice, whereas the blue arrow points to possible ice contamination. **B.** Zoom in to one of the grid holes of (A) and **C.** template definition where with the "A" letter (green), the acquisition area inside a hole is indicated, whereas, with the "F" letter, the focusing area is shown. Each grid hole has a 2 µm diameter.

2.2.8.4 Microscope adjustment and alignment

Before starting the protocols mentioned above, a series of alignments and adjustments were performed on the electron microscope. These steps ensure a centered, coherent and symmetric beam around the microscope's optical axis. All the alignments were done on the "eucentric focus", meaning the objective lens had finely focused on the eucentric plane. To achieve this situation, a part of the sample was brought to the eucentric height and then the beam was focused on the sample plane.

The basic alignment of modern electron microscopes like the 200 kV electron microscope Glacios, are extremely stable and because of this stability a fine tune just before starting acquisition is required. Just like full alignment of any electron microscope this partial fine alignment is done from top to bottom of the microscope. These alignments consist of: (1) Beam centering, (2) Real space pivot points, (3) Condenser aperture centering, (4) Condenser lens astigmatism, (5) Current centering, (6) Objective lens astigmatism, (7) Coma-free adjustment and (8) Parallel beam condenser beam adjustment.

It is worthy of note that all the alignments as mentioned above were done according to the Thermo Fisher Scientific workflow (Thermo Fisher Scientific (2019) Single Particle Analysis Workflow). One may consult with the references [173,174] for more details about these adjustments. This procedure is schematically presented in **Figure S3** in the **Appendix**.

After these alignments, a constant dose rate is expected. But the dose rate must be measured just before data acquisition and the exposure time should be adjusted based on the dose rate and the final desired cumulative dose. In all acquisitions presented in this work, the data were collected in a movie format, with a dose of more than 1 $e \cdot Å^{-2}$ and more than one frame (fraction) per second.

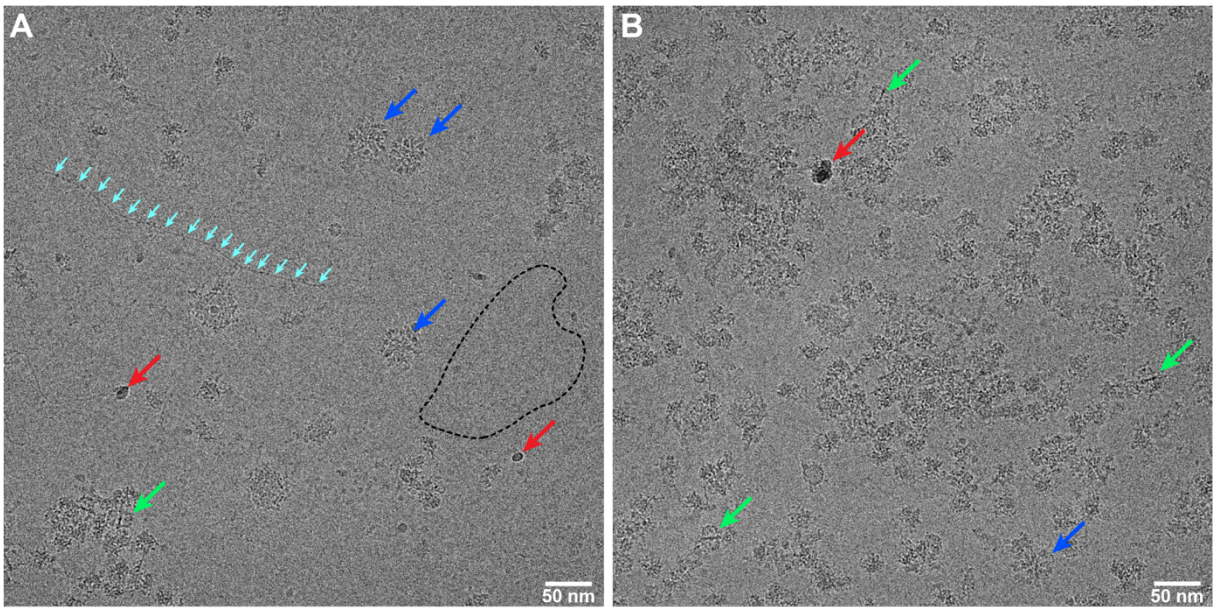


Figure 21. Representative micrographs of low A. and higher B. concentration samples from the same fraction and two different experiments.
In both **A.** and **B.** blue arrows the protein complex of interest (PDHc) is shown, red arrows show possible ice particles (contamination), green arrows point to FAS with possible binders, cyan arrows point to a filamentous like structural signature (maybe nucleic acids) and the black dotted area vitreous ice background.

2.2.9 Data analysis

2.2.9.1 Image processing

Overall, image processing of micrographs derived from (cryo-)electron microscopy observations aim to resolve structural signatures by focusing on deriving a 3D reconstruction of a molecule of interest. Steps are shown in **Figure 22**, which start with the movies acquired at the microscope level, under low dose and end with a validated cryo-EM map. Below the specific steps in image processing followed for reconstructing the PDHc and OGDHc maps from cryo-EM data are shown; Terminology and image analysis theory can be found in [175] and references therein. For negative staining data, a highly similar protocol, skipping the steps of contrast transfer function estimation and per-particle refinements due to the nature of the negative stain data was followed.

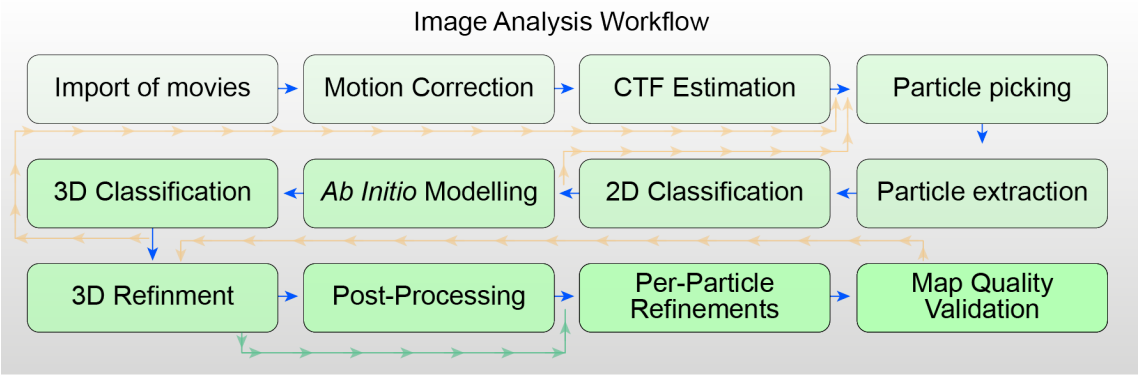


Figure 22. Main steps of image analysis workflow.
Steps of image analysis are illustrated starting with the movies acquired at the microscope level, under low dose, up until a validated cryo-EM map is generated.

2.2.9.2 Negative stain data image processing and volumetric calculations

All the collected micrographs of negatively stained fractions were processed in Scipion 2.0 image processing framework [166]. Scipion 2.0 is a meta-analysis software for processing microscopy images and includes various distinct software for each step in the image processing pipeline of electron micrographs and the imaged particles. The micrographs were collected in low dose mode and as single images for each fraction. Particle picking was performed using the combination of Xmipp-particle picking [171] and SPHIRE-crYOLO protocols [168]. The 2D, 3D classifications and 3D auto-refinement analysis were performed using RELION 3.0 [10] in Scipion. The selected micrographs were manually inspected to remove *e.g.,* highly contaminated or drifted ones and particles from datasets corresponding to the investigated fractions were then classified by iterative 2D classification (initial particle numbers in **Table S2 Appendix**).

Particularly for the studied fraction, a subset of particles exhibiting cubic or icosahedral symmetries were extracted and used for subsequent 3D classification without imposing symmetry. For 3D classification, 6CT0 and 1EEA models were low pass filtered using EMAN [159] and used as references for asymmetric reconstructions of PDHc and hybrid OGDHc/BCKDHc, respectively. The resulting subsets of particles for PDHc and hybrid OGDHc/BCKDHc per each fraction from 3D classes exhibiting distinguishable core characteristics were then subjected to 3D refinement. A spherical mask of 600 Å for PDHc and 450 Å for hybrid OGDHc/BCKDHc was applied to ensure that the full metabolons were included in the reconstruction. For fraction 5, additional runs of 3D refinement were performed by imposing I or O symmetry for PDHc or hybrid OGDHc/BCKDHc, correspondently, and spherical masks of 300 Å and 250 Å respectively. These smaller masks were applied to reconstruct each complex's E2 core. Resulted unprocessed, unmasked half-maps of PDHc/OGDHc per fraction were then uploaded to the EBI FSC server (https://www.ebi.ac.uk/pdbe/emdb/validation/fsc/) for unambiguous resolution estimation. Reported resolutions were based on an FSC threshold of 0.5 (**Table S2**).

For volumetric calculations, negative stain reconstructions of PDHc and hybrid OGDHc/BCKDHc from fraction 6 were thresholded at 0.07 and 0.066 σ values, respectively. Then, the SEGGER software [167] in Chimera was used, followed by merging of the segmented volumes of the core. Afterwards, the volume of each segment was measured using Chimera and plotted in a 100 % plot.

2.2.9.3 cryo-EM Data image processing

Image processing was performed using Scipion 2.0. Two datasets from 2 biological replicates (Dataset 1 and Dataset 2) were collected for fraction 6 and another dataset for fraction 4 (Dataset 3). Collected movies of all datasets were corrected for beam-induced motion using MotionCor2 [163] (**Table 5**), resulting in a dose-weighted, motion-corrected set of summed images. CTF estimation was performed with Gctf [161], using the dose-weighted set and used for the rest of the analysis steps. All subsequent analysis steps were performed using Xmipp3 and RELION 3.0 (**Table 5**). Global resolutions mentioned for Datasets 1 and 2 were calculated in RELION 3.0 and sphericity was evaluated by 3DFSC web server [158] (**Table 5**).

For Dataset 1, a total number of 513 particles were manually picked with Xmipp3, extracted and used to generate initial 2D classes that were afterwards used as templates for automated picking using RELION 3.0. This resulted in 126,739 particles

that were subjected to reference-free 2D classification. The best resulting classes were used as templates for a new round of refined auto-picking, leading to a final set of 99,434 particles that were then again extracted and underwent reference-free 2D classification. All classes that included falsely picked or junk particles were manually inspected and discarded. The remaining 89,263 particles were subjected to reference-based 3D classification into 5 classes without imposing symmetry. PDB ID: 6ct0 (**Table S8**) [145] was used as a reference after being lowpass-filtered to 40 Å. Again, particles belonging to junk classes were discarded and the remaining 12,103 particles were subjected to reference-free 2D classification and selected for quality, ending up with a final set of 9,795 particles that were used for all subsequent 3D refinement steps. Three different refinement procedures were performed to derive an icosahedral core of PDHc, an asymmetric core of PDHc and an asymmetric reconstruction of the full complex. Briefly, the first reference-based, by imposing Icosahedral symmetry and a mask isolating the PDHc core, resulted in a final core model of an overall resolution of 7.56 Å (FSC=0.143). The second refined this data without imposing symmetry, resulting in a core model of 12.03 Å (FSC=0.143). In the third, an initial model of the complete complex was generated with RELION 3.0 and used as a template for 3D refinement of the entire PDH complex after low pass filtering, leading to a model with an overall resolution of 34.52 Å (FSC=0.5).

For Dataset 2, the final particle set of 9,795 particles was reference-free 2D classified. The class averages were then used as templates for automated picking after low pass filtering, resulting in 105,817 particles which, after sorting for miss-picks and junk via 2D reference-free classification and 3D template-based classification, lead to a final set of 19,721 particles. These were used for all subsequent 3D refinement steps with the same parameters as Dataset 1. This time the three refinements lead to a symmetric core model of an overall resolution of 7.10 Å (FSC=0.143), an asymmetric core model of 8.82 Å and a complex model of 22.69 Å (FSC=0.5). The highest quality particles of both datasets were finally combined to a set of 29,516 particles and used for 3D refinement steps, once again with the same parameters as previously mentioned. This resulted in a final symmetric core model with an overall resolution of 6.90 Å (FSC=0.143), an asymmetric core model of 8.73 Å (FSC=0.143) and a complex model of 22.06 Å (FSC=0.5).

For Dataset 3, 3,098 images were collected and 41,843 particles were picked from the micrographs. Particles were 2D classified in 256 classes using RELION 3.0. 2D

classes with prominent structural features and the corresponding number of particles ending up in each class are shown in **Figure S5** A (**Appendix**).

2.2.10 Sequence alignments, homology modelling, interface energy statistics and model refinement.

The Needleman-Wunsch algorithm was used for identification of sequence identity and similarity [164]. Whole sequences, including potential signal peptides, were used for *H. sapiens* and *C. thermophilum* deposited in UniProt [176]. For the ctBCKDHc E1β protein (UniProt-ID G0SA92), an annotation discrepancy was found based on an erroneous genomic mapping. This error was reported and updated in the next UniProt update. For sequence alignment of this protein, amino acids 1 to 413 were used. For the generation of atomic models for the E2 cores of PDHc, OGDHc and BCKDHc and also for the PDHc E1p and the E3 protein complexes, MODELLER [162] was used. VTFM and MD optimisations were used as described in the manual (Section 2.2.2). Template structures used for modelling were 6CT0 for PDHc E2p core [145], 6H05 for OGDHc E2o core (10.2210/pdb6H05/pdb), 2II3 for BCKDH E2b core [177], 2OZL for PDHc E1 heterotetramer [178] and 6I4R for the E3 homodimer [146]. The E2 core proteins were modelled as a homotrimer, the E1 proteins as a heterotetramer and the E3 proteins as a homodimer to ensure structural integrity and optimise interface interactions. The dodecahedral or cubic structure of the cores were generated by superimposing the generated trimeric models to the biological assembly of the used template structure.

For calculating the interface properties, HADDOCK refinement server (**Table 5**) was used [115]. For the E2 core proteins, two monomers of one trimer (m1, m2) and a third monomer of a neighbouring trimer (m3) were isolated. For the intra-trimeric interface calculation, m1 was defined as chain A whereas both others as chain B. For inter-trimeric interface calculation, m1 and m2 were defined as chain A and m3 as chain B. For dimeric interface (α/β) of the E1 heterotetramer, a single α and a β subunit were isolated. For the heteromeric interface (αβ/αβ), each one α and a β subunit were combined to a single chain. For the homodimeric E3 complex, no modification was needed. For interface comparisons to the E3BP trimer, unique interfaces were extracted from the E2p/E3BP core and subsequently refined in HADDOCK.

For the refinement of the initial model for the *ct*PDHc E2p core in the cryo-EM density, the complete core model and the map were superimposed using ChimeraX

[170] and refined in real space with PHENIX [165] using standard settings. Regions of the model that were not covered by densities were removed, *i.e.,* the *N-ter* extension after Pro252. Before deposition, the core structure was condensed to a single chain using the PHENIX tool "*Find NCS operators*".

2.2.11 Localisation, fitting and refinements

For the identification of the folded region of E3BP, the protein sequence was analysed using InterPro [179]. The overall domain architecture of E3BP is highly similar to E2p: A single *N-ter* LD, a peripheral subunit binding domain and an ordered *C-ter* region, each connected by a disordered region. The ordered *C-ter* region (residues 270–442) was considered the core region and subsequently analysed by HHpred [180]. One suitable template structure (PDB-ID 2II3) [177] was identified that covers parts of the *C-ter* and *N-ter* core region and the *C. thermophilum* structure was produced, using MODELLER by default [162]. One out of the four trimeric densities observed inside the PDHc core was isolated from the C1 and the I reconstruction to generate the trimeric structure. The C1 density was then rotated by 120° to capture the underlying C3 symmetry of the inside trimeric protein complex and aligned to the symmetrised icosahedral density, followed by an unsupervised fitting of the initial model. All these geometric operations were performed using UCSF Chimera [169]. Due to clashes, the derived C3 symmetrised trimeric structure was truncated at the *C-ter* by 7 residues. Then, the structure was real-space refined in the C1 trimeric density with PHENIX (version 1.17) with additional morphing and simulated annealing by default.

2.2.12 Densitometric protein quantification

The recombinant protein was quantified by Bradford. For protein quantification, the same volume of sample and defined amounts of recombinant protein were analysed by Western-Blotting. Specifically, purified *C. thermophilum* E2p-His-Tag (aa 29-459) and E1α-His-Tag (aa 37-411) were used as the positive control (PC). Signal intensity was quantified using Image Lab 6.1 (**Table 2**, Bio-Rad Laboratories, Inc.). For direct comparison, derived concentration ($ng \cdot \mu L^{-1}$) was normalised by molecular weight. The marker used in immunodetection assays was Bio-Rad Precision Plus Protein™ All Blue Prestained Protein Standards.

2.2.13 Fit of proteins in C1 reconstructions and measured active site distances

To unambiguously localise the E1p, E2p and E3 proteins within the PDHc, an ensemble of 10,000 low-pass filtered maps of each at 20 Å was generated and placed inside the core of the PDHc reconstruction. Then, the Chimera Fitmap function [169] and the PDHc asymmetric map at 0.02 sigma with a gaussian filter of 2.5 σ was used. This resulted in fits in various parts of the PDHc asymmetric map, including inside the core, on the core and outside the core, in the peripheral densities. Using the Chimera metric "correlation" as readout, distribution of the hits classified as being part of the core, localised inside the core, or at external densities was performed. Among the top hits with high volume coverage were C2-symmetrically oriented positions of the E1p and the E3 at the large density cluster. All other hits were manually examined and the best competing hits had considerably lower scores, caused obvious clashes and were non-plausible, *e.g.*, mapped on or entirely inside the E2p core, or were fitted in regions of external density without sufficient density coverage.

The active sites were defined by residues His432 of each E2p, Glu81 of each E1β and Cys78 of each E3 monomer to calculate active site distances. The distances between these active site residues are the Euclidian distance of their Cα atoms.

2.2.14 Quantification and statistical analysis

For the kinetic experiments of PDHc and OGDHc, values for the enzyme kinetics measurements were measured in technical duplicates of biological triplicates. For counting the localised densities in the negative stain class averages of PDHc and hybrid OGDHc/BCKDHc, class averages were provided to five lab members. These class averages (55 for PDHc and 36 for the OGDHc/BCKDHc hybrid) were generated by Scipion 2.0 using the default threshold and shown in **Table S6 Appendix**. Each lab member was independently asked to count the external densities with a conservative and a non-conservative threshold. Results were statistically treated to derive distribution properties of the blind measurements (**Table S4 Appendix**). For validating the cryo-EM reconstructions of the core and the complex of PDHc, the structure calculations were performed in data coming from two independent biological replicates and then merged to report the final maps. In order to calculate the distances of active sites in the final E1p-E2p-E3BP-E3 model, all distances among all subunit active sites were measured and their average and standard deviation were considered.

3. Results

3.1 Enrichment and molecular characterisation of OADH complexes in native cell extracts

3.1.1 Single-step enrichment fractionation recovers active OADH complexes

Optimal growth conditions were set up at 45 °C and *C. thermophilum* mycelium was propagated as described in Chapter 2 (Materials and Methods) and adapted from *Kellner et al.* [97]. Applying a size exclusion chromatography step (**Figure 23** A), cell lysate was fractionated and complexes included in the high-molecular-weight fractions were isolated and vitrified. The fractions studied here and the complexes within them have been previously reproducibly identified by mass spectrometry (MS) [57]. In the cryo-EM micrographs, a variety of structural signatures of known and unknown molecules and complexes of intriguing architectures were observed, including the fatty acid synthase (FAS) metabolon [57], double and single membrane structures, liposomes with encapsulated biomolecules, filaments and other higher-order assemblies (**Figure 23** B **and Figure S5** A**, Appendix**).

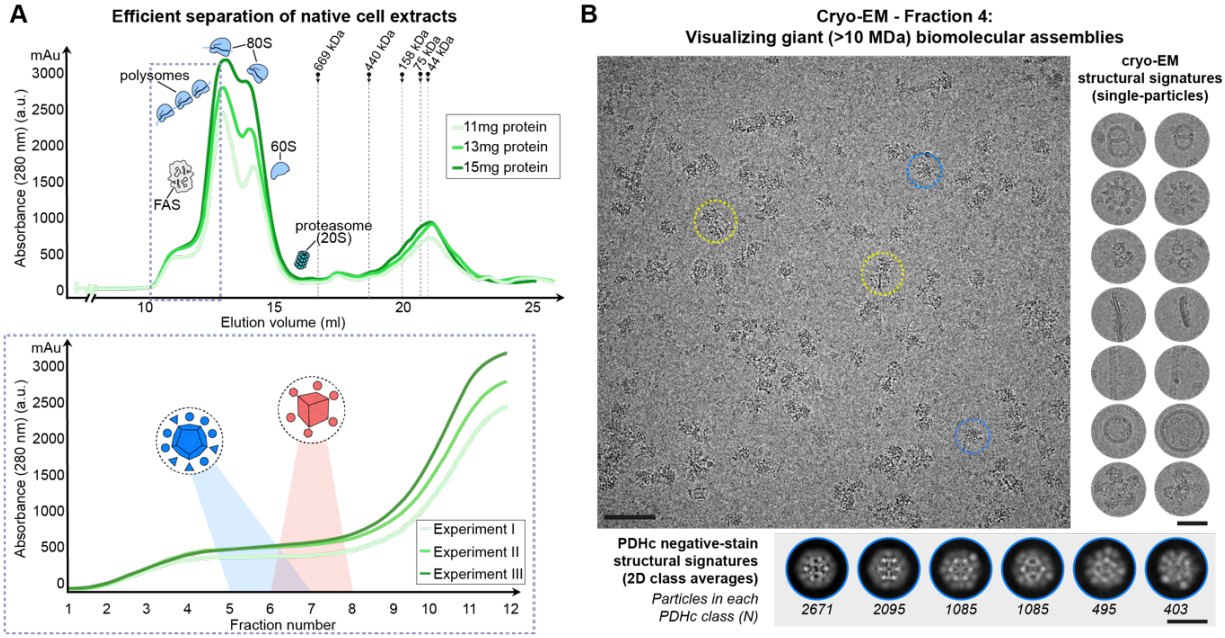


Figure 23. Biochemical characterisation of native cell extracts aiming to capture α-keto acid dehydrogenase complexes.
A. SEC profile of *C. thermophilum* native extract; Insert shows recovered protein material at high molecular weight fractions; **B.** Cryo-EM of fraction 4 (>10 MDa complexes); Inserts show: The metabolon of FAS with the bound carboxylase [57] and PDHc particles in yellow and blue

circles, respectively; On the right, single particles described in Materials and Methods; On the bottom, class averages of PDHc derived from NS. Scale bars: insert: 100 nm; extracted single-particles and PDHc class averages: 60 nm. (Source: Figure reproduced from [140])

By performing negative stain (NS) of these high-molecular-weight fractions, particles were processed and subsequent image processing to retrieve 2D class averages showed particles with diameter and subunit architecture similar to PDHc (**Figure 23** B). To further identify the studied complexes, custom antibodies (Abs) against the *C. thermophilum* PDHc E1α (UniProt-ID G0SHF3) and E2p PDHc E2p (UniProt-ID G0S4X6) protein constructs were produced and a co-migration profile for both was shown (**Figure 24**).

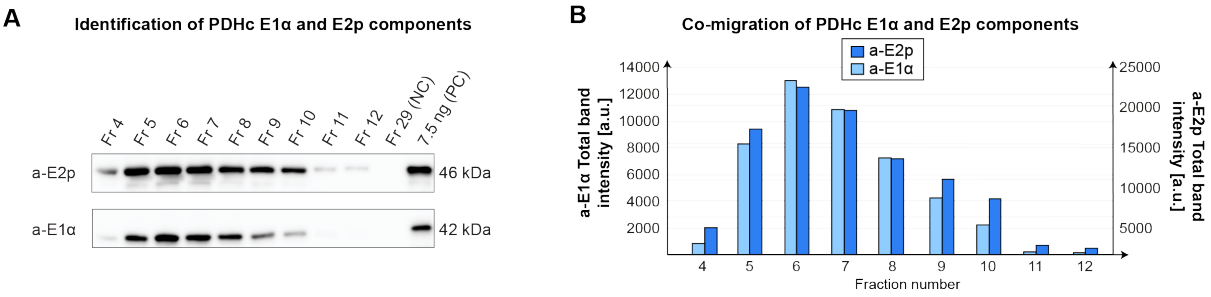


Figure 24. Immunodetection of PDHc E1α and E2p components.
A. Immunodetection of E2p and E1α within cell extracts; NC and PC stand for negative and positive controls respectively(see Materials and Methods Chapter 2). **B.** Calculation of total band intensity exhibiting co-migration of identified components across studied fractions (Source: Figure reproduced from [140])

The latest was also in agreement with the quantitative protein identification from MS of previous results [57]. Specifically, abundance from MS data was inferred using the iBAQ combined score for peptide quantification [181] reported previously [57]. As shown in **Figure 25**, a significant amount of co-eluting OADH complex components was identified and sequence and complex specific abundance analyses were performed (**Figure S5** B-J and **Table S3 Appendix**). These data point to a high abundance of PDHc components with a peak at fraction 6, whereas, for OGDHc and BCKDHc components, abundance showed a peak at fractions 7 and 8, respectively (**Figure 25** and **Figure S5 Appendix)**.

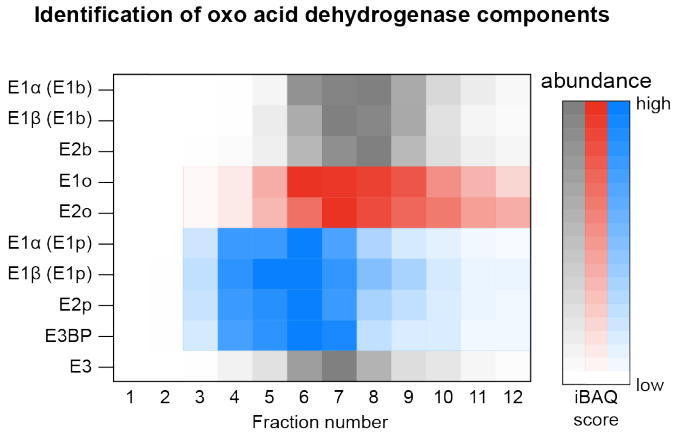


Figure 25. MS data showing the abundance of α-keto acid dehydrogenase subunits across fractions.
Data shown correspond to the different subunits and colour code shows the abundance of each subunit according to the calculated iBAQ score (Source: Figure reproduced from [140])

As far as the amino acid sequences and domain architecture of *C. thermophilum* OADH complexes compared to their human counterparts are concerned, a similarity ranging from 38 to 69 % was revealed (**Figure S5** B-D**, Appendix**). Of importance to note here, a significant difference between the two homologues is that the E2p of *C. thermophilum,* similar to yeast, has one LD domain, compared to the human counterpart, which has two (**Figure S5** B**, Appendix**).

Equally important was to investigate these complexes' activity within the identified fractions. This is because studying a cell extract architecture without knowing if protein complexes and their communities retain their function is not, of course, optimal. Thus, a coupled continuous spectrophotometric assay that targets the E1 activity of PDHc and OGDHc throughout the steps of the experiment and within the fractions was performed as described in Chapter 2 of the thesis.

From cell lysis, followed by ultracentrifugation, concentration (**Figure 11, Materials and Methods, Chapter 2.2.2** and **Figure S2 Appendix**) and to the collected fractions after SEC (**Figure 14, Materials and Methods, Chapter 2.2.4.1**), both PDHc and OGDHc metabolons retained activities. They showed a peak in activity at the corresponding fractions that were abundant (**Figure 26 and Figure 14, Materials and Methods, Chapter 2.2.4.1**) as determined by MS (**Figure 25, Figure S5** G-J **Appendix**). As shown in **Figure 26** highest activity of PDHc was detected in fraction 6. It is also known from the literature that an active E1 tetramer corresponds to a calculated molecular weight of only ~160 kDa [182]. Taking this into account and together with the observations that E1 was: (a) identified in the high MDa fractions; (b) co-eluted

with all other proteins of PDHc and OGDHc and (c) is measured to be catalytically active, it is bound in the respective higher-order complexes. This result clearly demonstrates the strength of combined biochemical and biophysical characterisation of endogenous cellular material optimally retrieved in native cell extracts.

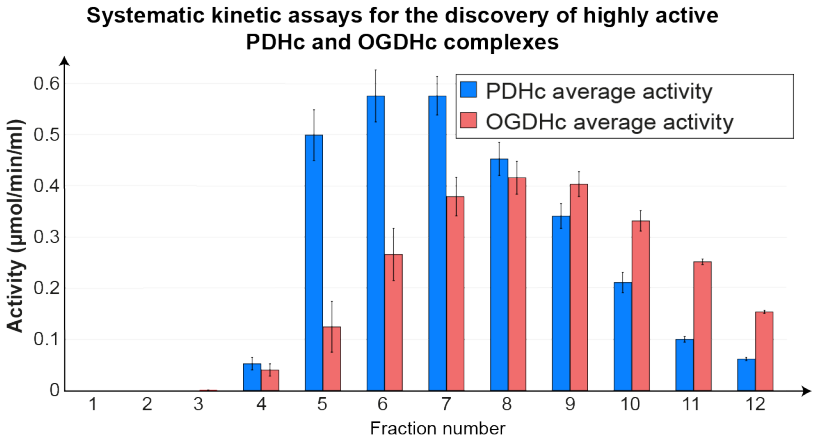


Figure 26. Activity assays of PDHc (blue) and OGDHc (red) E1.
Standard deviation is retrieved after averaging three biological replicates reported in **Figure 14, Materials and Methods, Chapter 2.3**. (Source: Figure reproduced from [140])

More specifically, for fraction 6, where PDHc is of the highest abundance and activity and to unambiguously correlate activity to the presence of all PDHc subunits, further immunoblotting experiments were performed, using this time antibodies specifically designed against each respective recombinant protein assembling PDHc (**Table 3, Materials and Methods Chapter** & **Figure 27**) [183]. These data provided an extra orthogonal validation and co-identification of co-eluting PDHc subunits that were identified by label-free MS in the same fractions [140]. In detail and as shown in **Figure 27**, all proteins of the PDHc metabolon are present in the high molecular-weight fraction 6, but not in the later fraction 22. As positive control recombinant proteins, used for antibody production were loaded on the same gel.

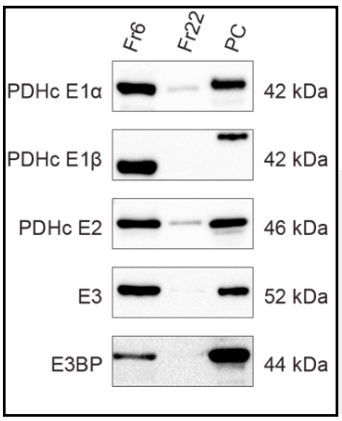


Figure 27. Detection of all components of the PDHc metabolon (n= 1).
Fraction 6 (Fr6), fraction 22 (Fr22; negative control) and recombinant protein, used for antibody production (PC positive control) were analysed. The displayed molecular weight corresponds to the recombinant protein. (Source: Figure reproduced from [183])

The apparent shift in molecular weight for PDHc E1β might be explained by either higher charge to mass ratio in the native protein by phosphorylation, or by a significant difference of the native protein in comparison to the recombinant protein due to signal-peptide cleavage on the native protein and the affinity tag on the recombinant protein.

Concerning the results reported in **Figure 27**, the positive control was overexpressed in *E. coli* and thereby lacks any post-translational modifications. On the other hand, the endogenous protein might be modified, *e.g.,* by phosphorylation which increases the negative charges during gel separation, resulting in a lower apparent molecular weight. We can also not rule out that due to signal-peptide cleavage on the endogenous protein, we have a significant shorter polypeptide than the one using for overexpression, which also carries an additional His6-tag. Overall, reported data in this Chapter unambiguously show that all protein components of the PDHc metabolon are identified in the same fraction, providing validation to the MS results previously shown (**Figure 25**) and pointing to the fact that the complete PDHc metabolon is fully recovered in native cell extracts.

3.1.2 Large-scale EM imaging and further image processing identifies multiple structural signatures in native cell extracts

Fractions were extensively screened with EM while identifying structural features of fully assembled OADH complexes, which of course, was the primary goal. A total number of 5,666 images of negatively stained samples for fractions 5-11 were collected and processed for subsequent image analysis **(Figures 28-30, Table S2 Appendix)**. The criteria for choosing these fractions for visual analysis were the following: (a) high molecular weight and thus, OADH complexes must be enriched in those (**Figure 12, Materials and Methods Chapter)**; (b) protein concentrations higher than in the previous fractions (**Figure S1** C**, Appendix)**, thus more particles expected per micrograph and (c) minor populations of amorphous aggregates present in the earlier fractions were avoided. Except monitoring the presence of aggregates by EM, observed absorbances among replicates at 320 nm were low, indicating fewer unwanted artifacts that might be introduced (**Figures 23** B**, 28** and **Figure S1** D).

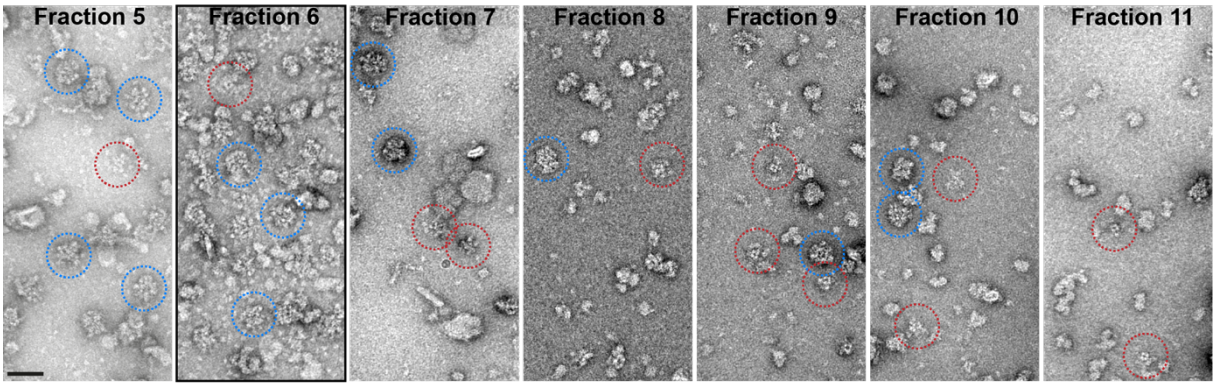


Figure 28. Large-scale negative stain imaging of consecutive fractions.
Representative micrographs displaying single particles present in fractions 5-11. Blue and red circles highlight prominent signatures of particles with dodecahedral and octahedral cores, respectively. Scale bar: 60 nm. (Source: Figure reproduced from [140])

The low OADH complexes observed in the acquired micrographs correlate to the MS data, where the highest protein abundance of summed OADH complexes does not overcome the 40 % level of overall fraction abundance (**Figure S5** G-J**, Appendix**). In total, two million (*N*=2,043,294) particles were classified in 2D and class averages of the two prominent particle cores emerged (**Figure 29**).

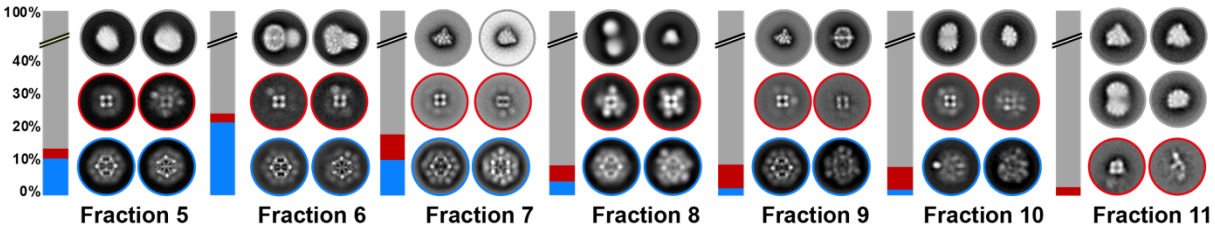


Figure 29. Recovery of reference-free 2D-class averages from consecutive cell extract fractions.
Quantification of particles included in the class averages showing characteristics of dodecahedral (blue), octahedral (red) and other types of structures (grey). Insert class averages show a representative class average from each category. All class averages of fraction 6 are shown in **Figure S6** A **Appendix**. (Source: Figure reproduced from [140])

Both distinct cores identified after the single-particle classification were grouped and 3D reconstructed with no symmetry at low resolution (**Table S2**). Still, cores were retrieved with prominent structural features which reflect their symmetry (correct amount of faces and vertices). *C. thermophilum* models were built based on their human counterparts for each E2p, E2o and E2b core and then superpositioned to the reconstructed densities. The overall density-model cross-correlation values calculated for every superposition showed high cross-correlation values (CC) as well as high overlap of *C. thermophilum* models to their respective volumes. Utilising the derived atomic model after homology modelling and reconstructing the overall core

architectures, each of the distinct cores derived from NS EM can be efficiently explained (**Figure S6** B-F **Appendix**). Still, a fraction of the cores' densities did not belong to core structures (**Figures 28-30, Figure S6** B-F **Appendix**). These surrounding densities localised asymmetrically, not in a diffuse pattern and for both structural species were reproducible in independent 3D reconstructions retrieved from the several fractions studied (**Figure 30**).

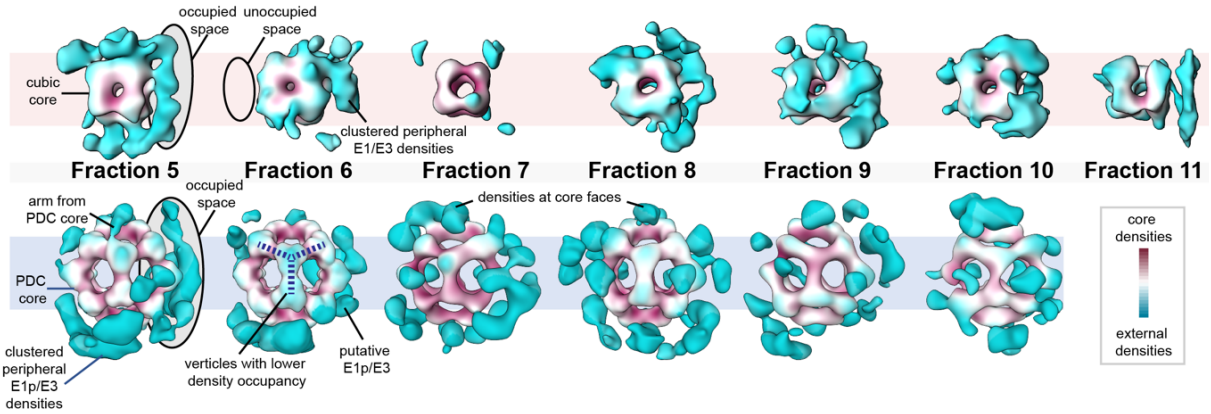


Figure 30. Systematic 3D reconstructions of PDHc and hybrid OGDHc/BCKDHc from consecutive cell extract fractions.
Cubic (top) and dodecahedral cores (bottom) of complexes 3D reconstructions with C1 symmetry. For volume presentation, applied density thresholds were: for PDHc F5-F10: 0.115, 0.0815, 0.0489, 0.107, 0.0759, 0.0909; and for OGDHc, F5-F11:0.114, 0.071, 0.02, 0.0957, 0.0201, 0.0932, 0.0472. (Source: Figure reproduced from [140])

Homology models of structurally-resolved regions related to all protein components were produced to validate structural relationships within the OADH complex family and crosslinking MS data from cell extracts previously communicated [57] showed that, for those OADH protein components, 118 crosslinks are identified in total and were subsequently analysed **(Figure S4** A-C **Appendix).** Within those crosslinks, 77 % (N=91) were mapped in flexible regions of OADH complexes agreeing with the underlying plasticity of those assemblies (**Figure S4** A **Appendix**), whereas the rest (N=27) map to ordered domains and recapitulate the log-normal distribution of crosslink distances (16.1±6.5 Å), with only one crosslink showing minor violation of the 30.0 Å cut-off distance (E1o, Lys542-Lys171, 33.2 Å) (**Figure S4** B-C **Appendix**).

Due to their similar architecture, the reconstructed particles with cubic cores were hard to assign to either OGDHc or BCKDHc (**Figure S6** B-E **Appendix**), always considering that they elute with similar abundance (**Figure S5** H-J **Appendix**). On the other hand, the dodecahedrally reconstructed particles had higher chances of being PDHc in combination with the immunodetection (**Figure 24** & **27**), quantitative MS

(Figure S5 G-J **Appendix**) and activity data shown before (**Figure 26**). Furthermore, 3D reconstructions from the same fraction strongly resemble the known eukaryotic E2p PDHc core (**Figure 30**) and, there is no other known complex with a comparable abundance in cells sharing similar architecture in contrast to the other OADH complexes. This was further validated by applying known symmetry (symmetry type: I) to resolve PDHc. After symmetry enforcement to reconstruct the EM density, the size and architecture of the respective E2p core (**Figure S4** D-I **Appendix**) were further corroborated. However, interestingly, comparisons with mesophilic counterparts and energy calculations did not reveal any significant contribution of calculated non-covalent forces to the known stabilisation forces governing thermophilic interfaces (**Table S5 Appendix**).

Finally, it is also reported here that the retention times in the SEC (**Figure 23** A), the activity analyses, the distinct activity peaks of PDHc and OGDHc E1 enzymes (**Figure 26**), EM analysis (**Figures 28-30**) and derived core models (**Figures S4** & **S6 Appendix**) illustrate that the native size of PDHc is larger compared to the rest of the OADH complexes.

3.1.3 Localised clustering of densities surrounding the PDHc and OADH complexes reconstructions

To further address the localisation of the peripheral densities observed surrounding the PDHc and OADHc cores and reported in the previous Chapter, a significant number of single particles was collected from fraction 6 (**Figure 28** & **29**). This led to a gallery of 55 and 36 reference-free class averages out of 282,698 particles and 17,649 particles for the annotated PDHc and remaining OADH complexes respectively, containing ~54 % of the single particles (**Figure S6** A **Appendix**). All these highly contrasty local densities observed surrounding the E2 core, represented individual spatial and molecular features, as shown in **Figure 31**. A quantification of this signal recapitulated 2-6 localised densities per class average of PDHc. Extrapolating this number to the total number of averaged particles therein resulted in at least 3.10±1.27 localised densities for each PDHc particle, following stringent outer density cluster classification (**Figure 31**). On the other hand, due to possibly higher flexibility for OADH remaining complexes, statistics derived were lower with mixed-signal and/or "diffused" external densities. In numbers, this was translated to at least 1.25±0.57 of localised densities observed per single-particle (**Figure 31** and **Table S4 Appendix**).

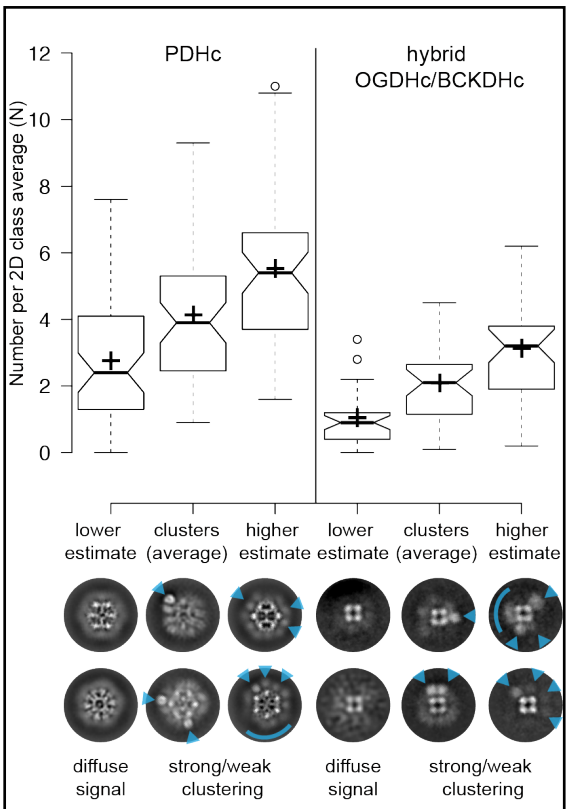


Figure 31. Quantification of localisation densities beyond the core structure.

Class averages exhibit strong external densities (clusters). Upper part boxplots of number of clusters per reference-free 2D class average and lower part class averages with diffuse, weak and strong signals (**Table S4 Appendix**). (Source: Figure reproduced from [140])

Asymmetric reconstructions containing 267,547 particles of PDHc and 9,000 OGDHc/BCKDHc, to localise in 3D those manifested densities resulted in ~26.6 Å and ~44.6 Å final resolutions, respectively (FSC=0.5, **Figure S4** I **Appendix**). Despite this limited resolution, asymmetric clustering (**Figure 30, Figure 32** A) and densities that frequently stem from the vertices of the cores were among the most remarkable observations. These densities of PDHc and the hybrid OGDHc/BCKDHc stemming from the core could correspond to the *N-ter* extensions of the E2 core proteins which tether E1 (**Figure 32** A), even though the possibility that they can correspond to E3BP flexible regions, which are known to tether the E3 cannot be disregarded due to the limited resolution. Considering that both these complexes were usually studied by applying some kind of symmetry, most of these densities are novel and have not been previously observed (shown in purple, **Figure 32)** and studied from a native source. Given these reconstructions, an overall volume of 50 % of the PDHc and 25 % of the OADH complexes correspond to these novel external densities (**Figure 32** B). Again, these results clearly show the strength of studying protein complexes and their higher-order assemblies, in this case OADH complexes to derive novel architectural insights. These additional external densities combined with the previously presented results for the identification of all protein components of the PDHc in particular bring us one step closer to the structural characterisation of the endogenous 10 MDa pyruvate oxidation metabolon.

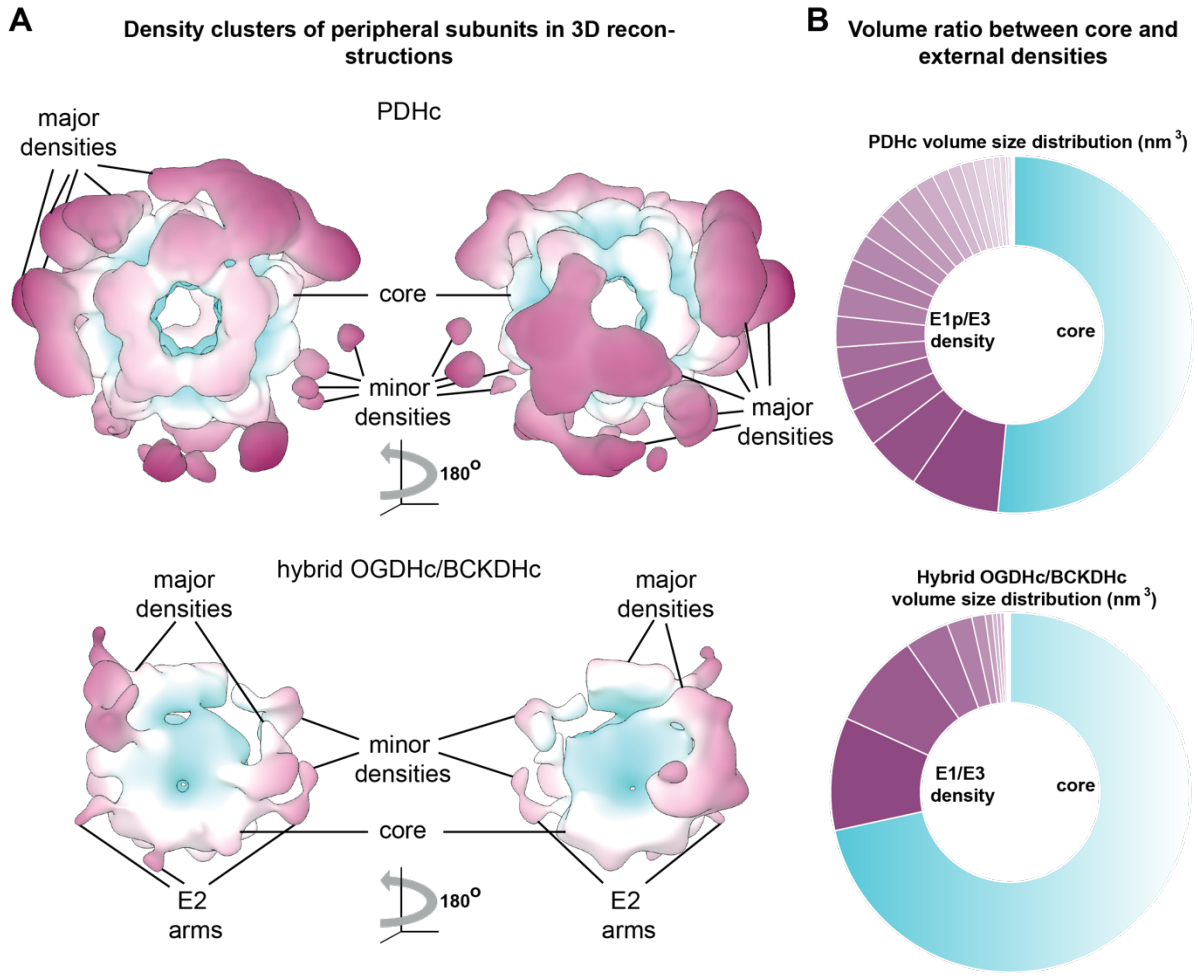


Figure 32. Density clusters of peripheral subunits in 3D reconstructions.
A. Many external densities stem from the core structure, like arms. Smaller densities are also present, constituting a minority of the overall structure. **B.** The ratio of volumetric data corresponding to the core (blue) and the outer part of the α-ketoacid dehydrogenase complexes (purple). 50 % and 25 % of the total volume is covered by external densities, where E1p and E3 dimers are supposedly localised, together with associated kinases and phosphatases. (Source: Figure reproduced from [140])

3.2 Cryo-EM of cell extracts elucidates the architecture of the endogenous PDHc core

3.2.1 Cryo-EM structure of PDHc reveals a stable interface and flexible folded domains

The OADH E2 core models did not reveal notable energetic interface contributions compared to their mesophilic counterparts after being modelled and refined in their respective NS EM 3D reconstruction (**Table S5**). The NS reconstructions showed that OGDHc results in lower resolution (32.7 Å, FSC=0.5, **Figures S4** I & **S6** C **Appendix**) compared to that of PDHc (14.2 Å, FSC=0.5**, Figure S4** I **Appendix**). Furthermore,

the *C. thermophilum* E2p core model fitted quite well in its respective density, even though some parts of the atomic models did not completely fit their expected densities in the NS 3D volumes. These namely being: (a) part of the *N-ter* region Gln227-Asn251 (Thr417 to Ile443 in human) of E2p (**Figure S4** F-G **Appendix**) forming a helix-turn-sheet-loop fold (**Figure S5** E-F **Appendix**) [145,184], covering the external top part of the trimeric asymmetric element of the core (**Figure S4** G-H **Appendix**) and (b) inter-trimeric contacts mediated by helix Gly341-Gly359 of the homodimer (Gly531-Gly551 in human) (**Figure S4** G-H **Appendix**).

To gain further insights and compare in more detail those observations and if they hold true due to decreased resolution retrieved by negative stain, fraction 6 was vitrified *via* plunge-freezing and the cryo-EM data analysis resulted in a resolution of the PDHc core at 6.9 Å (FSC=0.143, **Figures 33, 34, Table S6** & **Figure S7 Appendix**). Since this was a resolution where the α-helices can be distinguished (**Figure 33**) and the densities were uniform, the fitting of the previously resolved NS E2p model was successful (**Figure S4** F, H **Appendix**).

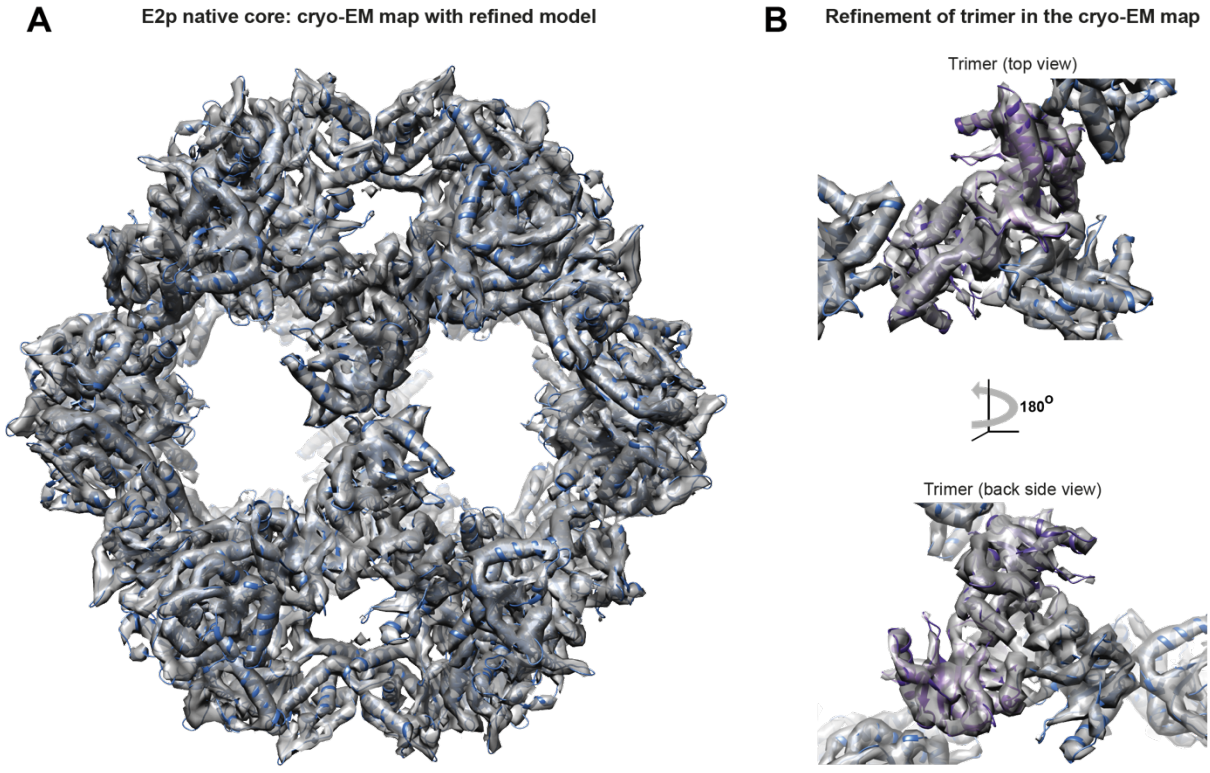


Figure 33. E2p core native cryo-EM map.
A. Architecture of the *C. thermophilum* E2p core. **B.** View of the asymmetric unit of the E2p trimer. (Source: Figure reproduced from [140]).

Amino acids Gly341-Gly359 of the E2p homodimer, present at the inter-trimeric interface, were now clearly resolved (**Figure 34, Figure S4** H **Appendix**), showing that

low-resolution or structural unfolding artifacts due to staining may have limited localisation of this structural region in the previously reported results.

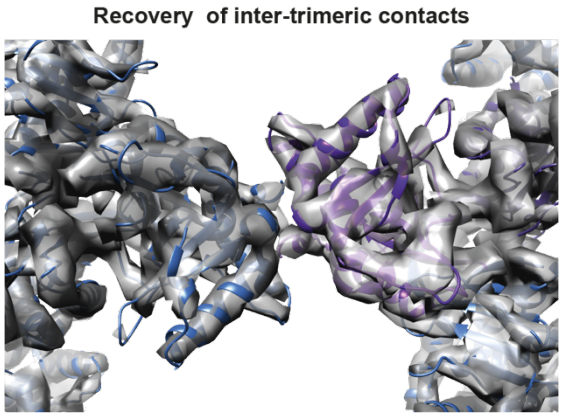


Figure 34. Recovery of inter-trimeric contacts and side helices which were not identified in the negative stain reconstruction are apparent in the cryo-EM map.
(Also **Figure S4 Appendix**) (Source: Figure reproduced from [140])

Species- and interface-specific differences were revealed by repeating energy calculations of the intra- and inter- trimeric interfaces. *C. thermophilum* E2p core has higher desolvation energies accompanied by a more extensive intra-trimeric interface compared to human, as shown in **Figure 35**. Also, the *C. thermophilum* inter-trimeric interface desolvation energy is almost two times higher than its counterpart. Especially, weaker electrostatics were calculated for the inter-trimeric interface (**Figure 35)**, possibly due to compensation effects for higher stability of the thermophilic interface. This biophysical adaptation related to the higher contribution of solvation reflects and, probably to an optimisation of the hydrophobic interactions in both E2p core interfaces and has also been reported for other thermophilic proteins [185] and complexes [186].

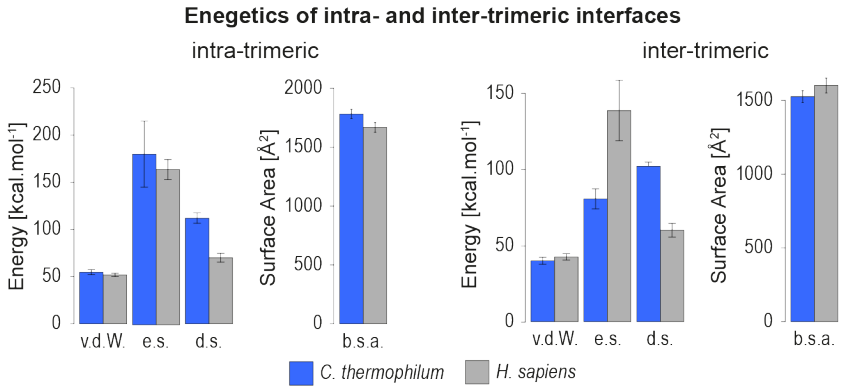


Figure 35. Energetics of intra- and inter-trimeric interfaces.
Energy calculations for thermophilic and mesophilic interfaces. Average and standard deviations show energetics for the top 4 calculated structures. v.d.w.: van der Waals; e.s.: electrostatics score; D.s.: desolvation score; b.s.a: buried surface area. (Source: Figure reproduced from [140])

There is another structural domain that is not observed in the native PDHc core, neither in the NS reconstruction nor in the reported cryo-EM map, even though it is highly conserved in *C. thermophilum* (residues Gln227-Asn251) (**Figure 36** A**, Figure S4** H **Appendix**), equivalent to the region of the human *N-ter* region (Thr417-Ile443). Even if E2p constructs used to date most of the times include that *N-ter* extension, it has been shown as not actively participating in the catalytic domain [184]. In contrast, the active site, which is also highly conserved between the two homologues, seems to interact with this region and becomes exposed if this *N-ter* extension is absent (**Figure 36** B). Using the crystallographically resolved molecules of oxidised CoA and the lipoyl derivative (6,8-dimercapto-octanoic acid amide) from the *Azotobacter vinelandii* PDHc (**Table S8**) [187] it was observed that both could be accommodated in the active site without violation, although the side-chain resolution and cofactor Coulomb potential maps were not resolvable (**Figure 36** B).

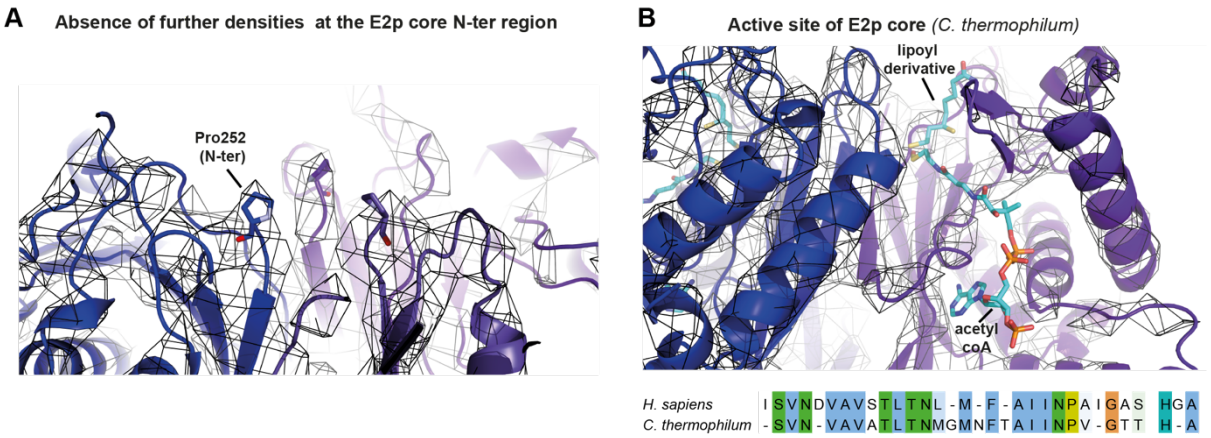


Figure 36. Densities at the E2p core *N-ter* region and active sites of E2p core.
A. The core *N-ter* density does not continue beyond Pro252 (also **Figure S4 Appendix**) **B.** The *C. thermophilum* E2p active site. Fitting a lipoyl derivative (lpm from: PDB-ID 1EAE) and CoA (CoA from: PDB-ID 1EAD) is possible, without clashes. Below the PyMOL image, conserved active site residues are shown. (Source: Figure reproduced from [140])

In this trans-acetylation reaction, substrates enter the active site from opposite directions. The acetylated lipoamide moiety from the outside, whereas CoA enters from the interior of the core. The active site region has a considerably larger surface than the human counterpart (**Figure 37**) because of the structural absence of the conserved additional back-folded structural element of the *N-ter* region in *C. thermophilum,* which makes it even more accessible. Finally, as shown by the *N-ter* E2p region flexibility, possible artificial states of back-folded structural elements covering parts of the E2p active sites are circumvented by capturing the complex in a native cell extract and as close to native as possible. It is highly intriguing that this structural element is not

resolved; various reasons could explain such a phenomenon, being either sub-stoichiometric, or transient, or, as mentioned above, possibly artificial due to the utilisation of a construct and not of the fully assembled, endogenous PDHc metabolon studied in this thesis.

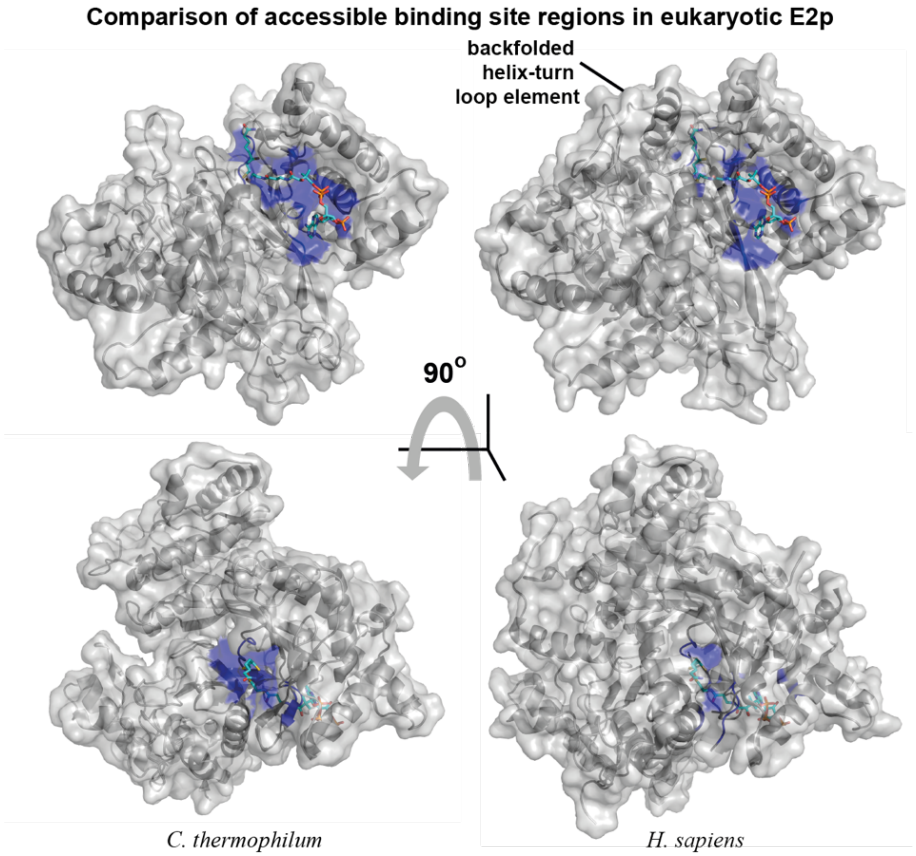


Figure 37. Comparison of the active site accessible surface areas of human and *C. thermophilum* E2p cores.
(Source: Figure reproduced from [140])

3.2.2 Resolving the native core at 3.85 Å from a eukaryotic PDHc metabolon

The fraction of interest was vitrified to further structurally characterise native PDHc and another dataset of 2,808 micrographs was acquired. Knowing that higher-order assemblies of considerable size exist in this fraction (**Figure 38** A), pixel size was updated to 1.57 Å rather than 3.17 Å which was used to derive the core reported in **Result Chapter 3.2.1**. Decrease of pixel size would lead to a possible increase in resolution of observed MDa complexes after image processing. Such an imaging protocol allowed to map a broad cell extract area of 1158.80 μm^2. As expected, and previously investigated, data showed that PDHc was already prominently present within those micrographs (**Figure 38** A), even in a native fraction where other

biomolecules of notable size co-elute. PDHc was also often participating in even larger protein communities [140]. Major 2D averages were retrieved for PDHc after localising the metabolon and performing image processing (**Figure 38** B) and ultimately included 10,249 single particles for further analysis. Densities around the cores were averaged to resolve the core scaffold (**Figure 38** B).

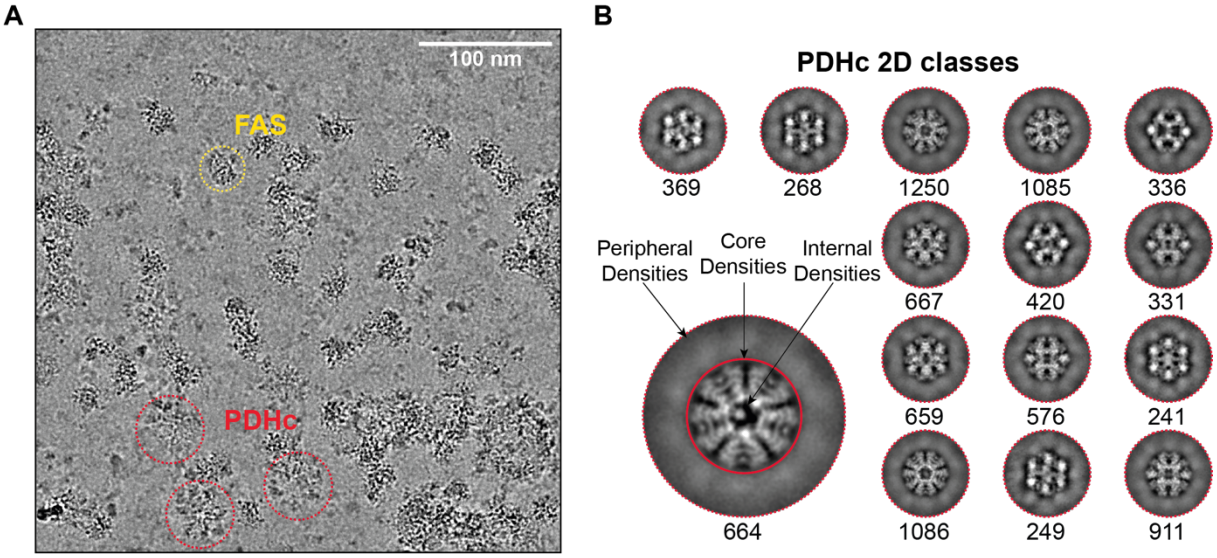


Figure 38. Representative micrograph of denoised micrograph and PDHc 2D classes. **A.** Representative micrograph of fraction 6 after denoising. PDHc in higher-order assemblies is visible in a heterogeneous environment, *e.g.*, including fatty acid synthase molecules (FAS). **B.** 2D classes of PDHc retrieved from cryo-EM single-particles. External, core and internal densities are clearly distinguishable. Numbers below representative class averages correspond to number of particles included in each class. (Source: Figure reproduced from [183])

The reconstruction of the core with an application of icosahedral symmetry was resolved at 3.85 Å resolution (FSC=0.143) (**Figure 39** A & **Figure S9**), where the backbone was distinguishable (*e.g.*, helical curvature) and partial side-chain resolution was visible (**Figure 39** B). The map's resolution allowed to build a refined model of the *C. thermophilum* PDHc core (**Table S7**). Although the model compares well to the previously lower-resolution core model, with an overall root-mean-square-deviation (RMSD overall) of 1.95 Å, atom placement was optimised, visible in map correlation coefficients of 0.9 ± 0.1 of the new model and 0.4± 0.2 of the previous model (**Figure 39** B).

Two interfaces stabilise the overall icosahedral core with distinct energetic signatures, calculated with HADDOCK (**Figure 39** C). Especially, the intra-trimeric interface buries a large surface area of ~5000 Å^2 whereas the inter-trimeric interface buries only ~1500 Å^2. Such differences point to the former sharing similarities to permanent interfaces [188], whereas the latter to interfaces found in transient protein-

protein interactions [189]. Sizes of the interfaces may support the conservation of an assembly mechanism for the core with the E2 trimer being initially formed and subsequently, trimers associating to form the assembly [190]. In addition, van der Waals (vdW) interactions are known to correlate qualitatively with the experimental affinities of known rigid protein-protein interactions [191]. Calculated values for both interfaces corroborate the more permanent nature of the intra-trimeric interface that has three times lower vdW energy than its inter-trimeric counterpart (**Figure 39** C).

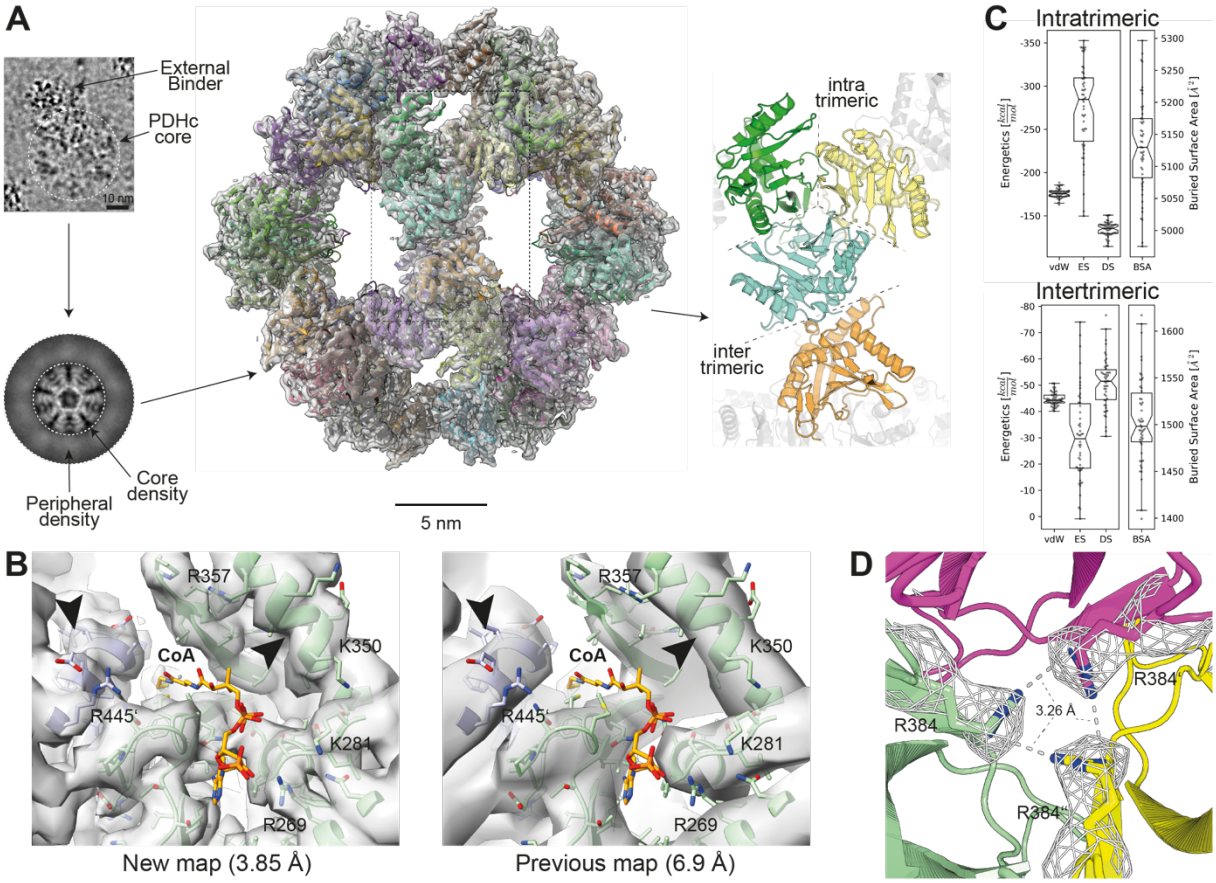


Figure 39. Near-atomic resolution reconstruction and characterisation of the PDHc core. A. Icosahedrally-averaged cryo-EM reconstruction of PDHc core. Densities corresponding to PDHc higher-order binders, visible in raw micrographs (scale bar represents 10 nm), as well as external PDHc densities of the metabolon, visible in 2D classes, are averaged to retrieve the high-resolution core structure. Each E2 monomer is coloured uniquely. **B.** Comparison of improvement in resolution, compared to reported results in **Chapter 3.2.1**. CoA molecule is not resolved but is here computationally placed and refined within the binding site as described in the **Methods Chapter**. Side chains in the CoA binding pocket of the active site are distinguishable. Helical pitch (arrowhead) allows now the unambiguous placement of the backbone. **C.** Energetic calculations within the E2 core structure from n = 50 HADDOCK models. The intra-trimeric building block with an extensive buried surface area (b.s.a) is stabilised by electrostatics (e.s.) and van der Waals (v.d.w.) energies, whereas desolvation (D.s.) energies have a minor contribution. In the dimeric inter-trimeric interaction, D.s. plays a major role, whereas v.d.w. and e.s. energies have decreasing contributions. The box minima represent the 25th percentile, the box maxima the 75th percentile, the Notch indicated the data's median, whiskers extend to the minimum and maximum value inside of a 1.5 interquartile range.; All data points are overlayed as a beeswarm plot. **D.** Potential Arg cluster

in the intra-trimeric interface. Arg384 of each subunit are in close contact, thereby contributing to the electrostatic binding energy. Contour levels applied for the cryo-EM maps are as follows: (A) 0.05, (B) 0.03 (new map) and 0.09 (previous map, (D) 0.041). (Source: Figure reproduced from [183])

The final 60-meric model of the core also shows excellent agreement with the cryo-EM map with a density-map cross-correlation value of 0.84, compared to the previous model with only a cross-correlation value of 0.53. In the intra-trimeric interface, a large contributor to the buried surface area is a well-resolved cavity. Inside this cavity, three Arg residues with a side chain-side chain distance of 3.26 Å are visible (**Figure 39** D). It cannot be ruled out that this proximity results from symmetrical averaging, but an Arg cluster is suspected. These are rare but may form a delocalised π-electron network, significantly improving protein stability [192]. This result is consistent with the previously performed energy calculations, deriving the favourable electrostatic energy contributions (**Figure 39** C). This Arg is unique for Fungi, as shown by a large-scale comparative sequence analysis that was performed [183]. In Metazoans, an Asn or His was found present at this position, while in Plants, a Gln is frequently found. As these residues can also form delocalised π-electron networks [192,193], this might be a conserved feature of the PDHc core exceptional stability.

3.2.3 Novel oligomerisation interfaces of four E3BP trimers embedded in the PDHc

For the E3BP protein, a minimal fold was constructed (**Figure S7** A **Appendix**) based on structural similarity with *Bos taurus C-ter* domain of E2b (**Table S8**) [177] and identification of structural homologs for ctE3BP *C-ter* regions (residues 270-442, **Figure 40**). Overall, the sequence organisation of E3BP is similar to E2, with an *N-ter* LD, a central peripheral subunit binding domain (PSBD) where E3 might bind, and the *C-ter* core region connected by disordered linkers.

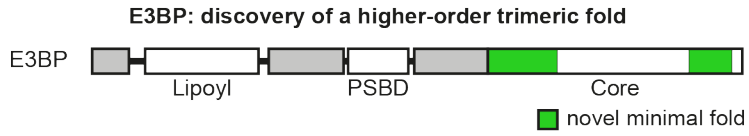


Figure 40. Structural homology of *C. thermophilum* **E3BP at the C-ter.**
(Source: Figure reproduced from [140])

Strong averaged densities are observed inside both reported icosahedrally-averaged PDHc cores previously assumed to belong to E3BP in yeast PDHc [194]. To further investigate this observation, an asymmetric reconstruction at 8.7 Å

(FSC=0.143, **Figure S7** & **Table S6 Appendix**) was performed that retrieved four robust densities within the core, each resembling trimeric complexes of C3 symmetry (**Figure 41** A-B and **Figure S7** C **Appendix**). Similarly, in *Neurospora crassa,* these densities were observed inside the E2p/E3BP core and shown to belong to E3BP [150]. This observation is further investigated and confirmed here in the case of a eukaryotic thermophile native and endogenous PDHc metabolon.

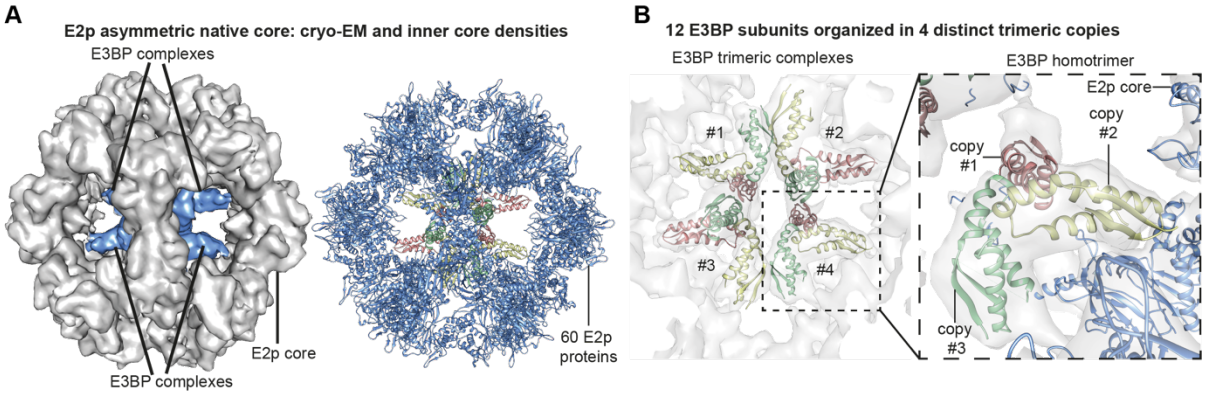


Figure 41. E2p asymmetric native core: cryo-EM and inner core densities.
A. A C1 PDHc core reconstruction unveils densities inside the E2p (blue density and red/yellow/green molecules). **B.** E3BP trimeric models reveal significant curvature, recapitulated by the density. (Source: Figure reproduced from [140])

One of the three arms of the inside trimeric density (**Figure S7** D **Appendix**) was recapitulated in the extended E3BP core model. By applying C3 symmetry, an E3BP trimer (**Figure 41** B) with a clear, clash-free interface after fitting (**Figure S7** D-E **Appendix**) and refinement (**Figure S7** F **Appendix**) was resolved. Except that the *ct*E3BP model validates the few previously modelled helical regions of its homolog in *N. crassa*, it further reveals the interaction interface and its energetics, compared to the E2p trimer and E2p dimer, respectively (**Figure 42**). In total, energetics from 144 commonly found protein-protein interfaces were considered [191] are shown as the background grey colour distributions (**Figure 42**) and are compared to the E2p/E3BP core interface energetics. It is observed that (a) the E2p trimer has a large interface with high similarities to permanent interfaces (extensive surface area and lower electrostatics [195]); and (b) the E3BP trimer shows favourable desolvation, pointing to the involvement of additional protein-protein interactions as previously shown for other complexes with favourable desolvation [196]. However, such differences were somehow expected since only the core part of E3BP was resolved and not the full E3BP, including the flexible regions.

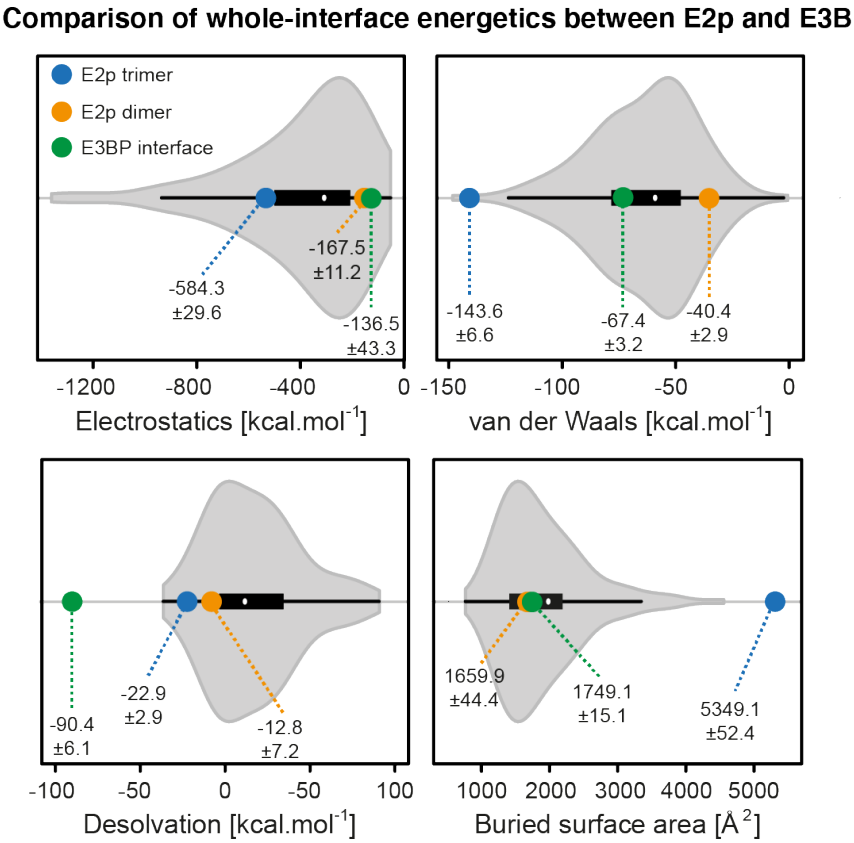


Figure 42. Energy calculations using HADDOCK and comparison of shown interfaces for the proteins of *C. thermophilum* E2p (in its core state) and E3Bp (in its trimer state). (Source: Figure reproduced from [140])

3.2.4 Stoichiometric investigation of PDHc

Two main stoichiometry models have been reported for human [149] and *N. crassa* [150] PDHc, observing the substitution [151] or the addition model [197], respectively and both exhibit variations of the E2p and E3BP core components. Diverse stoichiometries are proposed for the proteins decorating the core, namely E1α, E1β and E3. The MS data analysis performed, uncovered abundance ratios for PDHc components, capturing a distribution with a median of a correct 1:1 stoichiometry for E1p and E1b across fractions (**Figure 43** A). This analysis was performed by comparing the reported iBAQ abundances of each protein entry across the fractions (**Figure 43** A). PDHc component ratios with E1p activity in fractions were found being consistent and therefore, activity is not directly related to extensive absence/presence of PDHc components but primarily to PDHc concentration (**Figure 43** B). Therefore, combining ratios from MS data with cryo-EM models of *C. thermophilum* E2p/E3BP core a stoichiometry of a 60-meric E2p core and 4 E3BP trimers ("addition model"), in agreement with other reports in fungal PDHc [150,194], can be proposed. This stoichiometry is also consistent with the above-mentioned cryo-EM data localising the E3BP trimer (**Chapter 3.2.3**).

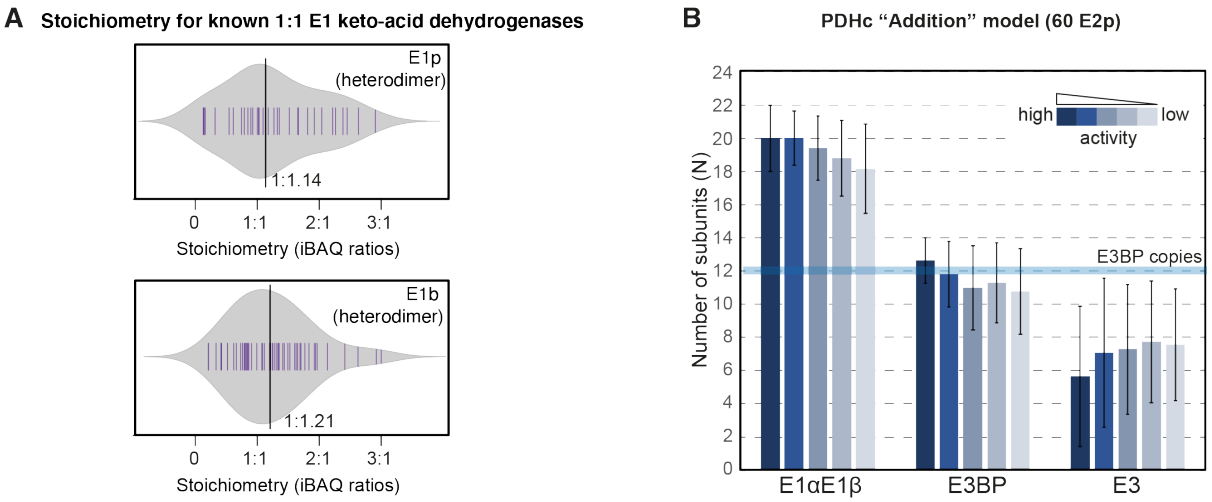


Figure 43. Stoichiometry of known E1 keto-acid dehydrogenases and PDHc "Addition" model.
A. iBAQ ratios of E1 complexes in several fractions across the replicates. **B.** Supposing 60 E2p, iBAQ ratios predict 12 E3BP subunits, with E1p and E3 dimer copy numbers being underestimated. Predictions are consistent across fractions with variable PDHc activity. Average and standard deviations shown correspond to stoichiometric calculations for fractions 5-11. The bar to the left includes measurements for fractions 5-7 (N=3) and to the right for fractions 5-8 (N=4), 5-9 (N=5), 5-10 (N=6) and 5-11 (N=7), respectively. (Source: Figure reproduced from [140])

Results are also derived for the stoichiometries of the E3 and E1p subunits: Even if the maximum number of bound complexes are predicted to be 12 E3 and 60 E1p, flexible arms of E3BP and E2p are not always saturated [198]. The prediction based on the MS data with ranges between 2-12 E3 monomers (6 E3s) and ~20 E1α E1β dimers (~10 E2p) might have underestimated their copy numbers because (a) E3 is present in all OADH complexes and (b) stoichiometry from iBAQ data is inherently limited [199]. By semi-quantifying E1α and E2p within fraction 6 with western blotting (**Figure 44, Figure S8** A **Appendix**) and performing densitometry analysis, E1p copy numbers were calculated. As a control for the validation of this result, E2p-His-Tag or E1α-His-Tag were immunoblotted using Abs against a-E1α and a-E2p, respectively (**Figure S8** A **Appendix**) and no band was detected, implying the specificity of the designed Ab and proving no cross-reactivity.

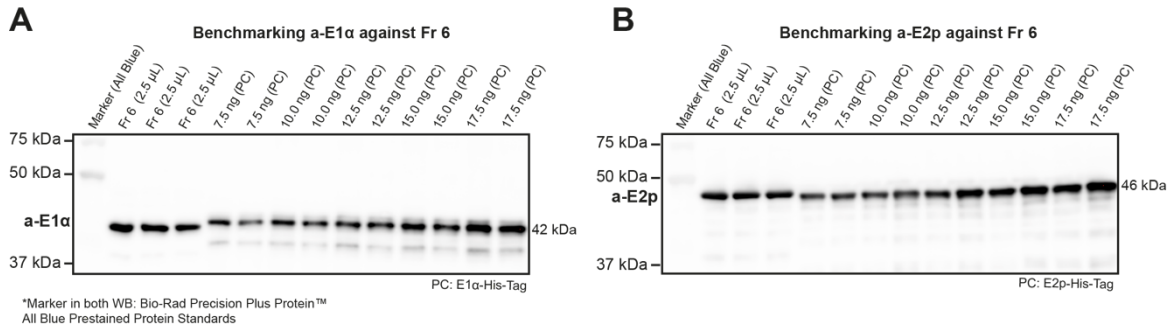


Figure 44. Benchmarking antibodies against E1α and E2p protein products.
(**A-B**) Benchmarking antibodies against E1α-His-Tag and E2p-His-Tag protein products with a technical triplicate of fraction 6 containing native *C. thermophilum* E1α and E2p permits semi-quantification by densitometry. (Source: Figure reproduced from [140])

Results showed that ~1.7 µg of E1α and ~3 µg of E2p in fraction 6, equivalent to monomeric molar concentrations of 166.2±20.6 nmol·L^{-1} E2p and 107.2±29.3 nmol·L^{-1} E1α and pointing to an average of ~19 copies of E1p per PDHc, ranging from ~16 to ~22 copies. To conclude, integration of cryo-EM, kinetic assays, MS and WB data reveal and propose a stoichiometry of the PDHc metabolon of: 1 E2p 60-mer, 4 E3BP trimers, ~20 E1p tetramers and max. 12 E3 dimers, summing up to a PDH complex composed of ~176 polypeptide chains.

3.3 Asymmetric cryo-EM reconstruction of the endogenous, 10 MDa PDHc metabolon

3.3.1 Looking at the outer densities of PDHc core (to localise E1p and E3 subunits)

An asymmetric reconstruction of the native PDHc core at a resolution of ~22 Å (FSC=0.5, **Table S6** & **Figure S7** K-L **Appendix**) was derived using the previously-mentioned particles that were derived from the combined Dataset with pixel size of 3.17 Å. Class averages and the volume are shown in **Figures 45** and **46**, respectively. These selected particles were used because their signal-to-noise ratio was higher than the ones derived from the micrographs with smaller pixel size

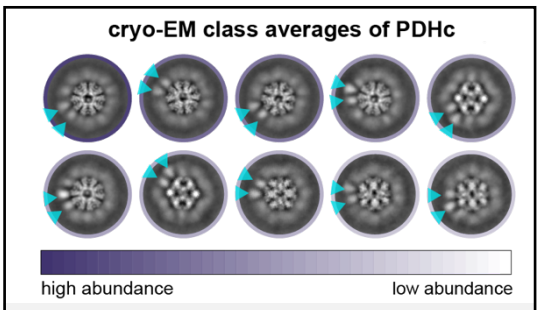


Figure 45. Asymmetric reconstruction of the full PDHc from native cell extracts.
Class averages showing localised binary densities (insert triangles); With purple to white, colour-coded class averages according to abundance. (Source: Figure reproduced from [140])

and led to structural insights into the E2p/E3BP cores (**Figures 33-37, 40-44** & **Figure S7** G-J **Appendix**). The C1 reconstruction of the complete PDHc showed a total calculated volume of 11070-13800 nm^3. By low-pass filtering the atomic models of all the PDHc components (60 E2p, E2p core, 12 E3BP, 20 E1p and 2-12 E3 dimers) at 20 Å and summing their individual volumes up, a total of 10608-13930 nm^3 is retrieved, validating that all components can be volumetrically accommodated in the PDHc 3D reconstruction of the complete metabolon. Densities above the faces represent possibly E3 due to the resolved localisation of the E3BP *C-ter* and the distance constraints imposed by the 50-residue long linker region. Furthermore, E1p components are expected to localise closer to the intersection of the dodecahedral edges, formed only by the E2p.

A systematic higher-order density outside the core, which in most of the cases occurred in pairs, was observed in the class averages (**Figure 45**). Except from the E2p/E3BP densities previously resolved and modelled (**Figures 33-37, 40-44** & **46**), the full complex asymmetric reconstruction seems to recover and recapitulate densities, previously reported at lower resolution in the negative stain C1 reconstruction (**Figures 30-32**). The validity of these densities is strengthened by the fact that they were captured in two independent biological replicates and in the negative stain reconstructions of particles from the same fraction (**Figure S8** B **Appendix**), showing a prominent conical shape surrounded by other conical, elongated and tubular densities (**Figure 46**). These MDa ultrastructure densities could include some kind of clustering of E1p/E3 subunits.

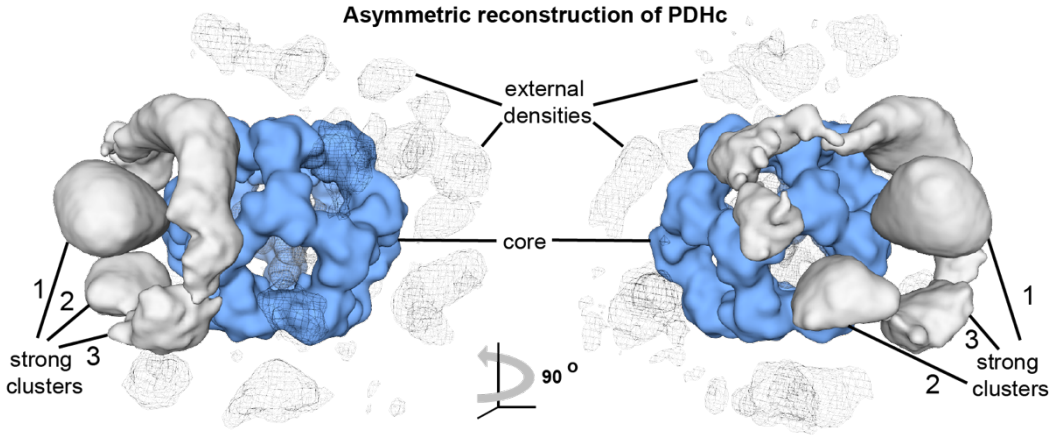


Figure 46. PDHc asymmetric reconstruction.
The PDHc map was segmented with SEGGER [167]; localised densities (wire), strong clusters of densities (silver) and E2p core (blue) are shown. Outer density clusters are denoted as 1, 2 and 3. (Source: Figure reproduced from [140])

Molecular models of the E1p and E3 subunits of *C. thermophilum*, which share sufficient sequence homology to their human homologues, were derived and together with the *C. thermophilum* E2p trimer (**Figures 33-37**), after low-pass-filtering all models of enzyme domains to 20 Å, each was fitted to the asymmetric PDHc density**,** calculating 10,000 solutions per model/enzyme domain (**Figure 47**). This means that one filtered copy per protein is fitted 10,000 times in the asymmetric PDHc density. Results for each fit per protein component were classified in three different categories named as "core" when fitting on the PDHc core, "internal" when fitting inside the core and "external" when the fit resulted in the external densities. This systematic fitting was performed based on the fact that at high cross-correlation (CC) values and density overlap, E1p from E3s at 22 Å resolution can be volumetrically distinguished (**Figure S8** C-F **Appendix**) and benchmarked thresholds for accurate fitting can be derived (**Figure S8** D-E **Appendix**). Since the flexible linker distances to their respective PSBDs were known after resolving the E2p/E3BP core, solutions that clash or violate these distances were discarded.

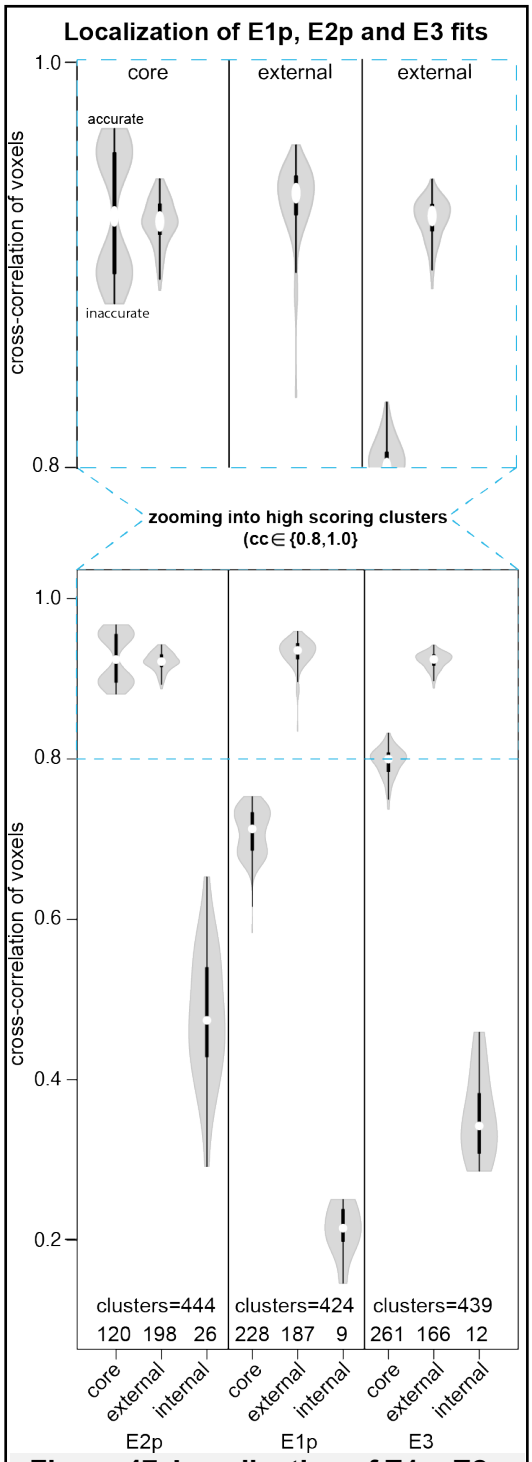


Figure 47. Localisation of E1p, E2p and E3 fits.

Systematic fitting of low-pass filtered E1p, E2p, and E3 enzymes in the reconstruction (**Figure S8** C-D). $E1p_i$, $E1p_{ii}$ and $E3_i$ correspond to the best-fitted conformation. (Source: Figure reproduced from [140])

As expected, E2p is localised in the core with high CC values. In contrast, it fits with lower CC to core densities (those being at the inter-trimeric interface), external densities, or entirely inside the core. Higher CC values for E1p and E3s matched to solutions at the non-

core densities (**Figure 47**), with E1p showing higher values compared to E3, maybe pointing to its abundance on non-core densities. Fit groups with high volume overlap with the map are displayed in **Figure 48** and ranked with CC values above threshold, corroborating their accurate placement (**Figure S8** D-F **Appendix**). E1p, with its conical overall shape, fits better in the conical density of the reconstruction (location 1, $E1p_i$ position) and gives an extra high-ranking fit in a less resolved density but of a similar conical shape (location 3, $E1p_{ii}$ position). For both of these densities, E3 fits are not satisfying and yield CC values below the threshold representing accurate fits and they are closer to the vertices than the faces of the native core. Above threshold values and high-ranking fit solutions for the E3 (CC>0.935) were observed close to the E1p fits, in a density rather elongated and over the face of the dodecahedral core (location 2, $E3_i$ position) (**Figure 48**). Position $E3_i$ yielded no significant solution for E1p (**Figure S8** F **Appendix**). Solutions for E1p and E3 were recovered with high CC values for all three positions due to the C2 applied symmetry and showed that the map holds information that can discriminate symmetric solutions. Based on similar symmetry analysis, the hNup107 subcomplex was correctly localised in the nuclear pore complex structure [47]. Reconstructions revealed no significant direct interactions of E1p or E3 subunits with the E2p/E3BP core, possibly due to the highly dynamic nature, rare or minority set of these events during the overall reaction. Overall, this observation further validates that E1 or E3 lipoyl domain interactions with the core must be highly transient, transferring the substrate acyl group rapidly to CoA and then releasing. In any event, the core would be expected to favour binding of acetylated lipoyl domains, likely in the presence of CoA and neither of them was added in the buffer to be then observed with cryo-EM imaging; instead, the active, native cell extract was vitrified which included the enriched PDHc metabolon.

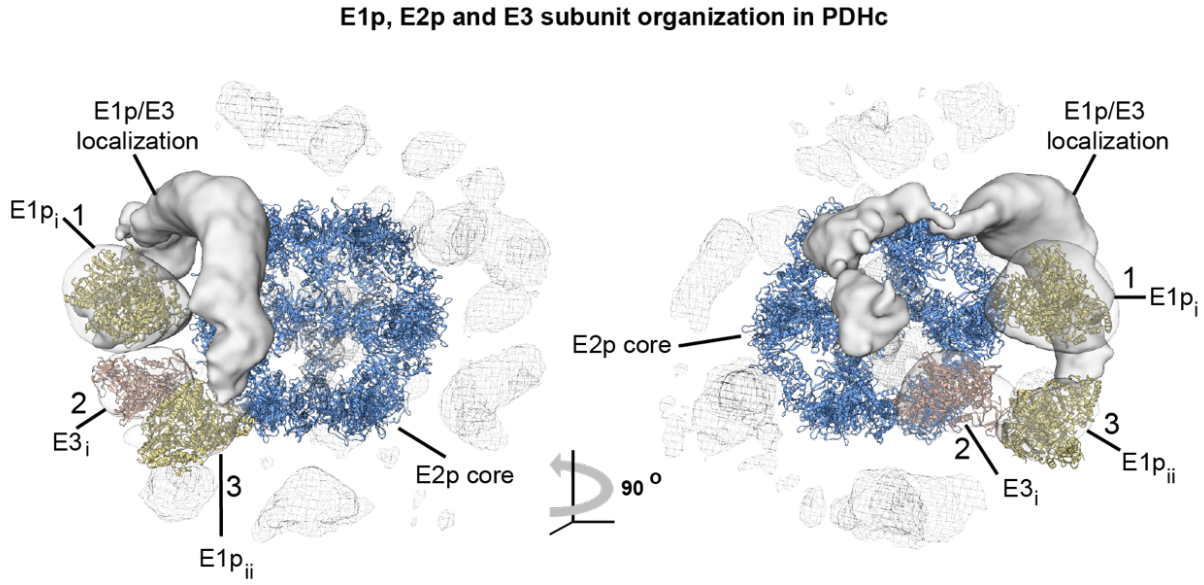


Figure 48. E1p, E2p and E3 subunit organisation in PDHc.
Top fits are shown for E1p, E2p and E3; The rest of the densities are not fitted due to ambiguity in localising the enzymes. (Source: Figure reproduced from [140])

3.3.2 Integrative structural biology reveals a localised nanocompartment proximal to the PDHc core - The "pyruvate dehydrogenase factory"

Considering that this could be a complete pathway for the PDHc reaction, forming a temporal reaction chamber, a model dubbed "dehydrogenase factory" (**Figure 49**) was introduced based on the above-mentioned results and the known biochemical data, for the first time. The density cloud composed of single E1p and E3 subunits restricts the flexible parts of the dehydrogenase factory as shown by cryo-EM (**Figures 46, 48** & **49**). The possibility of the presence of PDH-specific kinases and/or phosphatases or other binders within these densities cannot be eliminated, whereas on the other hand, based on the MS data (**Table S3 Appendix**), PDH kinases and phosphatases were of very low abundance and no other known binders of PDHc were identified. Therefore, the observed densities of the dehydrogenase factory must include E1 and E3 proteins only and possibly, LD domains stemming from the flexible arms of the E2p and the E3BP.

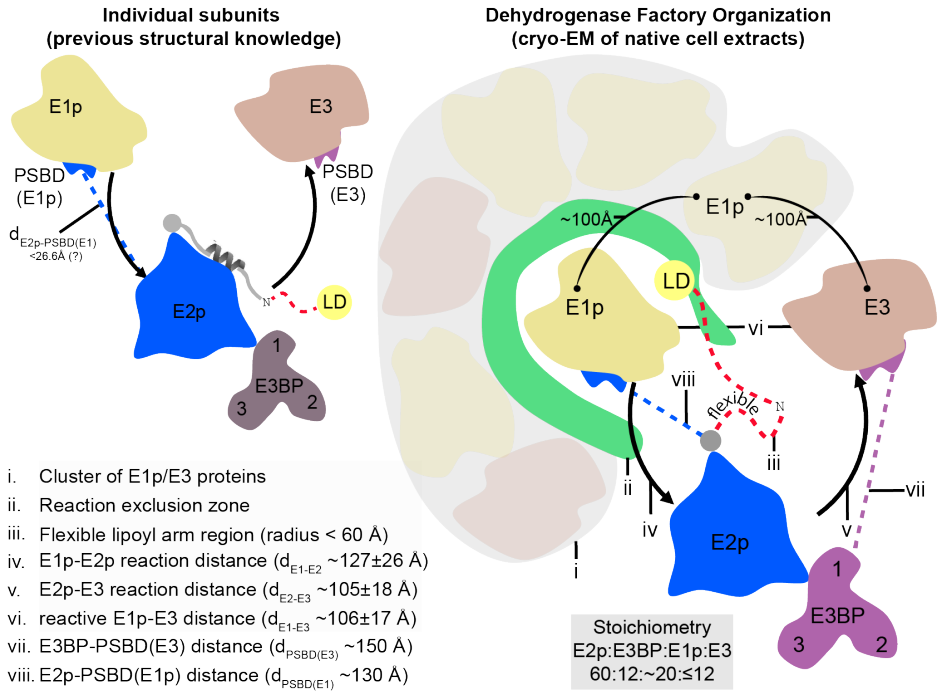


Figure 49. Previous knowledge about PDHc (on the left) and the knowledge derived in this thesis (Dehydrogenase Factory organisation, on the right).
On the left, the model of PDHc is shown with the LD *C-ter* stably attached to the E2p core, covering the binding site of the lipoyl. The arm transfers intermediates from E1p to E2p and then to E3. E3BP is described inside the E2p, but its intra-molecular interface is unknown. The pyruvate dehydrogenase factory model observed with the integrative methodology on the right. (Source: Figure reproduced from [140] with some colours modified)

E1p and E3 do not interact directly with the E2p core in the proposed model. Still, they are spatially confined in relative proximity specifying in such a way the flexible E2p and E3BP unresolved *N-ter* regions, which tether the E1p and E3 subunit to the core. Distances like the E1p-E2p active site minimum distance (d_{minAS}) of ~65 Å and the E2p-E3 d_{minAS} of ~75 Å are in the range of both flexible linker regions (core-PSBD; PSBD-LD; **Figure S5** B **Appendix**). On the same page, the E1p-E3 d_{minAS} of 80 Å is in the linker range as well (PSBD-LD). Additionally, the emerging distance between the last residue of the unresolved region of PSBD and the first residue of the E3BP core (Leu215-Ala271) is in a realistic range (138.8 Å or 151.1 Å), agreeing with a compacted state of the 55 residues structuring this region. Overall Euclidean distances are not sterically obstructed, especially the measured distances for E3BP cross the face of the dodecahedron unperturbed. A continuous reaction exclusion zone of ~100Å with an absence of observed densities is revealed in between the E1p, E2p and E3 enzymes, which possibly accommodates the flexible lipoyl domains and define a trail for the delivery of intermediates between the E1p, E2p and E3 active sites, forming the catalytically active reaction chamber of the pyruvate dehydrogenase factory.

4. Discussion

Years ago, it would have been difficult even to predict approaching a level where the biomolecular content of native cell extracts could be probed at a detail achievable in this thesis given their intrinsic complexity and heterogeneity. The prime example of this work, the Pyruvate Dehydrogenase Complex metabolon, was structurally characterised in its active endogenous state to demonstrate the usefulness of researching cellular function within native cell extracts. It is becoming more evident that native cell extracts hold a great promise in addressing the structural organisation of in-cell processes at an increasing level of detail. Of course, limitations are still there, meaning that cell extracts might not answer the complex biology at the organism level. Still, structures investigated in the cell extract environment pose the closest to native architecture compared to other high-resolution structural approaches and, as shown in the Results, provide insights into structural and architectural data that cannot be captured with other structural methods. This is because cryo-EM analysis does not necessarily require high enrichment as a more direct visualisation method: although the PDHc metabolon was enriched in the extract, it was present at comparatively lower numbers than other protein complexes in the imaged extract (**Figures 28** & **29**).

Among structural biology methods applied nowadays, cryo-EM holds an overall great potential: Its ability to study cell extracts and resolve structures from cellular fractions at near-atomic resolution is gradually established. Hand in hand in this direction, complementary proteomic methods are being reported that can handle such types of samples and may provide valuable data to understand cell extract biomolecular content and associated biomolecules [200]. Although no XL-MS data were generated in this thesis, a statistical treatment of previously published data [57] corroborated the molecular models of the *C. thermophilum* enzymes and their oligomerisation state. As also shown in the Results, docking approaches, combined with unsupervised fitting in the cryo-EM maps, and, when higher resolution is achievable, refinement and optimisation of biomolecules in the derived Coulomb potential map, can provide unprecedented insights into the PDHc metabolon core, internal architecture, as well as the overall proximity of E1p, E2p, E3BP and E3 subunits. In the future, cellular and molecular biology can provide further insights and strengthen the study of less abundant biomolecules within cell extracts, possibly even the PDHc-specific kinase and phosphatase biomolecules, which were not detected in

the native cell extract investigated in this work. However, overall, in combination with traditional structural methods, cell extracts can enhance our understanding of cellular processes.

We are gradually entering into ages where the structural characterisation of transient higher-order assemblies could become a common practice. The knowledge from such breakthroughs will help the scientific community further optimise and design molecules of great biotechnological and medical potential by justifying and understanding the underlying molecular origins of function and disease. Below, a discussion on the impact of the results from this thesis in the light of current developments in various experimental and computational methods, and especially in image processing, is provided. The discussion concludes with the functional implications of the discovered architecture of the "pyruvate dehydrogenase factory".

4.1 On the enrichment and the molecular characterisation of OADH complexes in native cell extracts

Recent advances in biochemical, biophysical, structural, and computational methods provide structural insights for biomolecular assemblies within cellular homogenates and thus a deeper understanding of their relationship, like the reported "dehydrogenase factory" organisation. Cell lysates are shown to retain functional aspects that can be easier manipulated, *e.g., C. thermophilum* extracts were optimised to perform pyruvate oxidation by retaining the functional PDHc metabolon. Other cell extracts can be studied utilising the methods reported in this thesis, such as those utilised as biotechnological tools for bioproduction [201], cell-free gene expression, transcription and translation [202] and, recently, molecular design [203]; These cell extracts have not yet been probed structurally; therefore, understanding the function of these cell extracts in detail by probing their intrinsic structure will be advantageous.

Up to now, our knowledge around cell extracts has been limited to low-resolution data [204]. Still, new methodological advances can already provide resolutions of 4.7 Å for protein communities and their members [57], and up to 3.85 Å, as reported in this thesis for the PDHc core (**Figure 39**). Even more recent studies increased the achievable resolution [155,205,206,] in particular for nuclear [206] and membrane extracts [205]. This thesis, in addition, also reports determined snapshots of higher-order organisation of in-extract flexible, functional metabolons (**Figures 30, 32, 41, 46** & **48**). In many

recent reviews [56,157,207] integrative structural studies of native extracts benefits, importance and challenges have been reported, and the correlation between structural disorder and enzymatic function for in-extract metabolons has been described. The core advance of this work is the direct visualisation and functional characterisation of an active, endogenous PDHc metabolon with multiple methods not previously combined to achieve such a result. However, various resulting data reported here can be further analysed in the light of recent methodological developments in the fields of biochemistry, structural biology and image processing.

4.1.1 Integration of biochemical, biophysical, and structural methods for understanding cell extract content

Investigating cell extracts and probing their biomolecular content includes a broad toolbox and a wide range of biochemical treatments and further biophysical analysis (**Figure 50**).

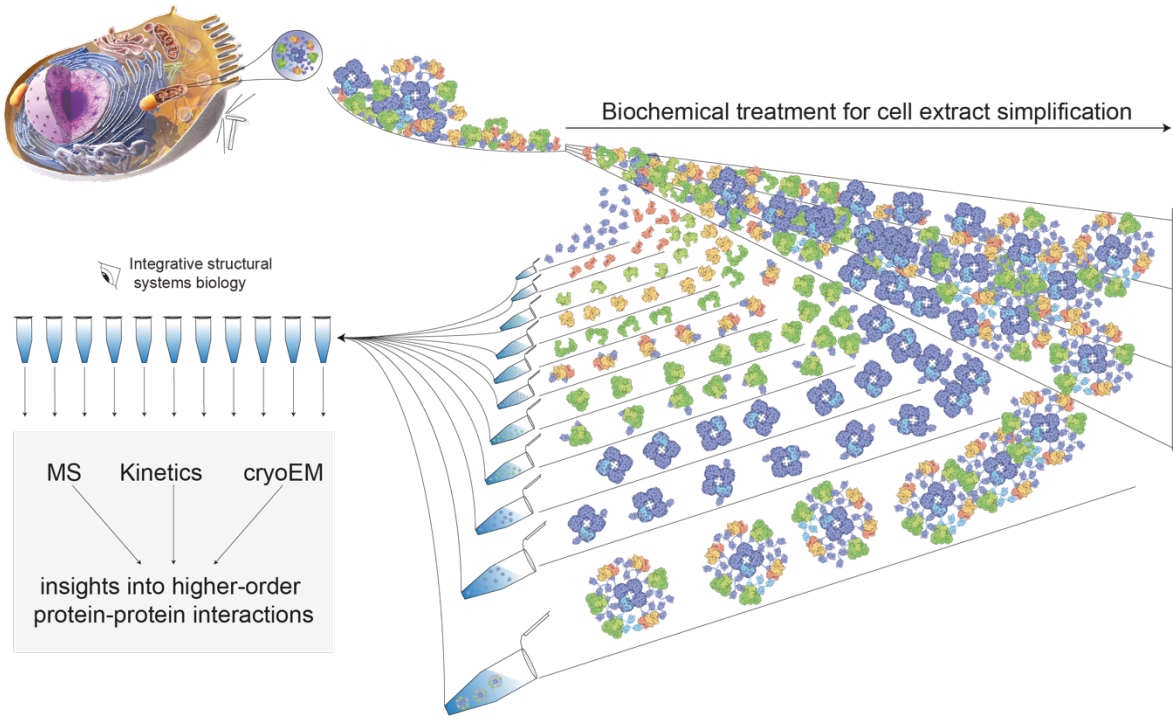


Figure 50. Methods to experimentally extract identity, structure and dynamics information of protein communities.
In short, the cell is lysed, and subsequent fractionation is applied to recover co-eluting protein material. In a large-scale manner, mass spectrometric, kinetic and cryo-EM analysis of the fractions leads to the characterisation of protein communities in native cell extracts. The example of the Pyruvate Dehydrogenase complex (PDHc) metabolon is shown. Molecular representations for PDHc are retrieved and further edited from PDB "Molecule of the month" section (Source: Figure reproduced from [208])

This toolbox, including high-end, modern equipment (**Table 2**), has not only been applied in this thesis to gain deeper insights into the biomolecular content of native cell extracts (**Figure 8**) but also focused on providing even more detailed analysis in metabolons involved in pyruvate oxidation, *i.e.,* PDHc and OGDHc (**Figures 24-26**). Experimental proteomic identification utilising mass spectrometry (MS) can provide information for thousands of protein sequences [209,210]; This information was harvested in this thesis from the published data [57] to annotate all protein community members and their co-elution in size exclusion chromatography fractions, showing that all component proteins can be found in the same biochemical fractions (**Figure 25**). However, such MS analysis only provides a list of proteins that might include detailed information on their relative abundance but not their interactions. The same detection, but supervised (*i.e.,* knowing which proteins must be found in the eluting fractions utilising MS), has been applied utilising immunoblotting identification of protein components from OGDHc and PDHc, further validating their abundance and specific elution profile during size exclusion chromatography (**Figure 24** & **27**). In particular, the antibodies produced in this study against all PDHc components (**Figure 27**) might serve in the future many more studies probing, for example, proteins of PDHc and their inter-relations in nuclear extracts [211,212].

Despite this, various other protein components might be additional binders of OADHc metabolons but were not identified. This can be addressed by performing a detailed network to predict more protein communities and their members in the cell extracts, combining external interaction data. Such proteomic analysis has been performed, for example, for interconnected yeast complexes discovered by tandem affinity purification (TAP) and mass spectrometry (MS) [59]. More recently, characterisation of protein-protein interactions (PPI) with experimental and computational methods combining *in vivo, in vitro* and *in silico* data [213] was also applied. Indeed, higher-order assemblies of PDHc have been identified in the cryo-EM micrographs where those additional bound biomacromolecules manifest as structural signatures in the cryo-EM data (**Figure 51**).

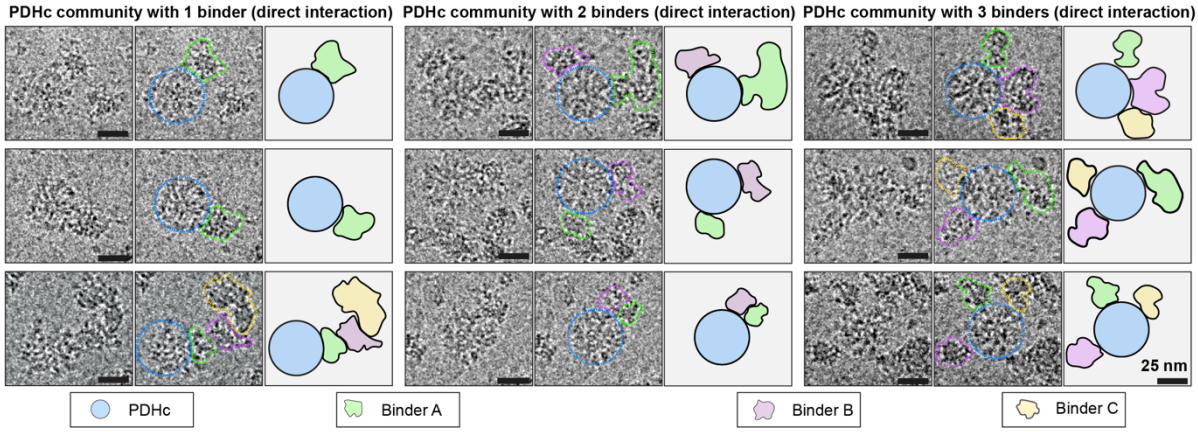


Figure 51. Higher-order interactions of the pyruvate dehydrogenase complex within native cell extracts captured by cryo-EM.
PDHc is occasionally found in protein communities with one, two, three or more MDa protein complex members when visualizing single particles that ended up in the PDHc reconstructions directly from the micrographs. PDHc may participate in a "pearl-string-like" arrangement, localizing at the edges of the community (bottom left panel). This architecture is reminiscent of other previously reported communities, *e.g.,* the communities of FAS with its binders [57] . Interactions of other MDa complexes with PDHc provides evidence for another layer of complexity regulating metabolons beyond the identified dehydrogenase factory. All scale bars are 25 nm. (Source: Figure reproduced from [140])

4.1.2. Biochemical and proteomic characterisation of protein communities in pyruvate oxidation: Implications for the discovery of additional communities

Various proteomics methods have been reviewed recently [214] as efficient ways to address and identify the protein content of fractionated cell extracts. Previous studies measuring enzymatic activities across cellular fractions, *e.g.,* probing the interactions of Krebs cycle enzymes in fractionated extracts of *E. coli* [215], has triggered the application of co-fractionation to monitor protein associations [71,79]. In this thesis, one of the main methods nowadays for efficient and well-resolved separation of cell extracts, *i.e.,* high-resolution size-exclusion chromatography (SEC), coupled to previously communicated MS data [214] and kinetic data for identification of the PDHc and other OADHc metabolons were applied.

Advantages of this workflow include: (a) simplification of cell extracts according to an intrinsic physical property of the contained biomolecules and biomolecular complexes, *i.e.,* size; (b) per-fraction quantitative data regarding protein abundance and co-detection utilising MS; and (c) robust per-protein elution profiles across the studied fractions, which may be used for later integration into a PPI network by optimising protein elution profiles. Interactors within protein communities can be

identified in this protein co-fractionation approach [57]. Recently, a comparison of PPI networks across species was constructed, highlighting in such a way similarity of higher-order assemblies in an evolutionary manner [60]. This thesis provided a solid example for a deeper characterisation of metabolons involved in pyruvate oxidation, identifying co-eluting proteins for OADHc (**Figures 23** & **28-30**) and understanding the community members by analysing the derived MS data (**Figure 25**) validated via WB (**Figures 24** & **27**). Given a PPI network, deeper insights into its communities and their members can be derived utilising the above-mentioned methods (**Figure 52**). Note that the final interaction network for *C. thermophilum* native cell extracts was previously communicated [57] and this thesis provided deeper analysis in the OADHc communities, being part of this PPI network. As a perspective, in the future, the identification and characterisation of a larger number of protein communities and their members are expected by utilising new combined computational and experimental methods, and as performed in previous studies by the integration of machine learning approaches, *e.g.,* random forest approaches [57]. One of the major challenges that remain would still be to recognise random co-elution events and distinguish them from real protein complexes and interactions, which is a purely computational task augmented by experimental validation.

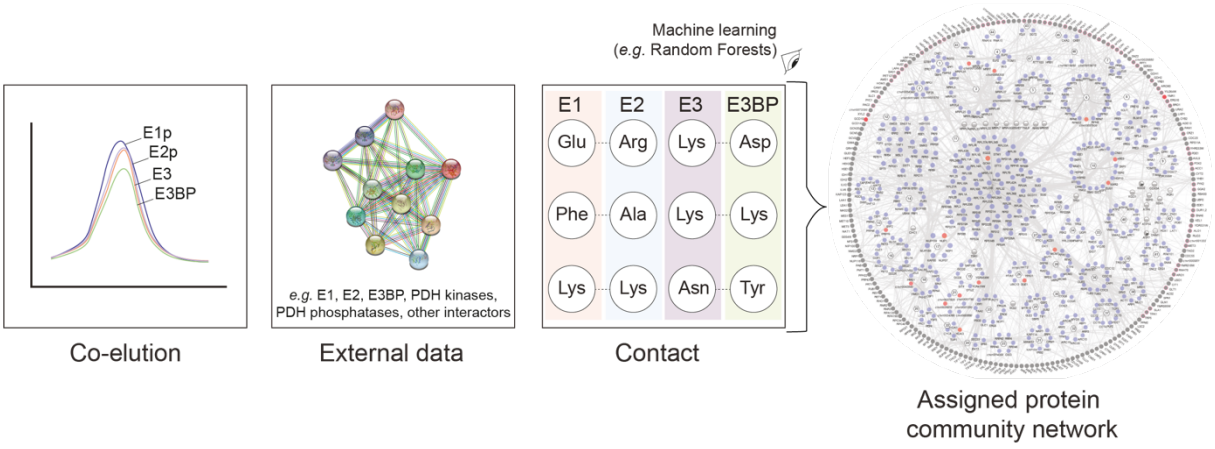


Figure 52. Detection of protein communities using experimental and computational data and their integration in an assigned protein community network.
Combined data regarding protein-protein interactions stemming from fractionation (co-elution), external database information (network data) and contact information prediction (*e.g.*, from co-evolution analysis, chemical crosslinking or mutagenesis experiments) among community members are used for deriving a protein-protein interaction network, even augmented by machine learning, *e.g.*, using a random forest. Finally, a network with interconnected protein communities is derived and insights into community members can be retrieved. External data shown are extracted from STRING (https://string-db.org/) and network shown from [57]. E1, E2, E3 and E3BP are the proteins structuring the 10-MDa complex of the PDHc metabolon, all involved in the complex reaction of pyruvate oxidation. (Source: Figure reproduced from [208])

4.2 cryo-EM of cell extracts and structural interpretation of endogenous protein communities and their members

4.2.1 Performance of traditional cryo-EM image processing methods for the visualisation of structural signatures of protein communities

Protein identification utilising cell extracts as a specimen can provide information regarding the sequence of the community members or their interactions, but not for the structure; thus, cryo-EM imaging of native cell extracts stands as an essential tool for the structural characterisation of protein communities. This was exactly what was performed in this work (**Methods Chapter 2.2.7**), after systematic optimisation of vitrification conditions and imaging the vitrified specimen under low electron dose, revealing an incredible complexity of very large biomolecular complexes (**Figure 23** B). However, even if these signatures are captured in the cryo-EM data, the readout could be limited to understanding the protein community architecture due to the absence of stoichiometric data. Derivation of stoichiometric data is actually not a trivial task. In the current thesis, a combination of cryo-EM, immunoblotting data, and analysis of published MS and crosslinking MS data in fractionated extracts was performed, and approximate stoichiometries for the higher-order structure of the endogenous PDHc were derived for the first time, reporting overall, 176 distinct polypeptide chains in a stoichiometry of 1 E2p 60-mer, 4 E3BP trimers, ~20 E1p tetramers and max. 12 E3 dimers. As a more direct method to observe cell extracts, electron microscopy has also been applied at low-resolution by integrating mass spectrometry data [204], but in this thesis the advancement was to derive novel, additional densities surrounding the ordered OADHc cores, pointing, therefore, to novel architectural principles of OADHc metabolons (**Figures 28-30**).

Nevertheless, via recent hardware and software advances in cryo-EM [31], recent high-resolution data [57,140] revealed novel protein communities involving, for example, fatty acid synthase (FAS) [57] or pyruvate dehydrogenase complex (PDHc) bound to other megadalton complexes, as shown in the current thesis. In addition, other results from studying cell extracts showed that abundant stable complexes could be *de novo* reconstructed from high-resolution cryo-EM data [155]. The current work is distinct from the previously communicated method [155], because the aim of this thesis is to

understand structures of protein communities and not of individual complexes. In addition, the thesis emphases on understanding architectural aspects of protein communities in pyruvate oxidation and does not focus on reaching the highest possible resolution by refining the most stable parts of a captured single particle. However, this was a challenging task because most developed algorithms for cryo-EM image processing and density reconstruction apply to individual, highly purified protein complexes rather than higher-order assemblies and their participation in native communities, where, when those image processing algorithms are applied, additional densities could be averaged out.

Another challenge that was surpassed in this thesis is the utilisation of the inherent low-resolution contrast of cryo-EM micrographs. Indeed, cryo-EM micrographs contain two-dimensional projections of the particles in different orientations but at low contrast, often including undesirable features or even contamination (see, *e.g.*, **Figure 53**). Even if multiple short exposures are recorded in cryo-EM, the signal-to-noise ratio is approx. 0.1 – almost comparable to the one existing in astronomy. The traditional methods of improving the contrast, like Wiener filtering or bandpass [216-218] are not that sensitive to the underlying noise properties. Despite those challenges, efficient 3D reconstructions were eventually reported in this thesis (**Figure 33, 34, 36, 39, 41, 46** & **48**). In addition, it is evident that nowadays, machine learning models within cryo-EM have a better capacity for training in noise characteristics learning and denoising at a better level than before ([219] and refs. therein). Such advances can be integrated in the future to understand the interconnectivity of protein community members in cell extracts and better define their structural inter-relations.

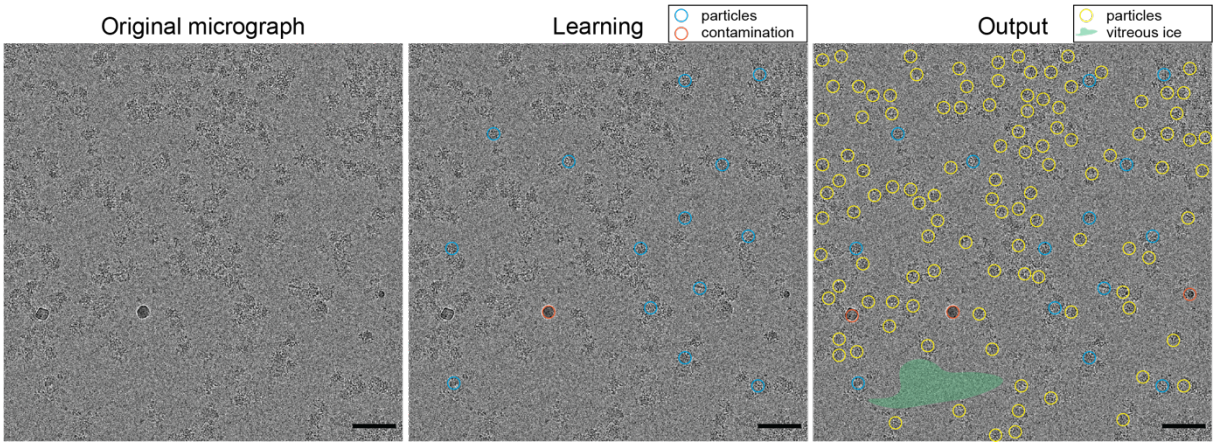


Figure 53. Machine learning to discover single particles in micrographs from native cell extracts.
A cryo-electron micrograph from *Chaetomium thermophilum* fractionated cell extracts is shown. During machine learning that could be applied in the future to the derived data of this work, the algorithm is being trained to discriminate particles from contamination, vitreous ice,

aggregation and noise. In the end, the algorithm optimally picks and selects learned features that were not previously recognised during learning. Red circles indicate contamination and blue and yellow circles indicate learned and predicted particles. The size of circles does not match particle size but represents correctly picked particles. Green highlighted area signifies regions of vitreous ice recognised by the algorithm. (Source: Figure reproduced from [208])

4.2.2 Future improvements in particle detection in cryo-electron micrographs via machine learning

In most traditional template-based approaches, *e.g.,* [220], particle candidates are considered by estimating similarities and cross-correlating between a reference (also known as a template) and an image region; The same idea was applied in this thesis to pick all particles from the negative-stain (**Figure 28**) and cryo-EM data (**Figures 23** & **38**), but instead of providing a template of the structures, contrasty regions of the micrograph were automatically picked via blob picking [221]. This means that to achieve a more precise particle selection, a template-matching technique with a cross-correlation similarity measure is usually employed in such approaches. However, still, there is a chance of introducing template-based bias and a high rate of false positives. A negative outcome could be that matching is performed over random regions, for example, over noise only, this noise will unavoidably be interpreted as a pattern leading to a phenomenon described as "Einstein-from-noise", introduced by *Glaeser* [222]. Although this effect was carefully evaluated during image processing in this thesis by recovering averages of much higher resolution than when selected templates, in the future, machine learning can drastically help in minimising such phenomena. For example, to select reference-free desirable regions, many deep learning algorithms have been developed [114,223-229]. Inspired by computer vision applications, they use convolutional neural networks (CNNs) [227,229], to demonstrate a per pixel-image segmentation of particle/non-particle regions (**Figure 54**). The main focus of designing these architectures is to avoid as many as possible undesirable features or learn to exclude them completely [114,223,225,226,228] and it is a big success that most developed recent machine learning and deep learning-based methods have immensely improved in terms of accuracy and low false-positive rates [114,223-229]. Advances in such methods can further help in recovering not only single particles of protein communities but also their individual members by identifying peripheral densities, even if the signal-to-noise ratio is decreased due to the smaller size of the peripheral proteins, *e.g.,* in the case of the PDHc metabolon.

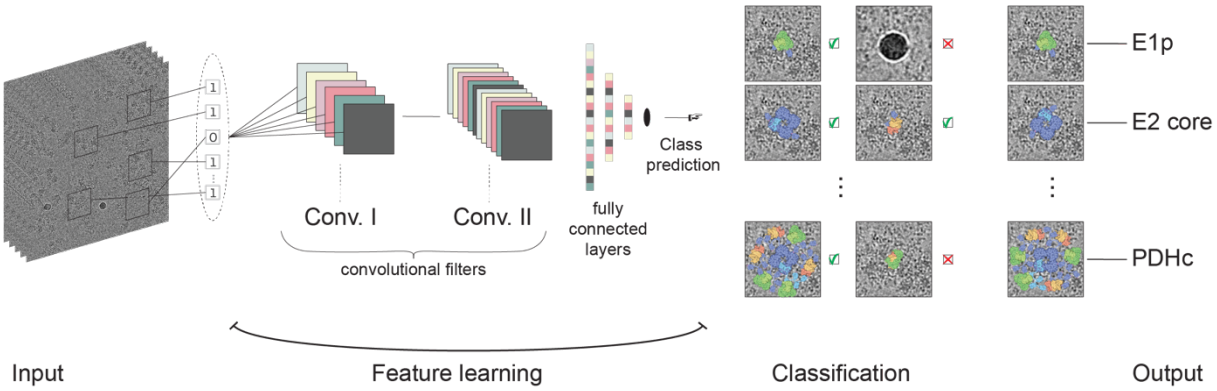


Figure 54. Convolutional neural network for image classification in heterogeneous cryo-EM samples.
Structure of a convolutional neural network algorithm frequently used to detect signals in cryo-EM micrographs. Input micrographs are used for feature learning during the convolution step of algorithm training. Optimal training would lead to efficient classification of single particles and/or their higher-order assemblies and discriminate those from noise, contamination and aggregates. A final output is achieved with metabolon members in their unbound and bound states as recognised by the CNN in heterogeneous cryo-EM micrographs of native cell extracts. (Source: Figure reproduced from [208])

CNNs can be declared as the most efficient models for image classification and particle picking since they are trained in learning thousands of 2D filters [230]. Keeping in mind that templates can be declared as filters if these algorithms are trained in the more complex environment of a heterogeneous mixture of cell extracts and not only in the purified to homogeneity single particle datasets, there is a high probability that they will be able to detect particles of varying shapes and sizes effectively and efficiently discriminate them from the artifacts in the micrographs of cell extracts to retrieve members of protein communities systematically.

4.2.3 Improvements in image processing of 2D projections and 3D reconstructions derived from native cell extracts

Many issues have to be addressed after the efficient recovery of single particles from heterogeneous specimens. The trained algorithm should segregate and cluster the particles into correctly assigned classes, incorporate rotational and Contrast Transfer Function (CTF) invariances and proceed to multiple 3D reconstructions yielding from heterogeneous 2D projections. In general, conventional cryo-EM classification methods can perform this in homogeneous specimens, but in the case of protein community discovery, the inherent higher variability has to be addressed for correct 3D shape classification and reconstruction. Although this work efficiently reconstructed two 3D shapes simultaneously out of the native extract (PDHc and

OGDHc), classes corresponding to each molecule were manually selected and merged after testing various combinations with the objective to increase global map interpretability. Current classification strategies, however, aim to achieve the highest possible resolution by optimising data homogenisation. This notion is itself contradictory to the understanding of protein communities and the interactions of their members. Focusing simply on retrieving high-resolution may lead to averaged densities of particles that may or may not participate in the same communities during data analysis, averaging chemically, stoichiometrically or conformationally distinct protein communities. This effect has been addressed experimentally in this thesis by correlating the 3D cryo-EM volume of the PDHc that was reconstructed with the expected volume of all 176 polypeptide chains that form this molecule, and both estimates agreed. The subsequent computational analysis also showed that the final 3D map of the endogenous PDHc metabolon could be interpreted and agreed with biochemical data regarding the proximity of its component proteins. However, such validation measures are sophisticated, and to avoid guessing the number of underlying 3D shapes present in the data, *Verbeke et al.* [156] applied projection-slice theorem principles to group particles into consistent subsets prior to 3D classification.

Even if the number of distinct conformational states of any biomolecule is usually unknown, during reconstructions, current methods often use cryo-EM data to assume that heterogeneity originates from a small number of independent, distinct states. A deep neural network-based algorithm called `CryoDRGN` [231] tries to solve this issue by continuously considering and estimating a 3D density function of the processed single particles. Due to the low number of OADHc particles of protein communities identified within *C. thermophilum* extracts, such a method is not possible to create such continuums for the reported data of this work. However, in the future, a greater number of data could be acquired utilising faster and more efficient detectors and therefore, classifying or discovering conformational states of protein communities could be a valid investigation.

Resolvability of reported PDHc cryo-EM maps could also be revisited in the future. More recent generative adversarial networks (GANs) trained on pairs of 3D atomic models and their noise-free cryo-EM maps seem to provide better results than traditional methods for map reconstruction followed in the thesis; GANs may generate a more realistic ground-truth 3D density map [232] improving this way the protein density of experimental cryo-EM maps. A deep investigation of these aspects and

implementation of neural networks for simulated cryo-EM 3D reconstructions [233] is performed by the *Scheres* laboratory. Furthermore, new machine-learning algorithms have been developed for post-processing cryo-EM maps and addressing resolution anisotropy [232,234]. Advances in those 3D reconstruction and refinement machine learning methods are expected to also impact cryo-EM processing of highly heterogeneous specimens, as those reported in this thesis, further improving both achievable resolution and localising more protein community members in better defined cryo-EM densities.

4.2.4 Structure prediction-based model building in cryo-EM maps from native cell extracts

Cryo-EM can provide high-resolution information for protein complexes when purified constructs are used. Still, when it comes to native cell extracts, things may become more complicated and challenging as the identity of the protein complexes and their interactors might be completely unknown and cannot be accurately matched to the MS data reported for the same specimen. Although cryo-EM can be applied to visualise protein communities, still, integration of complementary biochemical, proteomic, literature and other data into validated hypotheses are mandatory for interpreting the cryo-EM maps from protein communities.

The additional flexibility observed in protein communities, as the one derived for, *e.g.,* the external densities for OADHc metabolons (**Figures 30, 32, 46** & **48**), contributes to further challenges. This is because determining 3D models of isolated flexible biomolecules is already a challenge itself, and thus higher-order, flexible metabolons with chemical and conformational heterogeneity are even more difficult to determine. Only a few examples of discovering and structurally analysing protein communities retrieved at sub-nm resolution from native cell extracts are available, like that of FAS [57] and PDHc, reported in this thesis.

Methods like cryoID [155], which may perform *de novo* model building, are already available given that a high enough resolution is achieved and side chains are visible when of course, the proteome of the studied organism is also available. If the resolution is lower than ~4.0 Å, then side chain resolution is unattainable, and modelling methods must be ultimately employed (*e.g.,* [115,235]), as also the ones utilised for atomic modelling of the OADHc metabolons reported in this thesis. Another option, besides careful integration of molecular identity data to explain the cryo-EM density maps, can be, as proposed in [236], fold recognition within the reconstructed maps. Recently, the

Kastritis laboratory derived a *de novo* identification, fold recognition and machine-learning based approach to unambiguously characterise the structure of protein community members within their reconstructed cryo-EM maps, even when resolutions were worse than 4 Å, as in the case of the OGDHc core, which was resolved at 4.52 Å [237]. Such applications may become common practice in the future, especially when characterising protein communities and their members. Finally, when complexes include non-protein components, this task becomes even more challenging, but neural networks are being developed to localise nucleotides as well [238]. Although this is not the case for the pyruvate oxidation metabolons, future protein communities, including DNA or/and RNA, could exploit such functions for cryo-EM map interpretation.

Of course, to reach this level of interpretation of a cryo-EM map directly from cell extracts, protein complex abundance in the recovered cell extract fractions is of paramount importance. If the resolution is not high enough for *de novo* interpretation of the cryo-EM map, protein complex abundance within the sequential fractions can be correlated to the corresponding structural signatures that were recovered by negative staining or cryo-EM. This method assigns an identity to these signatures, which are also members of their respective protein communities [57]. Such an indirect approach has been previously applied for assigning abundant species of *C. thermophilum* complexes using simple cross-correlation functions [57] and was also shown to predict the abundance of OADHc complexes in this thesis (**Figures 24-27**).

In theory, if a T-squared distribution within a particularly thick micrograph (1,300x1,300x200 nm, pixel size of 3.17 Å) is assumed regarding the abundance of distinct single particles (**Figure 55**, left panel), then there can be an estimation for their relative abundance (**Figure 55**, left and middle panel). Suppose cell lysis is not applied as for example, in cryo-electron tomography of a cell section: To reach the required number of particles (~5,000) for FAS [57] or PDHc, reported in this thesis, for less abundant particles, a high number of tilt series will be required (**Figure 55**, right panel). This effect can be further magnified and become more intense in the case of native cell extracts if no further biochemical enrichment is applied to the sample because of dilution; thus, it is essential to note that in this thesis, protein complexes were selectively biochemically enriched, and therefore, conservative estimates are shown in **Figure 55**, right panel. This biochemical enrichment that was achieved by utilising cell extracts as a function of their elution in SEC was key to resolving cryo-EM

structures of pyruvate oxidation metabolons and their community members because enrichment was achieved according to size.

Conclusively, rare species will be algorithmically challenging; they are more difficult to capture by default, and usually, a considerable amount of data is further required. Therefore, abundant complexes are more likely to be captured and abundant community members to be structurally characterised. This key observation was applied for resolving, here, the communities involved in oxo acid metabolism.

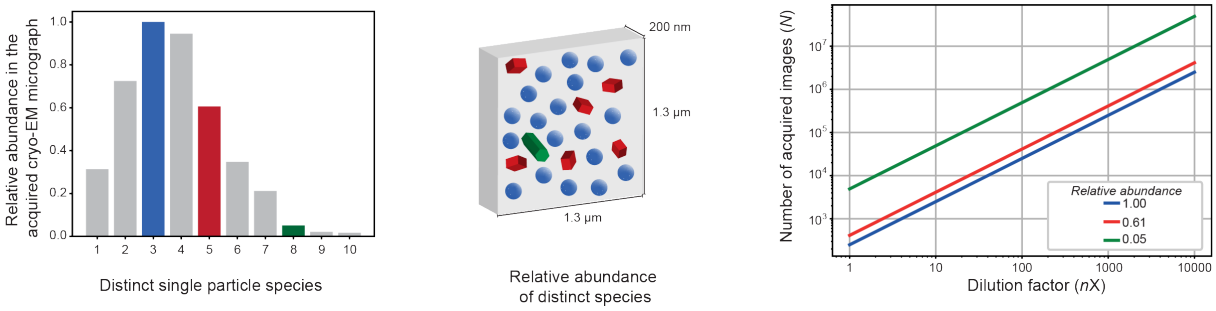


Figure 55. Conservative probabilities for particle detection based on abundance and dilution factor.
On the left, an example of 10 distinct single particle species is shown with their relative abundance following an assumed T-squared distribution. In the middle, illustration of relative particle abundance for three distinct particles (blue, green and red, representing high, medium and low abundant species in a calculated 4Kx4K micrograph with a pixel size of 3.17 Å and thickness of 200 nm. On the right, dependency of the number of images required to reach ~5,000 single particles on the dilution factor is shown (assuming no biochemical manipulation for particle enrichment). (Source: Figure reproduced from [208])

4.3 On the asymmetric cryo-EM reconstruction of the endogenous, 10 MDa PDHc metabolon

Overall, this work has shown that distinct image processing methods have been applied to reconstruct the cryo-EM maps of PDHc from native cell extracts: These methods included particle detection, 2D and 3D averaging and, finally, multiple methods to optimise and refine the retrieved 3D cryo-EM maps. In this methodology, traditional image processing approaches have been employed, mostly in RELION [10] and Scipion [166]. In the future, re-calculation of the acquired cryo-EM data utilising novel machine/deep learning approaches might eventually resolve profounder inter-relations of protein communities within the acquired micrographs. These new methods can find resolve issues related to the recovery and characterisation of protein communities at high-resolution, therefore, improving the interpretation of the endogenous PDHc metabolon and its core (**Figures 33, 34** & **Figures S7, S9** and **Tables S6** and **S7**).

To have an accurately predicted outcome of improved structural analysis of reported pyruvate oxidation metabolons, avoiding overfitting and performing cautious cross-validation, unambiguous and large training sets, true test sets, and overall systematic benchmarking are some of the prerequisites for developing any novel algorithm to handle tasks of image processing of heterogeneous data as those collected in this thesis. Besides improvements in single particle analysis of cell extract cryo-EM data, cryo-electron tomography could also be applied in the future, with the direct advantage of recovering low-resolution 3D reconstructions for each particle within the acquired tomogram. This 3D reconstruction from cryo-ET can serve as input for further discriminating 2D averages of single particle data and aid in their correct classification in distinct classes matching the tomographic 3D reconstruction. Such a method will resolve the ambiguity of "which 2D class average can match to which distinct protein complex". Moreover, analysis of protein communities of heterogeneous single particle datasets can be inspired by cryo-electron tomography of complicated specimens and related image processing methods for in-tomogram particle detection and classification [239-242].

As a perspective on the impact of this work on studying heterogeneous specimens, it should be pointed out that structural biology of native cell extracts may provide a complexity ideal for developing and testing artificial intelligence applications, as well as the transfer of image processing techniques from cryo-ET to single particle analysis to alleviate various challenges discussed above. Of course, one should keep in mind that the inherent flexibility and heterogeneity of biomolecules participating in protein communities set a realistic limit on the possible achievable resolution implying that high-resolution must not always be the priority when studying protein communities. If high-resolution is achieved for a protein community member, this might limit the structure/function interpretation of its interactors, which are flexible. An example is the PDHc core, resolved at 3.85 Å resolution (**Figure 39 & S9** and **Table S7**): Although the core was unambiguously assigned and structural insights into its endogenous structure were reported, at that resolution, all other densities for the proximal core biomolecules (E1p, E3, E3BP) were averaged out: This means that reaching high-resolution was excellent for understanding the core architecture of endogenous PDHc, but, to the expense, of understanding its inter-relations with the other members of its community. In this thesis, the PDHc reconstruction of the full complex was achieved and resulted in recovering densities for the complete 10-MDa complex: Although the

resolution was lower, it was sufficient to derive structural information about the interactions of the PDHc metabolon members. This combined approach of determining the structure of protein community members at high-resolution and understanding the architecture of the full protein community is inherently powerful because multiscale levels are eventually combined.

Finally, open-source data for biological systems will greatly accelerate the research on native cell extracts and the capturing of protein communities. In the future, dedicated databases like UNIPROT [243], PRIDE [244], CORUM [245], EMDB [246], EMPIAR [247], PDB [248] could increase/enrich the availability of datasets for heterogeneous specimens, like native cell extracts. This, in addition to the increasing availability of open-source data combined with the significant advances in computational hardware, will boost the efficiency of machine/deep learning algorithms for image analysis. Thus, many answers remain to be provided on how protein complexes interact and create all these interconnected communities. Machine learning methods can have a significant share in tackling some limitations mentioned above in investigating complex mixtures and homogenates of soluble or membrane extracts with notable success provided that available open-source data are gradually increasing in the field. Towards this direction, cryo-EM data reported in this thesis were made publicly available in EMPIAR (https://empiar.pdbj.org/entry/10625/), and the structural data are openly available through EMDB (EMDB ID: EMD-12181) and PDB (PDB ID: 7BGJ).

4.3.1 Relevance of the captured cryo-EM PDHc state.

This thesis reports high-resolution structural studies of native cell extracts with a focus on the PDHc metabolon. Here, a combined analysis of proteomic data with novel negative staining and cryo-EM data to computationally derive meaningful native architectures of the OGDHc and PDHc metabolons is performed. This thesis also connected activity profiles, which were never previously reported in such a context of a cell extract study, *i.e.*, were absent for the FAS complex [57]. It is mandatory that, when studying native cell extracts, the activity must be probed. This is because any purification or sample optimisation procedure may, ultimately, inactivate protein complexes within native cell extracts. Therefore, activity assays offer a view of what may constitute a functional protein community. In the case of PDHc, subcomplexes are E1p, E3BP, E2p and E3 (proteins directly involved in pyruvate oxidation partial reactions), and the metabolon, or protein community, is the entity performing the full catalytic reaction of pyruvate oxidation (PDHc). As mentioned in the discussion above

(**Chapter 4.2.4**), the study of cell extracts is ultimately confined by the relative abundance of recovered molecules. This is an important aspect to be considered when studying cell extracts not only of the same organism, but also from various organisms or cell types. Although, *e.g.,* ribosomes will always be abundant, other abundant biomolecules might be less enriched in other cell extracts by comparison. This, of course, limits quantitative comparisons of retrieved cell extracts and further deeper investigation of protein communities across cell types or species.

One might argue that various questions can be answered by studying protein organisation and architecture within cell extracts, and such methodology limits the impact of pure protein component investigations and their structure/function studies. However, this is not true: Cell extracts study the endogenous protein communities, while purified components study in detail the components of those protein communities. Therefore, data derived from pure specimens are extremely important for a deeper characterisation of the studied protein community because they are unambiguous. These data were used in this thesis to a very large extent: Firstly, pure specimens of OGDHc and PDHc components as positive controls for the kinetic assays were utilised (**Chapter 3.1.1, Results**). These served as critical observations for analysing the metabolon activities within the cell extract. Secondly, X-ray crystallographic structures of homologous proteins from other species were instrumental for deriving the atomic models of the equivalent proteins from *C. thermophilum* and their subsequent refinement procedures in the lower-resolution cryo-EM data. Without this data, accurate homology models of *C. thermophilum* proteins would have not been feasible. In addition, these models guided the understanding of the pyruvate dehydrogenase factory architecture and, ultimately, explaining structural aspects of the endogenous PDHc function (**Figure 49**). Thirdly, the pure constructs of E1α and E2p of PDHc were monitored via immunoblotting assays and the calculated amount was used as a proxy for estimating the approximate number of their copies within the endogenous PDHc metabolon. Lastly, the sequences of the pure constructs were used to calculate the participating domains and their distances based on homologous data. This information was pivotal to ultimately put forward the molecular architecture of the proposed pyruvate dehydrogenase factory.

As mentioned in the **Introduction Chapter**, cell lysis has several critical challenges that need to be addressed. Besides great dilution factors that might affect the protein interactions, aggregation could be introduced, loss of activity, *etc*. (**Chapter 1.4,**

Introduction). It would be extremely interesting to augment the results from this thesis for the PDHc metabolon with direct observations from *C. thermophilum* cells. This can be achieved by producing lamellae of cells utilising a (focused ion beam milling) cryo-FIB Scanning Electron Microscope (SEM), especially focusing on the organelles of interest (for PDHc, of course, this is mitochondria). However, extreme concentrations have been reported for proteins residing within mitochondria (up to ~560 $mg \cdot mL^{-1}$; [249]), and this is prohibitive for identifying any biomolecule residing within the imaged lamellae with cryo-ET. Ultimately, the true, native state of PDHc is still unknown if one considers as *native* only the one that PDHc acquires inside the cell. However, considering the limitations of the methods available to date to probe the in-cell structure of the PDHc metabolon, the integrative methodology applied in this thesis, with its advantages and limitations [56] provides indeed a closer view to the native PDHc biomolecular state; This is further strengthened by the fact that PDHc was fully active within the cell extract that it was imaged.

4.3.2 The eukaryotic, endogenous PDHc structure reveals adaptations of the native state

Most PDHc or its core structures are produced utilising icosahedral symmetry operations[145,147-149,194]. This suggests that parts of the structure that do not follow this type of symmetry will be averaged out or neglected as artificial densities introducing some kind of bias, as recently suggested in the work of *Forsberg et al.* [150]. Current knowledge indicates a "substitution model" for the mammalian core where protein subunit stoichiometries of 40:20 [250] or 48:12 E2p:E3BP [151] are described. Another notion for the core structure, favouring an "addition model", is described for the fungal PDHc with 60:12 E2p:E3BP proteins, localising E3BP in a lower-symmetry reconstruction of the E2p/E3BP core [150]. Additionally, up to now, all the E2p cores resolve an *N-ter* back-folded conformation of E2p, limiting the accessibility of the LD to the E2p active site. It remains to be determined how the LD interacts with the E2p (or any PDHc component) active sites since no interface data exists and only model compounds have been solved within the E2p active site or bound to LD. This is because the recent model of the E2p-LD interaction that the *Kastritis* laboratory put forward [183] comes out from integrative cryo-EM/molecular dynamics simulations data and not from high-resolution cryo-EM maps, and can certainly be improved. The applied symmetry is significant for the recovery of the positions of E1p and E3 relative to the core. If the flexible *N-ter* of the core E2p is folded, then the distance to E1p could

be only 26.6 Å (7 residues), which contradicts the previous knowledge hypothesising for more than three times this distance (up to max. 95 Å E1p-E2p) [148]. Another study suggests approx. 50 Å E1p from the E2p core, the first being close to E3, almost at an "interacting" distance [147].

Even if the E2p-E1p interaction could be feasible, agreeing, *e.g.,* with the *in vitro* E2p/E3BP/E1p complex shown in NS micrographs [149], external densities in the native *C. thermophilum* PDHc are observed at ~150 Å away from the PDHc core **(Figures 45-49, Results Chapter)** [140]. This folded state of the linker might imply an inactive form of native PDHc. On the other hand, an asymmetric reconstruction and placement of the E1p in the resolved density determine the proximity of its active site relative to E2p. Since E3BP localisation was never really addressed, the location of E3 in mammalian PDHc core is hypothetical. E3BP position was suggested in yeast [194] and shown with a partial model in *N. crassa* [150], with the hypothesis of E3 being close to the pentagonal openings of the core. However, in the current study [140] no densities are observed in this location and E3 is proposed to be located further above the pentagonal face. Data from the same reconstruction capture E3BP *C-ter* at a distance of 138.1 Å to its E3 PSBD. Based on the linker length (Leu215-Ala271), the linker is compacted, and the E3 is reasonably placed.

4.3.3 A justifiable model of a localised pyruvate oxidation reaction chamber dubbed as pyruvate dehydrogenase factory

The architecture of the pyruvate dehydrogenase factory is proposed by the 3D models, their placements in the cryo-EM map and the approximate distances between them. This overall architecture defines a low-resolution structural element of the native PDHc that was not previously observed. High molecular weight densities are revealed on the periphery of the PDHc core, proposing an attractive model that provides further mechanistic insights. As a result, the same enzymes that otherwise would participate as members of the reaction now spatially orient forming a reaction chamber that restricts the movement of the reacting enzymes. In such a manner, they define the relative proximity of involved binding sites and form a confined exclusion zone that can accommodate the flexible movement of the LD. The measured summation of minimum distances among the three active sites in this study proposes a distance of ~140 Å – that being the trafficking distance of the LD to eventually perform a full reaction.

Two main mechanistic models have been proposed for PDHc function, namely the "multiple random coupling" (MRC) [251] and the "division-of-labour" (DOL) [149,252], with the

pyruvate dehydrogenase factory proposed here being in favour of the latter and in agreement with a non-homogeneous distribution of different PDHc components. However, the pyruvate dehydrogenase factory mechanism does not agree with spatially organised and distinct E2p-E1p and E3BP-E3 clusters, but rather points to a more sophisticated clustering behaviour of E1p and E3 towards the PDHc core where E1p and E3 interdependently separate for some to be involved in the enzymatic reaction, and others to confine it spatially. Structures similar to the dehydrogenase factory might exist, perhaps for the mammalian PDHc or the remaining OADH complexes, but these are just assumptions, and further investigation is needed to prove its existence beyond being observed in *C. thermophilum* PDHc. However, the dehydrogenase factory architecture is not unique. Recently, similar reaction patterns have been revealed with lower resolution methods, such as a molecular reaction chamber responsible for modifications of chromatin utilising modern methods in light microscopy [253]. Overall, it is interesting to hypothesise that transient MDa reaction chambers could be additionally responsible for defining pathway/reaction regulation.

The overall spatial organisation of E1p, E2p, E3BP and E3 in the pyruvate dehydrogenase factory is intriguing, but many questions remain. For example, if localisation and determination of the dynamic lipoyl arms and a higher amount of E1p and E3 proteins is possible and if there are additional components of these MDa machines like shown in **Figure 51** for PDHc. In addition, it would be interesting to know if there is an influence in the PDHc ultrastructure due to (de-)phosphorylation and interaction with its various kinases and phosphatases. Furthermore, considering the higher-order structural complexity of OADHc metabolons, are higher resolution insights into native architectural and functional relationships of the 2-oxo-acid dehydrogenase complex family feasible at all?

Improved resolution in the fractionation of cell extracts combined by high-density crosslinking MS, utilising stable biomolecular assemblies, kinetic assays for all subunits and/or pure components, large-scale image acquisitions, and efficient asymmetric reconstruction methods could answer many of the remaining questions. In the long run, integrative biology of native cell extracts can shed light on the extensive complexity of the giant machines involved in cellular respiration. This pathway has fascinated biochemists since *Hans Krebs* [254] and continues doing so due to the involvement of those α-keto acid dehydrogenase complexes in numerous cellular pathways and diseases.

5. Summary

Nowadays, cryo-electron microscopy (cryo-EM) unambiguously represents a rising star among the methods available to understand the architecture of protein networks and is becoming a valuable system of methodologies for the scientific community. Until the last decade, cryo-electron microscopy was employed to understand the molecular organisation of purified to homogeneity or *in situ* biomolecules with significant limitations in achieved resolution. Due to the recent advances in the field, structural characterisation of large, flexible and heterogeneous complexes came into reach. A new era of studying the molecular organisation of native cell extracts tries to bridge this gap. This new approach holds a great promise in the more profound understanding of molecular signatures of ordered biological systems at high-resolutions – the molecules of life. The structurally resolved signature of the fungal fatty acid synthase (FAS) at 4.7 Å in 2017, directly from eukaryotic native cell extracts, stands among the best examples in this field, being the first one reaching close-to-atomic resolution of a highly complex biological sample. Surprisingly, higher-order binders, previously completely unknown, were observed in interaction with FAS and, even though FAS was lowly enriched (less than 50 %), finally was resolved to this resolution. Although a lysis step could introduce artifacts, cell extracts retain various functional aspects. In the case of cell extracts, in contrast to extensively purified or artificially over-expressed proteins of interest, all these higher-order biomolecular interactions, dubbed as protein communities, have higher chances of remaining in their (near-) native state. Such an approach greatly complements traditional molecular biology methods and goes hand-in-hand with the holistic understanding of molecular function.

In this thesis, a great example of further establishing, optimising, systematising and combining such an approach was described. The strength of native cell extracts was demonstrated by utilising them to explore megadalton (MDa) complexes involved in pyruvate oxidation, a core metabolic reaction and intrinsic part of cellular respiration. The oxidative decarboxylation of pyruvate involves one of the most critical and key playing complexes, the giant enzymatic assembly known as Pyruvate Dehydrogenase Complex (PDHc). PDHc consists of three core complexes E1, E2/E3BP and E3 in multiple copies, converting pyruvate into acetyl-CoA, providing CO_2 and NADH (H^+). Acetyl-CoA may then be used in the citric acid cycle to carry out cellular respiration,

thereby linking glycolysis to the citric acid cycle. Due to its heterogeneity, relatively enormous size and plasticity, the quaternary structure of PDHc has never been addressed in detail before, despite the fact that its components have been characterised in isolation. In the current work, fully assembled *Chaetomium thermophilum* α-keto acid dehydrogenase complexes in native cell extracts were identified. Their domain arrangements were characterised utilising proteomics, activity assays, crosslinking, electron microscopy and computational modelling. A cryo-EM structure of the PDHc core was reported and many unique and previously unknown features were revealed. Besides visualised folding and binding events -unique for the endogenous complex- an asymmetric reconstruction of a 10-MDa PDHc was illustrated. The spatial proximity of its components was described and, along with the proposed stoichiometric data (60 E2p:12 E3BP:~20 E1p:≤12 E3), suggested a minimum reaction path among component enzymes. Ultimately, the calculated trafficking distance for the lipoyl from the E1p, via E2p, to E3 is ~140 Å. In such a way, a dynamic pyruvate oxidation nanocompartment organised by core and peripheral protein species of PDHc was defined. A further understanding of PDHc and α-keto acid dehydrogenase complex structure and function can be achieved within this context. The strength of the entire approach/methodology of studying huge active complexes, dubbed as metabolons, directly from native cell extracts was proved in practice.

To conclude, the fact that we move towards automation in imaging cell extracts and their intrinsic organisation, combined with the integration of molecular/cell biology approaches, gives many hopes for remarkable achievements of pronounced biotechnological and medical importance. In this direction, distinct machine learning approaches can also be applied to discover protein-protein interactions within cell extracts, augmented by experimental data. Integration of machine learning, cryo-EM and complementary structural proteomic approaches can unambiguously provide the basis for a multiscale molecular description of protein communities within native cell extracts. Eventually, such steps tackle with the biophysical characterisation of cellular function in an integrative and holistic manner.

5.1 Zusammenfassung

Heutzutage ist die Kryo-Elektronenmikroskopie (Kryo-EM) ein aufsteigender Stern unter den Methoden der Wahl, um die Architektur der Prozesse des Lebens zu verstehen und wurde über die letzten Jahre zu einer bedeutenden Methode für die wissenschaftliche Gemeinschaft.

Bis zum letzten Jahrzehnt wurde die Kryo-Elektronenmikroskopie eingesetzt, um die molekulare Organisation von bis zur Homogenität gereinigten oder in-situ-exprimierten Proteinen zu verstehen, wobei die erreichte Auflösung stark eingeschränkt war. Dank der jüngsten Fortschritte auf diesem Gebiet ist die strukturelle Charakterisierung großer, flexibler und heterogener Komplexe in Reichweite gekommen. Eine neue Ära, in der die molekulare Organisation nativer Zellextrakte untersucht wird, versucht diese Lücke zu schließen und diese beiden Ansätze miteinander zu verknüpfen. Diese neue Ära verspricht ein tieferes Verständnis der molekularen Signaturen von geordneten biologischen Systemen mit hoher Auflösung - den Molekülen des Lebens. Die strukturell aufgelöste Signatur der Fettsäure-Synthase (FAS), von Hefe in einer Auflösung von 4.7 Å, die 2017, also vor gerade einmal fünf Jahren, direkt aus nativen eukaryotischen Zellextrakten gewonnen wurde. Diese gehört zu den besten Beispielen auf diesem Gebiet, da sie die erste war, die eine nahezu atomare Auflösung aus einer hochkomplexen biologischen Probe erreichte. Interessanterweise wurden Interaktionen höherer Ordnung, die zuvor völlig unbekannt waren, bei der Wechselwirkung mit FAS beobachtet. Hervorzuheben ist, dass diese Auflösung trotz einer geringeren Anreicherung—weniger als 50 %—, in der Probe, erreicht wurde. Obwohl ein Lyse-Schritt Artefakte verursachen kann, bleiben viele Aspekte der Zellfunktionen in Zellextrakten erhalten. Im Gegensatz dazu haben all diese biomolekularen Interaktionen höherer Ordnung, die als Protein-Gemeinschaften bezeichnet werden, eine größere Chance, in ihrem (fast) nativen Zustand zu bleiben, als extensiv gereinigte oder künstlich über-exprimierte Proteine von Interesse. Ein solcher Ansatz ergänzt in hohem Maße die traditionellen molekularbiologischen Methoden und geht Hand in Hand mit dem ganzheitlichen Verständnis der molekularen Funktion.

Ein großartiges Beispiel für die weitere Etablierung, Optimierung, Systematisierung und Kombination der oben genannten Ansätze wurde in den vorherigen Kapiteln beschrieben. Hier wurde die Stärke von Zellextrakten zur Erforschung von Megadalton-Komplexen (MDa), die an der Pyruvat-Oxidation, einer zentralen Stoffwechselreaktion und einem wesentlichen Bestandteil der Zellatmung, beteiligt sind, gezeigt. An der oxidativen Decarboxylierung von Pyruvat ist einer der kritischsten und wichtigsten Komplexe beteiligt, welcher als Pyruvat-Dehydrogenase-Komplex (PDHc) bekannt ist. PDHc ist ein riesiger Enzymkomplex aus drei Kernkomplexen E1, E2/E3BP und E3 in multiplen Kopien, die Pyruvat in Acetyl-CoA umwandeln. Acetyl-CoA kann dann im Krebszyklus, im Zuge der Zellatmung, weiterverwendet werden, so dass dieser Komplex die Glykolyse mit dem Krebszyklus verbindet. Die Pyruvat-Decarboxylierung wird auch als "Pyruvat-Dehydrogenase-Reaktion" bezeichnet, da sie auch die Oxidation von Pyruvat beinhaltet. Die PDHc ist aufgrund ihrer Heterogenität, ihrer relativ enormen Größe sowie ihrer Plastizität ein sehr komplexes System, dessen Bestandteile zwar isoliert charakterisiert wurden, jedoch dessen Quartärstruktur nie im Detail untersucht wurde. In den vorangegangenen Kapiteln wurden vollständige Chaetomium thermophilum α-Ketosäure-Dehydrogenase-Komplexe in nativen Zellextrakten identifiziert. Ihre Domänenanordnungen wurden mit Hilfe von Massenspektrometrie, Aktivitätsuntersuchungen, Crosslinking, Elektronenmikroskopie und computergestützter Strukturmodellierung charakterisiert. Es wurde über eine Kryo-EM-Struktur des PDHc-Kerns berichtet, die viele einzigartige und bisher unbekannte Merkmale offenbarte. Interessanterweise zeigte eine

asymmetrische Rekonstruktion einer 10-MDa-PDHc die räumliche Nähe ihrer Komponenten und legt zusammen mit den vorgeschlagenen stöchiometrischen Daten (60 E2p:12 E3BP:~20 E1p:≤12 E3) einen minimalen Reaktionsweg zwischen den Komponentenenzymen nahe. Auf diese Weise wurde ein dynamisches Pyruvat-Oxidationskompartiment definiert, das durch Kern- und periphere Proteinspezies der PDHc organisiert wird. In diesem Zusammenhang konnte ein tieferes Verständnis der Struktur, sowie Funktion des PDHc- und α-Ketosäure-Dehydrogenase-Komplexes erreicht werden. Des Weiteren wird die Stärke des gesamten Ansatzes/der Methodik zur Untersuchung großer aktiver Komplexe, welche als Metabolone bezeichnet werden, direkt aus nativen Zellextrakten in der Praxis bewiesen.

Andere Ansätze wie moderne proteomische Methoden, Crosslinking-experimente, Netzwerkbiologie und biophysikalische Modellierung können nun auf Zellextrakte angewandt werden und bieten neue Einblicke in diese Art von Studien. Durch die Automatisierung der Bildgebung von Zellextrakten mit ihrer intrinsischen Organisation, lässt in Verbindung mit der Integration von molekular- und zellbiologischen Ansätzen, auf bemerkenswerte Erfolge bei der Untersuchung lebensnaher biomolekularer Zustände von ausgeprägter biotechnologischer und medizinischer Bedeutung hoffen.

In dieser Richtung können auch verschiedene Ansätze des maschinellen Lernens angewandt werden, um Protein-Protein-Interaktionen in Zellextrakten zu entdecken. Dedizierte biologische Netzwerke in denen Mitglieder der Proteingemeinschaften unterschiedlicher Organismen rekonstruiert und identifiziert werden. Die Validierung ist nach wie vor von großer Bedeutung, z.B. bei der Crosslinking-Massenspektrometrie oder bei zellbiologischen Methoden. Die Anwendung von Bildverarbeitungsabläufen, die von Techniken des maschinellen Lernens inspiriert sind, können bei der Unterscheidung von strukturellen Signaturen helfen. Die Korrelation von Proteom- und Netzwerkdaten mit diesen Signaturen wird die Rekonstruktion von Kryo-EM-Karten verbessern und gleichzeitig helfen, unentdeckte und völlig unbekannte Proteingemeinschaften mit hoher Auflösung zu charakterisieren. Die Integration von maschinellem Lernen, Kryo-EM und ergänzenden strukturellen proteomischen Ansätzen kann eindeutig die Grundlage für eine ganzheitliche molekulare Beschreibung von Proteingemeinschaften in nativen Zellextrakten bilden. Letztendlich nähern sich solche Schritte der biophysikalischen Charakterisierung von Zellfunktionen auf integrative und ganzheitliche Weise.

5.2 Περίληψη

Σήμερα η κρυο-ηλεκτρονική μικροσκοπία (κρυο-ΗΜ) αποτελεί αναμφισβήτητα μια ανερχόμενη μέθοδο επιλογής στην κατανόηση της αρχιτεκτονικής των μοριακών συμπλόκων, των αλληλεπιδράσεων τους και των διαφορετικών στερεοδιαμορφώσεών τους σε σχεδόν ατομικό επίπεδο ανάλυσης. Μέχρι πρόσφατα, η κρυο-ηλεκτρονική μικροσκοπία χρησιμοποιούταν για την κατανόηση της μοριακής οργάνωσης βιομορίων που είχαν απομονωθεί ομοιογενώς ή παρατηρηθεί *in situ*. Παρά ταύτα, τα εμπόδια προς την επίτευξη υψηλής ανάλυσης και συνεπώς λεπτομέρειας παρέμεναν. Πρόσφατα ο δομικός χαρακτηρισμός μεγάλου μεγέθους ετερογενών συμπλόκων έγινε πιο προσιτός εξαιτίας των εξελίξεων στον τομέα της ηλεκτρονικής μικροσκοπίας. Μια νέα εποχή όπου μελετάται η μοριακή οργάνωση των κυτταρικών εκχυλισμάτων προσπαθεί να γεφυρώσει το χάσμα των δύο προαναφερθέντων δομικών προσεγγίσεων: Ενώ οι μέθοδοι *in situ* παρατηρούν τα βιομόρια στο περιβάλλον τους, δεν είναι ικανές να τα ταυτοποιήσουν με ακρίβεια, και οι μέθοδοι ομοιογενούς απομόνωσης των βιομορίων περιγράφουν επακριβώς τα συστατικά τους, αλλά εκτός του κυτταρικού περιβάλλοντος. Η νέα αυτή εποχή, λοιπόν, της δομικής μελέτης των κυτταρικών εκχυλισμάτων, υπόσχεται τη βαθύτερη κατανόηση των «μοριακών υπογραφών» των βιολογικών συστημάτων -των μορίων της ζωής- σε υψηλή ανάλυση. Το 2017, η επίλυση της δομής της συνθετάσης των λιπαρών οξέων (FAS) στα 4.7 Å χρησιμοποιώντας ως οργανισμό μοντέλο ένα μύκητα, απευθείας από ευκαρυωτικά κυτταρικά εκχυλίσματα, συγκαταλέγεται μεταξύ των καλύτερων παραδειγμάτων σε αυτόν τον τομέα. Αποτελεί ουσιαστικά την πρώτη δομή που έφτασε σε σχεδόν ατομική ανάλυση από ένα εξαιρετικά πολύπλοκο βιολογικό δείγμα. Παραδόξως, παρατηρήθηκαν αλληλεπιδράσεις υψηλότερες από αυτή της τριτοταγούς δομικής οργάνωσης με την FAS, που ήταν προηγουμένως εντελώς άγνωστες. Στο παραπάνω δείγμα η FAS ήταν εμπλουτισμένη σε χαμηλά ποσοστά (λιγότερο από 50 %), όμως επιλύθηκε σε υψηλή ανάλυση. Παρά το γεγονός ότι κατά το στάδιο λύσης των κυττάρων για την παραγωγή των κυτταρικών εκχυλισμάτων θα μπορούσαν να δημιουργηθούν «artifacts», τα κυτταρικά εκχυλίσματα διατηρούν πολλές πτυχές και πληροφορίες της κυτταρικής λειτουργίας. Οι «πρωτεϊνικές κοινότητες» και όλες αυτές οι αλληλεπιδράσεις υψηλότερης τάξης, έχουν μεγαλύτερες πιθανότητες να διατηρηθούν στην (σχεδόν) ενδοκυττάρια τους κατάσταση από ό,τι οι εκτενώς και ομοιογενώς απομονωμένες ή τεχνητά υπερεκφρασμένες μορφές τους. Μια τέτοια προσέγγιση συμπληρώνει σε μεγάλο βαθμό τις παραδοσιακές μεθόδους μοριακής βιολογίας και συμβάλλει προς την κατεύθυνση της περαιτέρω κατανόηση της μοριακής λειτουργίας.

Ένα εξαιρετικό παράδειγμα επιπλέον καθιέρωσης, βελτιστοποίησης, συστηματοποίησης και συνδυασμού των παραπάνω προσεγγίσεων περιγράφεται στην παρούσα διδακτορική διατριβή. Στην παρούσα εργασία αποκαλύπτονται οι δυνατότητές των κυτταρικών εκχυλισμάτων σε σχέση με τη μελέτη μεγάλων συμπλόκων (της τάξης των MDa) που εμπλέκονται στην οξείδωση του πυροσταφυλικού. Η διαδικασία αυτή αποτελεί μια βασική μεταβολική αντίδραση και σημαντικό κομμάτι της κυτταρικής αναπνοής, στην οποία εμπλέκεται ένα γιγαντιαίο ενζυμικό σύμπλοκο γνωστό ως αφυδρογονάση του πυροσταφυλικού (PDHc). Η PDHc είναι ένα σύμπλοκο τριών βασικών λειτουργικών υπο-μονάδων (πρωτεϊνών) , των E1, E2/E3BP και E3 ενζύμων, σε πολλαπλά αντίγραφα το καθένα, που ως σκοπό έχουν τη μετατροπή του πυροσταφυλικού σε ακετυλο-συνένζυμο A (Acetyl-CoA) παρέχοντας διοξείδιο του άνθρακα (CO_2) και NADH (H^+). Το ακετυλο-συνένζυμο A μπορεί στη συνέχεια να χρησιμοποιηθεί στον κύκλο του κιτρικού οξέος για τη διεξαγωγή της κυτταρικής αναπνοής και έτσι η PDHc συνδέει τη γλυκόλυση με τον κύκλο του κιτρικού οξέος. Όσων αφορούν στον δομικό χαρακτηρισμό της PDHc, λόγω της ετερογένειάς της, του σχετικά μεγάλου μεγέθους

της, της πλαστικότητάς της, και μολονότι τα συστατικά της έχουν χαρακτηριστεί μεμονωμένα, η τεταρτοταγής δομή του πλήρους συμπλόκου δεν είχε ποτέ εξεταστεί λεπτομερώς.

Στην παρούσα διδακτορική διατριβή, μελετήθηκαν διάφορα σύμπλοκα α-κετοξικών αφυδρογονασών απευθείας από κυτταρικά εκχυλίσματα, με τη χρήση ενός θερμόφιλου οργανισμού μοντέλου που ονομάζεται *Chaetomium thermophilum*. Η οργάνωση των υποσυμπλόκων τους μελετήθηκε χρησιμοποιώντας φασματοφωτομετρία μάζας, δοκιμασίες ενζυμικής ενεργότητας, ηλεκτρονική μικροσκοπία και υπολογιστικά μοντέλα. Η δομή του πυρήνα της PDHc με τη χρήση κρυο-ΗΜ παρουσιάστηκε σε αυτή τη διδακτορική διατριβή και αποκαλύφθηκαν πολλά μοναδικά και άγνωστα μέχρι πρότινος χαρακτηριστικά της. Ενδιαφέρον παρουσιάζει το γεγονός ότι η δομική ανασύσταση της PDHc (συνολικού μεγέθους 10-MDa), χωρίς τη χρήση συμμετρίας, αποκάλυψε τη χωρική εγγύτητα των συστατικών της. Έτσι σε συνδυασμό με τα προτεινόμενα στοιχειομετρικά δεδομένα (60 E2p:12 E3BP:~20 E1p:≤12 E3), παρουσιάστηκε η ελάχιστη απαιτούμενη διαδρομή για την ολοκλήρωση ενός κύκλου αντίδρασης μεταξύ των συστατικών ενζύμων. Με αυτόν τον τρόπο, ορίζεται ένα δυναμικό υπο-διαμέρισμα που οργανώνεται από τις πρωτεΐνες του πυρήνα και των πρωτεϊνών περιφερικά του πυρήνα του συμπλόκου της PDH, όπου λαμβάνει χώρα η οξείδωση του πυροσταφυλικού. Μέσω αυτής της προσέγγισης μπορεί να κατανοηθεί πιο αποτελεσματικά η δομή και η λειτουργία του συμπλόκου της PDH και των α-κετοξικών αφυδρογονασών. Έτσι αποδεικνύεται στην πράξη η δυναμική της προσέγγισης/μεθοδολογίας αυτής στη μελέτη τεράστιων ενεργών συμπλόκων, που ονομάζονται «μεταβολώνια», απευθείας από κυτταρικά εκχυλίσματα.

Το γεγονός ότι οδεύουμε προς μια εποχή όπου η αυτοματοποίηση συμβάλλει όλο και περισσότερο στην απεικόνιση των κυτταρικών εκχυλισμάτων και της ενδοκυττάριας οργάνωσής τους, σε συνδυασμό με εκσυγχρονισμένες προσεγγίσεις μοριακής/κυτταρικής βιολογίας, δίνει πολλές ελπίδες για αξιοσημείωτα επιτεύγματα. Μέσω της μελέτης πιο ρεαλιστικών βιομοριακών δεδομένων προερχόμενων από κυτταρικά εκχυλίσματα, οδηγούμαστε σε ευρήματα με μεγάλη βιοτεχνολογική και ιατρική σημασία. Επίσης, ιδιαίτερη βαρύτητα θα πρέπει να δοθεί στη συμβολή της τεχνητής νοημοσύνης στη μελέτη πρωτεϊνικών αλληλεπιδράσεων εντός κυτταρικών εκχυλισμάτων. Πολύπλοκα δίκτυα βιολογικών δεδομένων μπορούν να ανακατασκευαστούν και να συγκριθούν μετέπειτα με μέλη πρωτεϊνικών κοινοτήτων διαφόρων οργανισμών. Η επαλήθευση τέτοιων δεδομένων παραμένει εξαιρετικά σημαντική, όπως για παράδειγμα, με φασματοφωτομετρία μάζας ή με μεθόδους κυτταρικής βιολογίας. Επιπλέον η επεξεργασία εικόνων με τη χρήση τεχνητής νοημοσύνης μπορεί να βοηθήσει στην αποκάλυψη νέων δομικών χαρακτηριστικών. Με τέτοια βήματα προσεγγίζεται με έναν συνδυαστικό και ολιστικό τρόπο ο βιοφυσικός χαρακτηρισμός των διαφόρων κυτταρικών λειτουργιών.

6. References

1 Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666, doi:10.1038/181662a0 (1958).

2 Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77, doi:10.1038/nature09750 (2011).

3 Kaptein, R., Zuiderweg, E. R., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J Mol Biol* **182**, 179-182, doi:10.1016/0022-2836(85)90036-1 (1985).

4 Williamson, M. P., Havel, T. F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J Mol Biol* **182**, 295-315, doi:10.1016/0022-2836(85)90347-x (1985).

5 Kline, A. D., Braun, W. & Wuthrich, K. Determination of the complete three-dimensional structure of the alpha-amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. *J Mol Biol* **204**, 675-724, doi:10.1016/0022-2836(88)90364-6 (1988).

6 Castellani, F. *et al.* Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* **420**, 98-102, doi:10.1038/nature01070 (2002).

7 Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* **213**, 899-929, doi:10.1016/S0022-2836(05)80271-2 (1990).

8 Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *Elife* **2**, e01345, doi:10.7554/eLife.01345 (2013).

9 Zhang, X. *et al.* Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proc Natl Acad Sci U S A* **105**, 1867-1872, doi:10.1073/pnas.0711623105 (2008).

10 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, doi:10.7554/eLife.42166 (2018).

11 Schur, F. K. *et al.* An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-508, doi:10.1126/science.aaf9620 (2016).

12 Turonova, B., Schur, F. K. M., Wan, W. & Briggs, J. A. G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A. *J Struct Biol* **199**, 187-195, doi:10.1016/j.jsb.2017.07.007 (2017).

13 Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat Methods* **15**, 955-961, doi:10.1038/s41592-018-0167-z (2018).

14 Levitt, M. The birth of computational structural biology. *Nat Struct Biol* **8**, 392-393, doi:10.1038/87545 (2001).

15 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590, doi:10.1038/267585a0 (1977).

16 Vangunsteren, W. F. & Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol Phys* **34**, 1311-1327, doi:10.1080/00268977700102571 (1977).

17 Levitt, M. & Lifson, S. Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure. *Journal of Molecular Biology* **46**, 269-79, doi:10.1016/0022-2836(69)90421-5 (1969).

18 Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* **103**, 227-249, doi:10.1016/0022-2836(76)90311-9 (1976).

19 Chothia, C. & Janin, J. Principles of protein-protein recognition. *Nature* **256**, 705-708, doi:10.1038/256705a0 (1975).

20 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

21 Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034, doi:10.1101/2021.10.04.463034 (2021).

22 Weingarth, M. & Baldus, M. Solid-state NMR-based approaches for supramolecular structure elucidation. *Acc Chem Res* **46**, 2037-2046, doi:10.1021/ar300316e (2013).

23 Frank, J. Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nat Protoc* **12**, 209-212, doi:10.1038/nprot.2017.004 (2017).

24 Raunser, S. Cryo-EM Revolutionizes the Structure Determination of Biomolecules. *Angew Chem Int Ed Engl* **56**, 16450-16452, doi:10.1002/anie.201710679 (2017).

25 McMullan, G., Faruqi, A. R., Clare, D. & Henderson, R. Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 156-163, doi:10.1016/j.ultramic.2014.08.002 (2014).

26 Carragher, B. *et al.* Leginon: an automated system for acquisition of images from vitreous ice specimens. *J Struct Biol* **132**, 33-45, doi:10.1006/jsbi.2000.4314 (2000).

27 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).

28 Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J Struct Biol* **157**, 47-55, doi:10.1016/j.jsb.2006.07.003 (2007).

29 Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J Mol Biol* **415**, 406-418, doi:10.1016/j.jmb.2011.11.010 (2012).

30 de la Rosa-Trevin, J. M. *et al.* Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J Struct Biol* **184**, 321-328, doi:10.1016/j.jsb.2013.09.015 (2013).

31 Kuhlbrandt, W. The Resolution Revolution. *Science* **343**, 1443-1444, doi:10.1126/science.1251652 (2014).

32 Allegretti, M., Mills, D. J., McMullan, G., Kuhlbrandt, W. & Vonck, J. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *Elife* **3**, e01963, doi:10.7554/eLife.01963 (2014).

33 Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. *Elife* **4**, e06980, doi:10.7554/eLife.06980 (2015).

34 Khoshouei, M., Danev, R., Plitzko, J. M. & Baumeister, W. Revisiting the Structure of Hemoglobin and Myoglobin with Cryo-Electron Microscopy. *Journal of Molecular Biology* **429**, 2611-2618, doi:10.1016/j.jmb.2017.07.004 (2017).

35 Neumann, P., Dickmanns, A. & Ficner, R. Validating Resolution Revolution. *Structure* **26**, 785-795 e784, doi:10.1016/j.str.2018.03.004 (2018).

36 Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205-214, doi:10.1016/j.str.2011.12.014 (2012).

37 van Heel, M. & Schatz, M. Reassessing the Revolution's Resolutions. *bioRxiv*, 224402, doi:10.1101/224402 (2017).

38 Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J Struct Biol* **197**, 191-198, doi:10.1016/j.jsb.2016.06.007 (2017).

39 Kunz, M. & Frangakis, A. S. Three-dimensional CTF correction improves the resolution of electron tomograms. *J Struct Biol* **197**, 114-122, doi:10.1016/j.jsb.2016.06.016 (2017).

40 Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969-972, doi:10.1126/science.aad8857 (2016).

41 Beck, M. & Baumeister, W. Cryo-Electron Tomography: Can it Reveal the Molecular Sociology of Cells in Atomic Detail? *Trends Cell Biol* **26**, 825-837, doi:10.1016/j.tcb.2016.08.006 (2016).

42 Beck, M., Mosalaganti, S. & Kosinski, J. From the resolution revolution to evolution: structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Curr Opin Struct Biol* **52**, 32-40, doi:10.1016/j.sbi.2018.07.012 (2018).

43 von Appen, A. & Beck, M. Structure Determination of the Nuclear Pore Complex with Three-Dimensional Cryo electron Microscopy. *J Mol Biol* **428**, 2001-2010, doi:10.1016/j.jmb.2016.01.004 (2016).

44 Hampoelz, B., Andres-Pons, A., Kastritis, P. & Beck, M. Structure and Assembly of the Nuclear Pore Complex. *Annu Rev Biophys* **48**, 515-536, doi:10.1146/annurev-biophys-052118-115308 (2019).

45 Lin, D. H. & Hoelz, A. The Structure of the Nuclear Pore Complex (An Update). *Annu Rev Biochem* **88**, 725-783, doi:10.1146/annurev-biochem-062917-011901 (2019).

46 Schwartz, T. U. The Structure Inventory of the Nuclear Pore Complex. *J Mol Biol* **428**, 1986-2000, doi:10.1016/j.jmb.2016.03.015 (2016).

47 Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233-1243, doi:10.1016/j.cell.2013.10.055 (2013).

48 Stuwe, T. *et al.* Nuclear pores. Architecture of the nuclear pore complex coat. *Science* **347**, 1148-1152, doi:10.1126/science.aaa4136 (2015).

49 von Appen, A. *et al.* In situ structural analysis of the human nuclear pore complex. *Nature* **526**, 140-143, doi:10.1038/nature15381 (2015).

50 Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363-365, doi:10.1126/science.aaf0643 (2016).

51 Schuller, A. P. *et al.* The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667-671, doi:10.1038/s41586-021-03985-3 (2021).

52 Allegretti, M. *et al.* In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796-800, doi:10.1038/s41586-020-2670-5 (2020).

53 Zimmerli, C. E. *et al.* Nuclear pores dilate and constrict in cellulo. *Science*, eabd9776, doi:10.1126/science.abd9776 (2021).

54 Mosalaganti, S. *et al.* In situ architecture of the algal nuclear pore complex. *Nat Commun* **9**, 2361, doi:10.1038/s41467-018-04739-y (2018).

55 Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**, 161, doi:10.1186/s13059-018-1547-5 (2018).

56 Kyrilis, F. L., Meister, A. & Kastritis, P. L. Integrative biology of native cell extracts: a new era for structural characterization of life processes. *Biol Chem* **400**, 831-846, doi:10.1515/hsz-2018-0445 (2019).

57 Kastritis, P. L. *et al.* Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Mol Syst Biol* **13**, 936, doi:10.15252/msb.20167412 (2017).

58 Kastritis, P. L. & Gavin, A. C. Enzymatic complexes across scales. *Essays Biochem* **62**, 501-514, doi:10.1042/Ebc20180008 (2018).

59 Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147, doi:10.1038/415141a (2002).

60 Wan, C. H. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339-344, doi:10.1038/nature14877 (2015).

61 Rost, H. L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741-748, doi:10.1038/nmeth.3959 (2016).

62 Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem (Palo Alto Calif)* **8**, 61-80, doi:10.1146/annurev-anchem-071114-040325 (2015).

63 Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* **38**, 1225-1237, doi:10.1002/jms.559 (2003).

64 Rappsilber, J. The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *J Struct Biol* **173**, 530-540, doi:10.1016/j.jsb.2010.10.014 (2011).

65 O'Reilly, F. J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* **25**, 1000-1008, doi:10.1038/s41594-018-0147-0 (2018).

66 Sinz, A. Cross-Linking/Mass Spectrometry for Studying Protein Structures and Protein-Protein Interactions: Where Are We Now and Where Should We Go from Here? *Angew Chem Int Ed Engl* **57**, 6390-6396, doi:10.1002/anie.201709559 (2018).

67 Liu, F., Lossl, P., Rabbitts, B. M., Balaban, R. S. & Heck, A. J. R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* **17**, 216-232, doi:10.1074/mcp.RA117.000470 (2018).

68 Chavez, J. D. *et al.* Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. *Cell Syst* **6**, 136-141e5, doi:10.1016/j.cels.2017.10.017 (2018).

69 Liu, F., Rijkers, D. T., Post, H. & Heck, A. J. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat Methods* **12**, 1179-1184, doi:10.1038/nmeth.3603 (2015).

70 Iacobucci, C. *et al.* A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein-protein interactions. *Nature Protocols* **13**, 2864-2889, doi:10.1038/s41596-018-0068-8 (2018).

71 Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068-1081, doi:10.1016/j.cell.2012.08.011 (2012).

72 Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631-636, doi:10.1038/nature04532 (2006).

73 Hogeboom, G. H., Schneider, W. C. & Pallade, G. E. Cytochemical studies of mammalian tissues; isolation of intact mitochondria from rat liver; some biochemical properties of mitochondria and submicroscopic particulate material. *J Biol Chem* **172**, 619-635 (1948).

74 Satge, D. & Benard, J. Carcinogenesis in Down syndrome: what can be learned from trisomy 21? *Semin Cancer Biol* **18**, 365-371, doi:10.1016/j.semcancer.2008.03.020 (2008).

75 Baudhuin, P., Beaufay, H. & De Duve, C. Combined biochemical and morphological study of particulate fractions from rat liver. Analysis of preparations enriched in lysosomes or in particles containing urate oxidase, D-amino acid oxidase, and catalase. *J Cell Biol* **26**, 219-243, doi:10.1083/jcb.26.1.219 (1965).

76 Haselkorn, R., Fernandez-Moran, H., Kieras, F. J. & Van Bruggen, E. F. Electron microscopic and biochemical characterization of fraction I protein. *Science* **150**, 1598-1601, doi:10.1126/science.150.3703.1598 (1965).

77 Oudet, P., Gross-Bellard, M. & Chambon, P. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**, 281-300, doi:10.1016/0092-8674(75)90149-x (1975).

78 Mowbray, J. & Moses, V. The tentative identification in Escherichia coli of a multienzyme complex with glycolytic activity. *Eur J Biochem* **66**, 25-36, doi:10.1111/j.1432-1033.1976.tb10421.x (1976).
79 Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nat Methods* **9**, 907-909, doi:10.1038/nmeth.2131 (2012).
80 Razvi, A. & Scholtz, J. M. Lessons in stability from thermophilic proteins. *Protein Sci* **15**, 1569-1578, doi:10.1110/ps.062130306 (2006).
81 Takagi, J., Strokovich, K., Springer, T. A. & Walz, T. Structure of integrin alpha5beta1 in complex with fibronectin. *EMBO J* **22**, 4607-4615, doi:10.1093/emboj/cdg445 (2003).
82 Verbeke, E. J., Mallam, A. L., Drew, K., Marcotte, E. M. & Taylor, D. W. Classification of Single Particles from Human Cell Extract Reveals Distinct Structures. *Cell Rep* **24**, 259-268 e253, doi:10.1016/j.celrep.2018.06.022 (2018).
83 Yi, X., Verbeke, E. J., Chang, Y., Dickinson, D. J. & Taylor, D. W. Electron microscopy snapshots of single particles from single cells. *Journal of Biological Chemistry* **294**, 1602-1608, doi:10.1074/jbc.RA118.006686 (2019).
84 Sigworth, F. J. A maximum-likelihood approach to single-particle image refinement. *J Struct Biol* **122**, 328-339, doi:10.1006/jsbi.1998.4014 (1998).
85 Luchinat, E. & Banci, L. In-cell NMR: a topical review. *IUCrJ* **4**, 108-118, doi:10.1107/S2052252516020625 (2017).
86 Selenko, P. *et al.* In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat Struct Mol Biol* **15**, 321-329, doi:10.1038/nsmb.1395 (2008).
87 Theillet, F. X. *et al.* Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nat Protoc* **8**, 1416-1432, doi:10.1038/nprot.2013.083 (2013).
88 Inomata, K. *et al.* High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **458**, 106-109, doi:10.1038/nature07839 (2009).
89 Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185-190, doi:10.1038/nature23002 (2017).
90 Gipson, P. *et al.* Direct structural insight into the substrate-shuttling mechanism of yeast fatty acid synthase by electron cryomicroscopy. *Proc Natl Acad Sci U S A* **107**, 9164-9169, doi:10.1073/pnas.0913547107 (2010).
91 Boehringer, D., Ban, N. & Leibundgut, M. 7.5-A cryo-em structure of the mycobacterial fatty acid synthase. *J Mol Biol* **425**, 841-849, doi:10.1016/j.jmb.2012.12.021 (2013).
92 Fischer, M. *et al.* Cryo-EM structure of fatty acid synthase (FAS) from Rhodosporidium toruloides provides insights into the evolutionary development of fungal FAS. *Protein Sci* **24**, 987-995, doi:10.1002/pro.2678 (2015).
93 D'Imprima, E. *et al.* Protein denaturation at the air-water interface and how to prevent it. *Elife* **8**, doi:10.7554/eLife.42747 (2019).
94 Elad, N. *et al.* Structure of Type-I Mycobacterium tuberculosis fatty acid synthase at 3.3 A resolution. *Nat Commun* **9**, 3886, doi:10.1038/s41467-018-06440-6 (2018).
95 Amlacher, S. *et al.* Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* **146**, 277-289, doi:10.1016/j.cell.2011.06.039 (2011).
96 Lin, D. H. *et al.* Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015, doi:10.1126/science.aaf1015 (2016).
97 Kellner, N. *et al.* Developing genetic tools to exploit Chaetomium thermophilum for biochemical analyses of eukaryotic macromolecular assemblies. *Sci Rep* **6**, 20937, doi:10.1038/srep20937 (2016).

98 Deduve, C. & Berthet, J. The Use of Differential Centrifugation in the Study of Tissue Enzymes. *Int Rev Cytol* **3**, 225-275, doi:10.1016/S0074-7696(08)60440-4 (1954).
99 Pettit, F. H. & Reed, L. J. Pyruvate-Dehydrogenase Complex from Bovine Kidney and Heart. *Method Enzymol* **89**, 376-386 (1982).
100 Daum, B., Walter, A., Horst, A., Osiewacz, H. D. & Kuhlbrandt, W. Age-dependent dissociation of ATP synthase dimers and loss of inner-membrane cristae in mitochondria. *P Natl Acad Sci USA* **110**, 15301-15306, doi:10.1073/pnas.1305462110 (2013).
101 O'Reilly, F. J. *et al.* In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554-557, doi:10.1126/science.abb3758 (2020).
102 Nirenberg, M. & Matthaei, J. H. Dependence of Cell-Free Protein Synthesis in E Coli Upon Naturally Occurring or Synthetic Polyribonucleotides. *P Natl Acad Sci USA* **47**, 1588-1602, doi:10.1073/pnas.47.10.1588 (1961).
103 Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **99**, 162-168, doi:10.1016/j.biochi.2013.11.025 (2014).
104 Cupello, S., Richardson, C. & Yan, S. Cell-free Xenopus egg extracts for studying DNA damage response pathways. *Int J Dev Biol* **60**, 229-236, doi:10.1387/ijdb.160113sy (2016).
105 Hoogenboom, W. S., Douwel, D. K. & Knipscheer, P. Xenopus egg extract: A powerful tool to study genome maintenance mechanisms. *Dev Biol* **428**, 300-309, doi:10.1016/j.ydbio.2017.03.033 (2017).
106 Hodgman, C. E. & Jewett, M. C. Cell-free synthetic biology: Thinking outside the cell. *Metab Eng* **14**, 261-269, doi:10.1016/j.ymben.2011.09.002 (2012).
107 Calhoun, K. A. & Swartz, J. R. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol Bioeng* **90**, 606-613, doi:10.1002/bit.20449 (2005).
108 Noren, C. J., Anthonycahill, S. J., Griffith, M. C. & Schultz, P. G. A General-Method for Site-Specific Incorporation of Unnatural Amino-Acids into Proteins. *Science* **244**, 182-188, doi:10.1126/science.2649980 (1989).
109 Kigawa, T. *et al.* Cell-free production and stable-isotope labeling of milligram quantities of proteins. *Febs Lett* **442**, 15-19, doi:10.1016/S0014-5793(98)01620-2 (1999).
110 Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem* **16**, 2420-2431, doi:10.1002/cbic.201500340 (2015).
111 Zhu, Z., Kin Tam, T., Sun, F., You, C. & Percival Zhang, Y. H. A high-energy-density sugar biobattery based on a synthetic enzymatic pathway. *Nat Commun* **5**, 3026, doi:10.1038/ncomms4026 (2014).
112 Huang, C., Rossi, P., Saio, T. & Kalodimos, C. G. Structural basis for the antifolding activity of a molecular chaperone. *Nature* **537**, 202-206, doi:10.1038/nature18965 (2016).
113 Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc Natl Acad Sci U S A* **111**, 15635-15640, doi:10.1073/pnas.1418377111 (2014).
114 Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun Biol* **2**, 218, doi:10.1038/s42003-019-0437-z (2019).
115 van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol* **428**, 720-725, doi:10.1016/j.jmb.2015.09.014 (2016).
116 Webb, B. *et al.* Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci* **27**, 245-258, doi:10.1002/pro.3311 (2018).

117 Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-921, doi:10.1107/s0907444998003254 (1998).

118 Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-1737, doi:10.1021/ja026939x (2003).

119 van Zundert, G. C. P., Melquiond, A. S. J. & Bonvin, A. Integrative Modeling of Biomolecular Complexes: HADDOCKing with Cryo-Electron Microscopy Data. *Structure* **23**, 949-960, doi:10.1016/j.str.2015.03.014 (2015).

120 Hopf, T. A. *et al.* Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife* **3**, doi:10.7554/eLife.03430 (2014).

121 Karaca, E. & Bonvin, A. M. On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys. *Acta Crystallogr D Biol Crystallogr* **69**, 683-694, doi:10.1107/S0907444913007063 (2013).

122 Karaca, E., Rodrigues, J., Graziadei, A., Bonvin, A. & Carlomagno, T. M3: an integrative framework for structure determination of molecular machines. *Nat Methods* **14**, 897-902, doi:10.1038/nmeth.4392 (2017).

123 Koukos, P. I., Faro, I., van Noort, C. W. & Bonvin, A. M. J. J. A Membrane Protein Complex Docking Benchmark. *Journal of Molecular Biology* **430**, 5246-5256, doi:10.1016/j.jath.2018.11.005 (2018).

124 Bonvin, A. M. J. J., Karaca, E., Kastritis, P. L. & Rodrigues, J. P. G. L. M. Defining distance restraints in HADDOCK. *Nature Protocols* **13**, 1503-1503, doi:10.1038/s41596-018-0017-6 (2018).

125 Orban-Nemeth, Z. *et al.* Structural prediction of protein models using distance restraints derived from cross-linking mass spectrometry data. *Nature Protocols* **13**, 1724-1724, doi:10.1038/s41596-018-0024-7 (2018).

126 Vangone, A. *et al.* Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. *Proteins* **85**, 417-423, doi:10.1002/prot.25198 (2017).

127 Kastritis, P. L. & Bonvin, A. M. J. J. Are Scoring Functions in Protein-Protein Docking Ready To Predict Interactomes? Clues from a Novel Binding Affinity Benchmark. *J Proteome Res* **9**, 2216-2225, doi:10.1021/pr9009854 (2010).

128 Kastritis, P. L., Rodrigues, J. P. G. L. M. & Bonvin, A. M. J. J. HADDOCK(2P2I): A Biophysical Model for Predicting the Binding Affinity of Protein-Protein Interaction Inhibitors. *J Chem Inf Model* **54**, 826-836, doi:10.1021/ci4005332 (2014).

129 Reed, L. J. A trail of research from lipoic acid to alpha-keto acid dehydrogenase complexes. *J Biol Chem* **276**, 38329-38336, doi:10.1074/jbc.R100026200 (2001).

130 Patel, M. S., Nemeria, N. S., Furey, W. & Jordan, F. The pyruvate dehydrogenase complexes: structure-based function and regulation. *J Biol Chem* **289**, 16615-16623, doi:10.1074/jbc.R114.563148 (2014).

131 Perham, R. N. Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. *Biochemistry* **30**, 8501-8512, doi:10.1021/bi00099a001 (1991).

132 Bisswanger, H. Substrate specificity of the pyruvate dehydrogenase complex from Escherichia coli. *J Biol Chem* **256**, 815-822 (1981).

133 Roche, T. E. & Reed, L. J. Function of the nonidentical subunits of mammalian pyruvate dehydrogenase. *Biochem Biophys Res Commun* **48**, 840-846, doi:10.1016/0006-291x(72)90684-5 (1972).

134 Patel, M. S., Roche, T. E. & Harris, R. *Alpha-Keto Acid Dehydrogenase Complexes*. 317 (Springer-Verlag, 1996).

135 Bassendine, M. F., Jones, D. E. & Yeaman, S. J. Biochemistry and autoimmune response to the 2-oxoacid dehydrogenase complexes in primary biliary cirrhosis. *Semin Liver Dis* **17**, 49-60, doi:10.1055/s-2007-1007182 (1997).

136 Liu, Y. *et al.* N (6)-methyladenosine RNA modification-mediated cellular metabolism rewiring inhibits viral replication. *Science* **365**, 1171-1176, doi:10.1126/science.aax4468 (2019).

137 Mkrtchyan, G. *et al.* Molecular mechanisms of the non-coenzyme action of thiamin in brain: biochemical, structural and pathway analysis. *Sci Rep* **5**, 12583, doi:10.1038/srep12583 (2015).

138 Jakkamsetti, V. *et al.* Brain metabolism modulates neuronal excitability in a mouse model of pyruvate dehydrogenase deficiency. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aan0457 (2019).

139 Palmieri, E. M. *et al.* Nitric oxide orchestrates metabolic rewiring in M1 macrophages by targeting aconitase 2 and pyruvate dehydrogenase. *Nat Commun* **11**, 698, doi:10.1038/s41467-020-14433-7 (2020).

140 Kyrilis, F. L. *et al.* Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. *Cell Rep* **34**, 108727, doi:10.1016/j.celrep.2021.108727 (2021).

141 Ciszak, E. M. *et al.* How dihydrolipoamide dehydrogenase-binding protein binds dihydrolipoamide dehydrogenase in the human pyruvate dehydrogenase complex. *J Biol Chem* **281**, 648-655, doi:10.1074/jbc.M507850200 (2006).

142 Klyuyeva, A., Tuganova, A., Kedishvili, N. & Popov, K. M. Tissue-specific kinase expression and activity regulate flux through the pyruvate dehydrogenase complex. *J Biol Chem* **294**, 838-851, doi:10.1074/jbc.RA118.006433 (2019).

143 White, P. J. *et al.* The BCKDH Kinase and Phosphatase Integrate BCAA and Lipid Metabolism via Regulation of ATP-Citrate Lyase. *Cell Metab* **27**, 1281-1293 e1287, doi:10.1016/j.cmet.2018.04.015 (2018).

144 Kato, M. *et al.* Structural basis for inactivation of the human pyruvate dehydrogenase complex by phosphorylation: role of disordered phosphorylation loops. *Structure* **16**, 1849-1859, doi:10.1016/j.str.2008.10.010 (2008).

145 Jiang, J. *et al.* Atomic Structure of the E2 Inner Core of Human Pyruvate Dehydrogenase Complex. *Biochemistry* **57**, 2325-2334, doi:10.1021/acs.biochem.8b00357 (2018).

146 Szabo, E. *et al.* Underlying molecular alterations in human dihydrolipoamide dehydrogenase deficiency revealed by structural analyses of disease-causing enzyme variants. *Hum Mol Genet* **28**, 3339-3354, doi:10.1093/hmg/ddz177 (2019).

147 Yu, X. *et al.* Structures of the human pyruvate dehydrogenase complex cores: a highly conserved catalytic center with flexible N-terminal domains. *Structure* **16**, 104-114, doi:10.1016/j.str.2007.10.024 (2008).

148 Zhou, Z. H., McCarthy, D. B., O'Connor, C. M., Reed, L. J. & Stoops, J. K. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proc Natl Acad Sci U S A* **98**, 14802-14807, doi:10.1073/pnas.011597698 (2001).

149 Prajapati, S. *et al.* Structural and Functional Analyses of the Human PDH Complex Suggest a "Division-of-Labor" Mechanism by Local E1 and E3 Clusters. *Structure* **27**, 1124-1136 e1124, doi:10.1016/j.str.2019.04.009 (2019).

150 Forsberg, B. O., Aibara, S., Howard, R. J., Mortezaei, N. & Lindahl, E. Arrangement and symmetry of the fungal E3BP-containing core of the pyruvate dehydrogenase complex. *Nat Commun* **11**, 4667, doi:10.1038/s41467-020-18401-z (2020).

151 Hiromasa, Y., Fujisawa, T., Aso, Y. & Roche, T. E. Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding

protein and their capacities to bind the E1 and E3 components. *J Biol Chem* **279**, 6921-6933, doi:10.1074/jbc.M308172200 (2004).

152 Brautigam, C. A. *et al.* Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex. *Structure* **14**, 611-621, doi:10.1016/j.str.2006.01.001 (2006).

153 Lengyel, J. S. *et al.* Extended polypeptide linkers establish the spatial architecture of a pyruvate dehydrogenase multienzyme complex. *Structure* **16**, 93-103, doi:10.1016/j.str.2007.10.017 (2008).

154 Cheng, J. *et al.* Thermophile 90S Pre-ribosome Structures Reveal the Reverse Order of Co-transcriptional 18S rRNA Subdomain Integration. *Mol Cell* **75**, 1256-1269 e1257, doi:10.1016/j.molcel.2019.06.032 (2019).

155 Ho, C. M. *et al.* Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat Methods* **17**, 79-85, doi:10.1038/s41592-019-0637-y (2020).

156 Verbeke, E. J. *et al.* Separating distinct structures of multiple macromolecular assemblies from cryo-EM projections. *J Struct Biol* **209**, 107416, doi:10.1016/j.jsb.2019.107416 (2020).

157 McCafferty, C. L., Verbeke, E. J., Marcotte, E. M. & Taylor, D. W. Structural Biology in the Multi-Omics Era. *J Chem Inf Model* **60**, 2424-2429, doi:10.1021/acs.jcim.9b01164 (2020).

158 Tan, Y. Z. *et al.* Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods* **14**, 793-796, doi:10.1038/nmeth.4347 (2017).

159 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

160 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).

161 Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).

162 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* **Chapter 5**, Unit-5 6, doi:10.1002/0471250953.bi0506s15 (2006).

163 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

164 Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443-453, doi:10.1016/0022-2836(70)90057-4 (1970).

165 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).

166 de la Rosa-Trevin, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol* **195**, 93-99, doi:10.1016/j.jsb.2016.04.010 (2016).

167 Pintilie, G. & Chiu, W. Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in cryo-EM density maps. *Biopolymers* **97**, 742-760, doi:10.1002/bip.22074 (2012).

168 Wagner, T. & Raunser, S. The evolution of SPHIRE-crYOLO particle picking and its application in automated cryo-EM processing workflows. *Commun Biol* **3**, 61, doi:10.1038/s42003-020-0790-y (2020).

169 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

170 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).

171 Sorzano, C. O. *et al.* XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148**, 194-204, doi:10.1016/j.jsb.2004.06.006 (2004).

172 *EPU User Manual*. PN 1025707 edn, (ThermoFisher Scientific, 2021).

173 Zemlin, F., Weiss, K., Schiske, P., Kunath, W. & Herrmann, K. H. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy* **3**, 49-60, doi:10.1016/S0304-3991(78)80006-0 (1978).

174 Edington, J. W. *The operation and calibration of the electron microscope*. (MacMillan, 1974).

175 Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* **294**, 5181-5197, doi:10.1074/jbc.REV118.005602 (2019).

176 Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **49**, D480-D489, doi:10.1093/nar/gkaa1100 (2020).

177 Kato, M. *et al.* A synchronized substrate-gating mechanism revealed by cubic-core structure of the bovine branched-chain alpha-ketoacid dehydrogenase complex. *EMBO J* **25**, 5983-5994, doi:10.1038/sj.emboj.7601444 (2006).

178 Seifert, F. *et al.* Phosphorylation of serine 264 impedes active site accessibility in the E1 component of the human pyruvate dehydrogenase multienzyme complex. *Biochemistry* **46**, 6277-6287, doi:10.1021/bi700083z (2007).

179 Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211-215, doi:10.1093/nar/gkn785 (2009).

180 Zimmermann, L. *et al.* A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**, 2237-2243, doi:10.1016/j.jmb.2017.12.007 (2018).

181 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).

182 Kresze, G. B. & Ronft, H. Pyruvate dehydrogenase complex from baker's yeast. 2. Molecular structure, dissociation, and implications for the origin of mitochondria. *Eur J Biochem* **119**, 581-587, doi:10.1111/j.1432-1033.1981.tb05647.x (1981).

183 Tuting, C. *et al.* Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. *Nat Commun* **12**, 6933, doi:10.1038/s41467-021-27287-4 (2021).

184 Mattevi, A. *et al.* Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. *Science* **255**, 1544-1550, doi:10.1126/science.1549782 (1992).

185 Sawle, L. & Ghosh, K. How do thermophilic proteins and proteomes withstand high temperature? *Biophys J* **101**, 217-227, doi:10.1016/j.bpj.2011.05.059 (2011).

186 Ma, B. G., Goncearenco, A. & Berezovsky, I. N. Thermophilic adaptation of protein complexes inferred from proteomic homology modeling. *Structure* **18**, 819-828, doi:10.1016/j.str.2010.04.004 (2010).

187 Mattevi, A., Obmolova, G., Kalk, K. H., Teplyakov, A. & Hol, W. G. Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p). *Biochemistry* **32**, 3887-3901, doi:10.1021/bi00066a007 (1993).

188 Janin, J., Bahadur, R. P. & Chakrabarti, P. Protein-protein interaction and quaternary structure. *Q Rev Biophys* **41**, 133-180, doi:10.1017/S0033583508004708 (2008).

189 Dey, S., Pal, A., Chakrabarti, P. & Janin, J. The subunit interfaces of weakly associated homodimeric proteins. *J Mol Biol* **398**, 146-160, doi:10.1016/j.jmb.2010.02.020 (2010).

190 Behal, R. H., DeBuysere, M. S., Demeler, B., Hansen, J. C. & Olson, M. S. Pyruvate dehydrogenase multienzyme complex. Characterization of assembly intermediates by sedimentation velocity analysis. *J Biol Chem* **269**, 31372-31377 (1994).

191 Kastritis, P. L., Rodrigues, J. P., Folkers, G. E., Boelens, R. & Bonvin, A. M. Proteins feel more than they see: fine-tuning of binding affinity by properties of the non-interacting surface. *J Mol Biol* **426**, 2632-2652, doi:10.1016/j.jmb.2014.04.017 (2014).

192 Vernon, R. M. *et al.* Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**, doi:10.7554/eLife.31486 (2018).

193 Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W. & Huang, R. B. The multiple roles of histidine in protein interactions. *Chem Cent J* **7**, 44, doi:10.1186/1752-153X-7-44 (2013).

194 Stoops, J. K. *et al.* On the unique structural organization of the Saccharomyces cerevisiae pyruvate dehydrogenase complex. *J Biol Chem* **272**, 5757-5764, doi:10.1074/jbc.272.9.5757 (1997).

195 Bahadur, R. P., Chakrabarti, P., Rodier, F. & Janin, J. A dissection of specific and non-specific protein-protein interfaces. *J Mol Biol* **336**, 943-955, doi:10.1016/j.jmb.2003.12.073 (2004).

196 Casasoli, M. *et al.* Integration of evolutionary and desolvation energy analysis identifies functional sites in a plant immunity protein. *Proc Natl Acad Sci U S A* **106**, 7666-7671, doi:10.1073/pnas.0812625106 (2009).

197 Sanderson, S. J., Khan, S. S., McCartney, R. G., Miller, C. & Lindsay, J. G. Reconstitution of mammalian pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase complexes: analysis of protein X involvement and interaction of homologous and heterologous dihydrolipoamide dehydrogenases. *Biochem J* **319 (Pt 1)**, 109-116, doi:10.1042/bj3190109 (1996).

198 Smolle, M. *et al.* A new level of architectural complexity in the human pyruvate dehydrogenase complex. *J Biol Chem* **281**, 19772-19780, doi:10.1074/jbc.M601140200 (2006).

199 Wohlgemuth, I., Lenz, C. & Urlaub, H. Studying macromolecular complex stoichiometries by peptide-based mass spectrometry. *Proteomics* **15**, 862-879, doi:10.1002/pmic.201400466 (2015).

200 Piersimoni, L., Kastritis, P. L., Arlt, C. & Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions horizontal line A Method for All Seasons. *Chem Rev*, doi:10.1021/acs.chemrev.1c00786 (2021).

201 Karim, A. S. & Jewett, M. C. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metab Eng* **36**, 116-126, doi:10.1016/j.ymben.2016.03.002 (2016).

202 Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nat Rev Genet* **21**, 151-170, doi:10.1038/s41576-019-0186-3 (2020).

203 Hammerling, M. J. *et al.* In vitro ribosome synthesis and evolution through ribosome display. *Nat Commun* **11**, 1108, doi:10.1038/s41467-020-14705-2 (2020).

204 Han, B. G. *et al.* Survey of large protein complexes in D. vulgaris reveals great structural diversity. *Proc Natl Acad Sci U S A* **106**, 16580-16585, doi:10.1073/pnas.0813068106 (2009).

205 Su, C. C. *et al.* A 'Build and Retrieve' methodology to simultaneously solve cryo-EM structures of membrane proteins. *Nat Methods* **18**, 69-75, doi:10.1038/s41592-020-01021-2 (2021).

206 Arimura, Y., Shih, R. M., Froom, R. & Funabiki, H. Nucleosome structural variations in interphase and metaphase chromosomes. *bioRxiv*, 2020.2011.2012.380386, doi:10.1101/2020.11.12.380386 (2020).

207 Skalidis, I., Tuting, C. & Kastritis, P. L. Unstructured regions of large enzymatic complexes control the availability of metabolites with signaling functions. *Cell Commun Signal* **18**, 136, doi:10.1186/s12964-020-00631-9 (2020).

208 Kyrilis, F. L., Belapure, J. & Kastritis, P. L. Detecting Protein Communities in Native Cell Extracts by Machine Learning: A Structural Biologist's Perspective. *Front Mol Biosci* **8**, 660542, doi:10.3389/fmolb.2021.660542 (2021).
209 Titeca, K., Lemmens, I., Tavernier, J. & Eyckerman, S. Discovering cellular protein-protein interactions: Technological strategies and opportunities. *Mass Spectrom Rev* **38**, 79-111, doi:10.1002/mas.21574 (2019).
210 Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol Syst Biol* **7**, 549, doi:10.1038/msb.2011.82 (2011).
211 Ferriero, R. *et al.* Pyruvate dehydrogenase complex and lactate dehydrogenase are targets for therapy of acute liver failure. *Journal of Hepatology* **69**, 325-335, doi:10.1016/j.jhep.2018.03.016 (2018).
212 Sutendra, G. *et al.* A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation. *Cell* **158**, 84-97, doi:10.1016/j.cell.2014.04.046 (2014).
213 Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. Protein-protein interaction detection: methods and analysis. *Int J Proteomics* **2014**, 147648, doi:10.1155/2014/147648 (2014).
214 Salas, D., Stacey, R. G., Akinlaja, M. & Foster, L. J. Next-generation Interactomics: Considerations for the Use of Co-elution to Measure Protein Interaction Networks. *Mol Cell Proteomics* **19**, 1-10, doi:10.1074/mcp.R119.001803 (2020).
215 Barnes, S. J. & Weitzman, P. D. Organization of citric acid cycle enzymes into a multienzyme cluster. *Febs Lett* **201**, 267-270, doi:10.1016/0014-5793(86)80621-4 (1986).
216 Sindelar, C. V. & Grigorieff, N. An adaptation of the Wiener filter suitable for analyzing images of isolated single particles. *J Struct Biol* **176**, 60-74, doi:10.1016/j.jsb.2011.06.010 (2011).
217 Xie, J., Xu, L. & Chen, E. in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1* 341–349 (Curran Associates Inc., Lake Tahoe, Nevada, 2012).
218 Jain, V. & Seung, H. S. in *Proceedings of the 21st International Conference on Neural Information Processing Systems* 769–776 (Curran Associates Inc., Vancouver, British Columbia, Canada, 2008).
219 Bepler, T., Kelley, K., Noble, A. J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat Commun* **11**, 5208, doi:10.1038/s41467-020-18952-1 (2020).
220 Huang, Z. & Penczek, P. A. Application of template matching technique to particle detection in electron micrographs. *J Struct Biol* **145**, 29-40, doi:10.1016/j.jsb.2003.11.004 (2004).
221 Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S. & Carragher, B. DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166**, 205-213, doi:10.1016/j.jsb.2009.01.004 (2009).
222 Shatsky, M., Hall, R. J., Brenner, S. E. & Glaeser, R. M. A method for the alignment of heterogeneous macromolecules from electron microscopy. *J Struct Biol* **166**, 67-78, doi:10.1016/j.jsb.2008.12.008 (2009).
223 Wang, F. *et al.* DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. *J Struct Biol* **195**, 325-336, doi:10.1016/j.jsb.2016.07.006 (2016).
224 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

225 Zhu, Y., Ouyang, Q. & Mao, Y. A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy. *BMC Bioinformatics* **18**, 348, doi:10.1186/s12859-017-1757-y (2017).

226 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Res Comput Mol Biol* **10812**, 245-247 (2018).

227 Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152, doi:10.1038/s41592-019-0580-y (2019).

228 Zhang, J. *et al.* PIXER: an automated particle-selection method based on segmentation using a deep neural network. *BMC Bioinformatics* **20**, 41, doi:10.1186/s12859-019-2614-y (2019).

229 Sanchez-Garcia, R., Segura, J., Maluenda, D., Sorzano, C. O. S. & Carazo, J. M. MicrographCleaner: A python package for cryo-EM micrograph cleaning using deep learning. *J Struct Biol* **210**, 107498, doi:10.1016/j.jsb.2020.107498 (2020).

230 Rawat, W. & Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput* **29**, 2352-2449, doi:10.1162/NECO_a_00990 (2017).

231 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185, doi:10.1038/s41592-020-01049-4 (2021).

232 Sanchez-Garcia, R. *et al.* DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *bioRxiv*, 2020.2006.2012.148296, doi:10.1101/2020.06.12.148296 (2020).

233 Kimanius, D. *et al.* Exploiting prior knowledge about biological macromolecules in cryo-EM structure determination. *IUCrJ* **8**, 60-75, doi:10.1107/S2052252520014384 (2021).

234 Ramirez-Aportela, E., Mota, J., Conesa, P., Carazo, J. M. & Sorzano, C. O. S. DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. *IUCrJ* **6**, 1054-1063, doi:10.1107/S2052252519011692 (2019).

235 Russel, D. *et al.* Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* **10**, e1001244, doi:10.1371/journal.pbio.1001244 (2012).

236 Saha, M. & Morais, M. C. FOLD-EM: automated fold recognition in medium- and low-resolution (4-15 A) electron density maps. *Bioinformatics* **28**, 3265-3273, doi:10.1093/bioinformatics/bts616 (2012).

237 Skalidis, I. *et al.* Cryo-EM and artificial intelligence visualize endogenous protein community members. *Structure*, doi:10.1016/j.str.2022.01.001 (2022).

238 Mostosi, P., Schindelin, H., Kollmannsberger, P. & Thorn, A. Haruspex: A Neural Network for the Automatic Identification of Oligonucleotides and Protein Secondary Structure in Cryo-Electron Microscopy Maps. *Angew Chem Int Ed Engl*, doi:10.1002/anie.202000421 (2020).

239 Xu, M., Beck, M. & Alber, F. Template-free detection of macromolecular complexes in cryo electron tomograms. *Bioinformatics* **27**, i69-76, doi:10.1093/bioinformatics/btr207 (2011).

240 Xu, M. *et al.* De Novo Structural Pattern Mining in Cellular Electron Cryotomograms. *Structure* **27**, 679-691 e614, doi:10.1016/j.str.2019.01.005 (2019).

241 Zhou, B. *et al.* One-Shot Learning With Attention-Guided Segmentation in Cryo-Electron Tomography. *Front Mol Biosci* **7**, 613347, doi:10.3389/fmolb.2020.613347 (2020).

242 Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J. M. & Forster, F. Fast and accurate reference-free alignment of subtomograms. *J Struct Biol* **182**, 235-245, doi:10.1016/j.jsb.2013.03.002 (2013).

243 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).
244 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450, doi:10.1093/nar/gky1106 (2019).
245 Giurgiu, M. *et al.* CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res* **47**, D559-D563, doi:10.1093/nar/gky973 (2019).
246 Lawson, C. L. *et al.* EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* **39**, D456-464, doi:10.1093/nar/gkq880 (2011).
247 Iudin, A., Korir, P. K., Salavert-Torres, J., Kleywegt, G. J. & Patwardhan, A. EMPIAR: a public archive for raw electron microscopy image data. *Nat Methods* **13**, 387-388, doi:10.1038/nmeth.3806 (2016).
248 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242, doi:10.1093/nar/28.1.235 (2000).
249 Srere, P. A. The infrastructure of the mitochondrial matrix. *Trends in Biochemical Sciences* **5**, 120-121 (1980).
250 Brautigam, C. A., Wynn, R. M., Chuang, J. L. & Chuang, D. T. Subunit and catalytic component stoichiometries of an in vitro reconstituted human pyruvate dehydrogenase complex. *J Biol Chem* **284**, 13086-13098, doi:10.1074/jbc.M806563200 (2009).
251 Hackert, M. L., Oliver, R. M. & Reed, L. J. Evidence for a multiple random coupling mechanism in the alpha-ketoglutarate dehydrogenase multienzyme complex of Escherichia coli: a computer model analysis. *Proc Natl Acad Sci U S A* **80**, 2226-2230, doi:10.1073/pnas.80.8.2226 (1983).
252 Milne, J. L. *et al.* Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy. *J Biol Chem* **281**, 4364-4370, doi:10.1074/jbc.M504363200 (2006).
253 Gilbert, N. & van Leeuwen, F. Chromatin modified in a molecular reaction chamber. *Nature* **579**, 503-504, doi:10.1038/d41586-020-00638-9 (2020).
254 Krebs, H. A. & Johnson, W. A. The role of citric acid in intermediate metabolism in animal tissues. *Enzymologia* **4**, 148-156 (1937).

7. Appendix

7.1 Supplementary figures

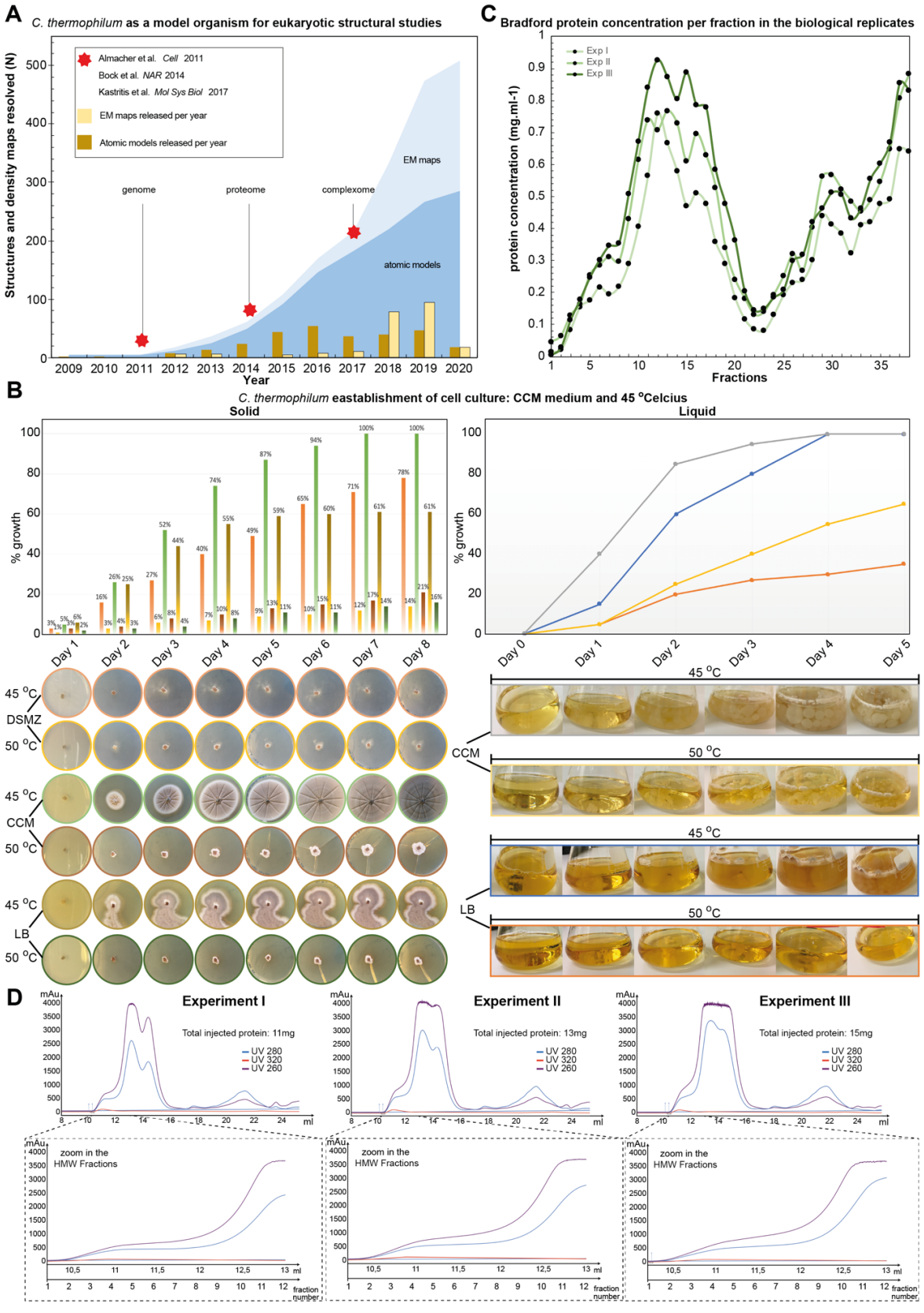


Figure S1. *C. thermophilum* **as a model system for cell extract fractionation.**
A. Yearly cumulative increase in deposited EM electron optical density maps (light blue) and atomic models deposited in PDB (dark blue) from *C. thermophilum*. Bar plots indicate EM maps published per year (light brown) and atomic structures published per year (dark brown). Stars represent publications communicating the genome, proteome and interactome of *C. thermophilum*. **B.** Establishment of *C. thermophilum* culture at 45^{o}C. On the left, solid culture is shown; on the right, liquid culture is shown; Plots show relative growth as function of day. Below the plots, Images acquired per day of *C. thermophilum* in different media is shown and a visual comparison of their growth. **C.** Determination of protein concentration in the fractions after size exclusion chromatography; The 3 different biological replicates are shown, also reflecting the relative increase in protein concentration as function of injected protein amount of native extract. **D.** Chromatography profiles of the 3 experiments and their UV absorbance at 260, 280 and 320 nm wavelengths; Below, zoom ins into the high molecular weight fractions and the amount of aggregation, as monitored by 320 nm. (Source: Figure reproduced from [140])

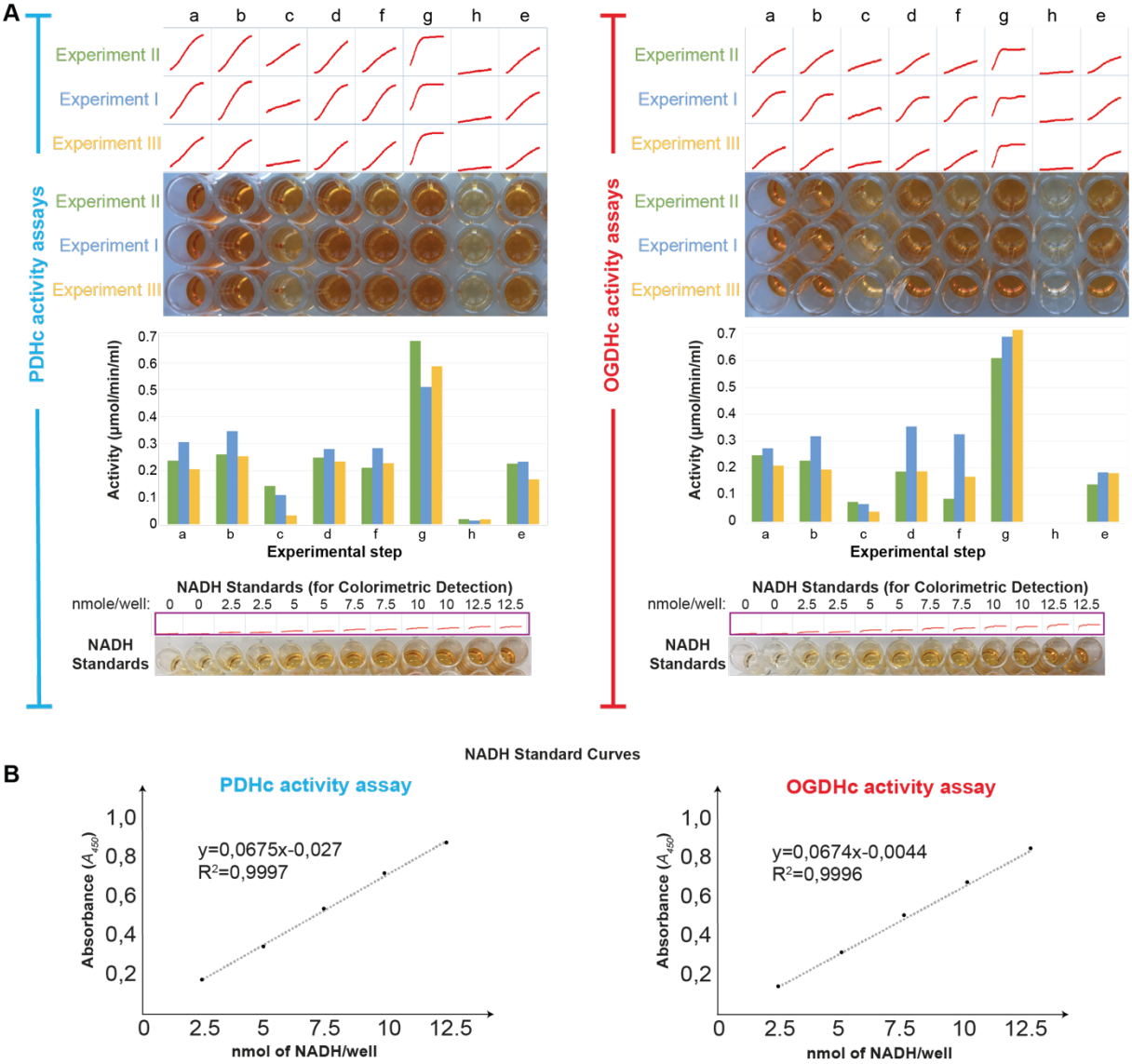


Figure S2. Activity assays of PDHc and OGDHc during the steps of the biochemical preparation.

A. On the left, for the PDHc activity assays, original spectrophotometric curves are shown for all steps of the experiment shown in Figure 11 (a-e) in biological replicates and the colouration of the wells, showing successful reactions; Below the image of the wells, quantification and analysis of the signal correspond to various activities of E1p at different steps of the biochemical preparation is shown. A similar illustration is depicted on the right for the kinetic assays performed concerning E1o. **B.** The linear NADH standard curves are shown, used in the PDHc and OGDHc activity analysis. (Source: Figure reproduced from [140])

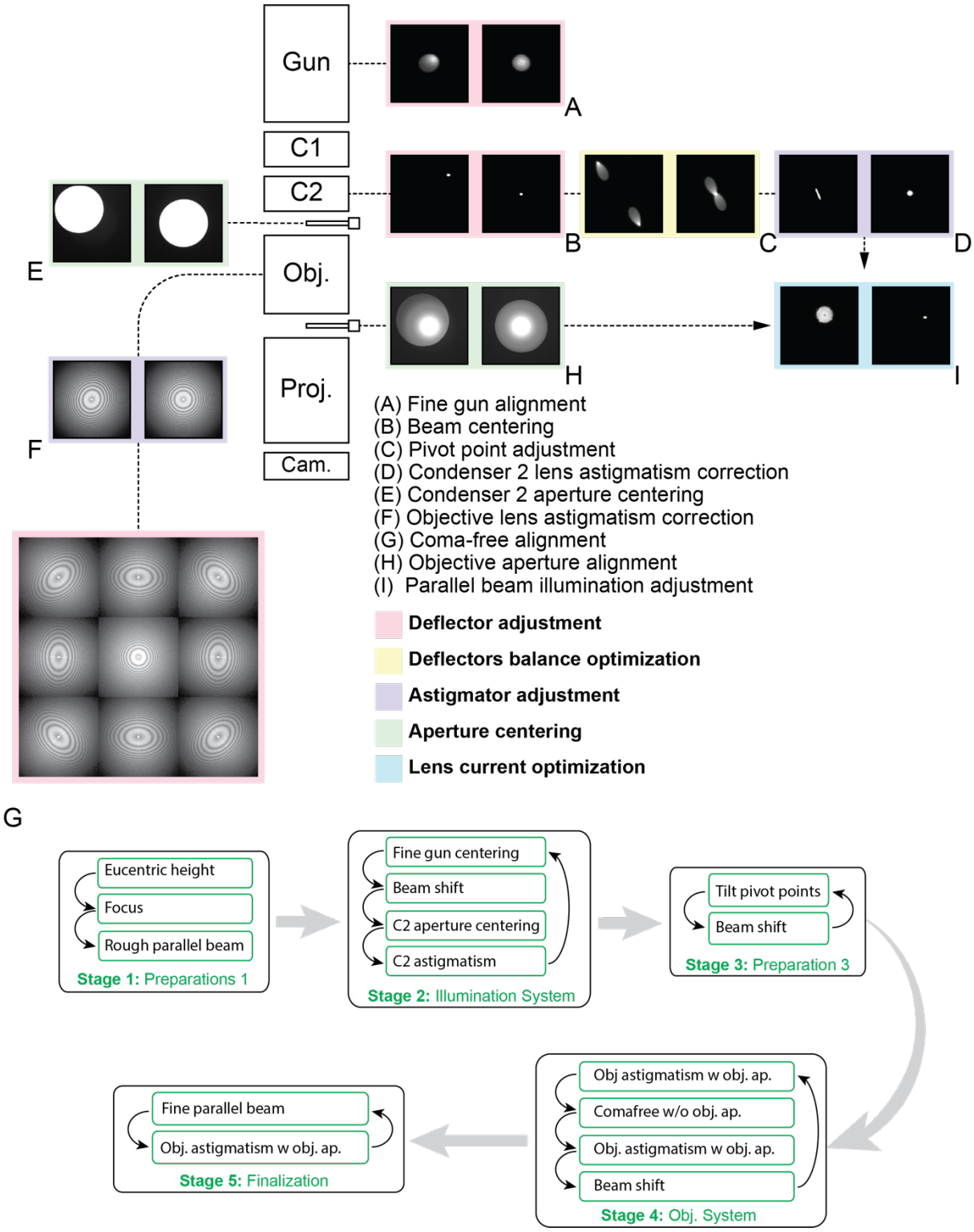


Figure S3. Schematic representation of the workflow and all basic steps required to ensure a centered, coherent and symmetric beam around the microscope's optical axis. (Courtesy of Farzad Hamdi, MLU Halle-Wittenberg)

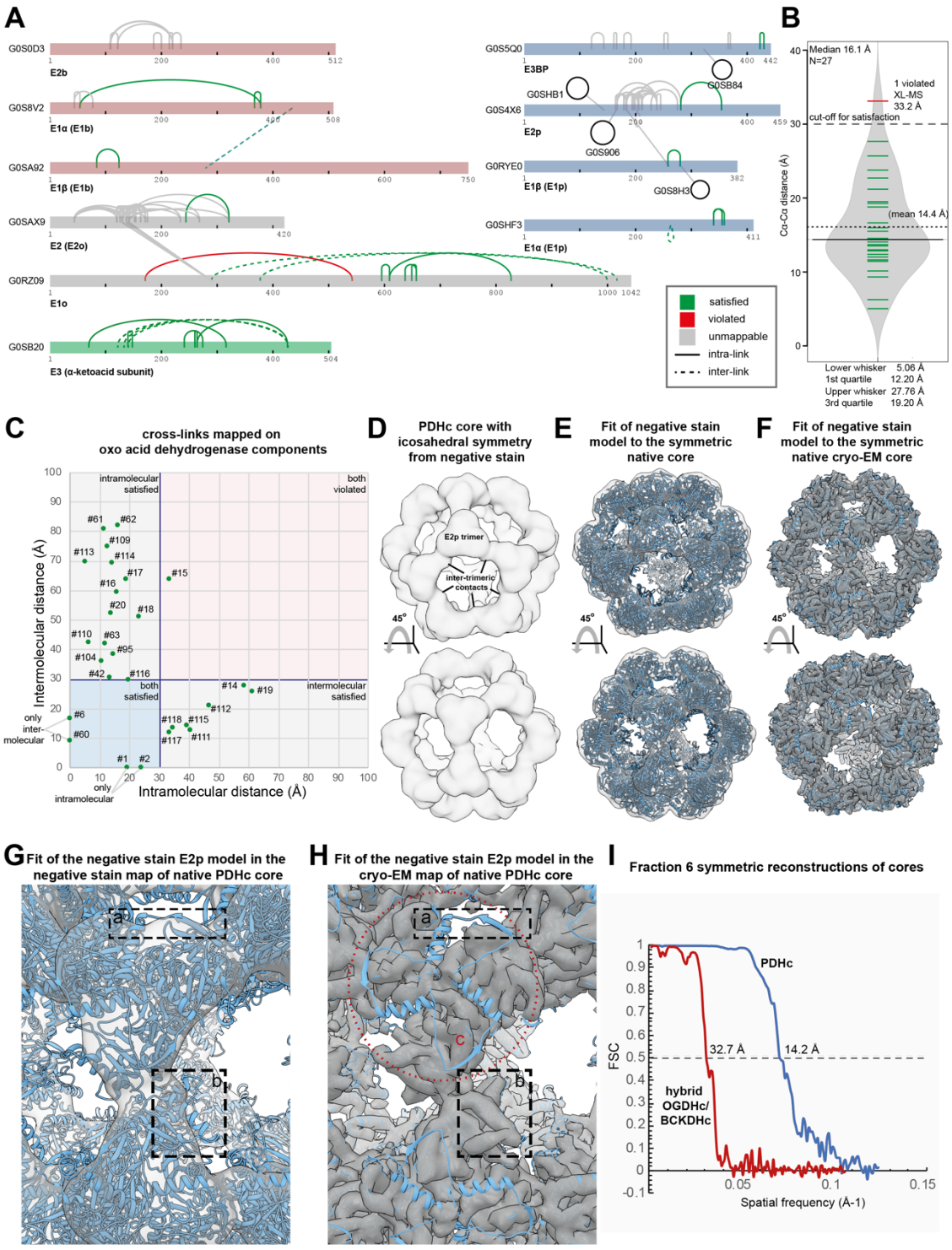


Figure S4. Analysis of crosslinks from previously communicated data [57] for OADH complexes and reconstructed core density of native PDHc core with the fit of derived molecular models.

A. mapping of crosslinks on top of α-keto acid dehydrogenase complex protein sequences; (grey: unmappable crosslinks; green: mappable crosslinks, satisfied; red: mappable crosslink, violated). **B.** Density distribution plot of all crosslinks mapping on top of the predicted of α-keto acid dehydrogenase complex protein models. **C.** Mapping of inter- and intra- molecular distances of derived crosslinks and satisfaction in either but not both of the states. Satisfied crosslinks can be either intra-molecular or inter-molecular, but results show that none is satisfied in both states: 9 correspond to satisfied inter-XL (18.4±7.3 Å), while 17 to intra-XL (14.5±5.2 Å). **D.** Density from negative stain showing features of vertices and faces, as well as the trimer shape, along with distinct densities for the inter-trimeric contacts of the E2 proteins.

E. Predicted E2 model based on the negative stain density and **F.** fit of the negative stain E2 atomic model in the cryo-EM map of the E2 core. **G.** Zoom into (E), where the fit of the negative stain atomic model in the negative stain map of the E2 core is shown. Box (a) shows missing coverage of densities for the corresponding helix-loop-strand element (see text) and box (b) shows corresponding missing density for the helices connecting the two trimers. **H.** Zoom into (F), where, clearly, densities of (a) are still missing, while densities for the helices present in the inter-trimer interface are recovered (b); (c) The N-terminal sequence of the E2 trimer is not covered by the native cryo-EM map of the PDHc core and therefore, this region is natively flexible. **I.** FSC plots for the symmetrized cores of PDHc and hybrid OGDHc/BCKDHc. (Source: Figure reproduced from [140])

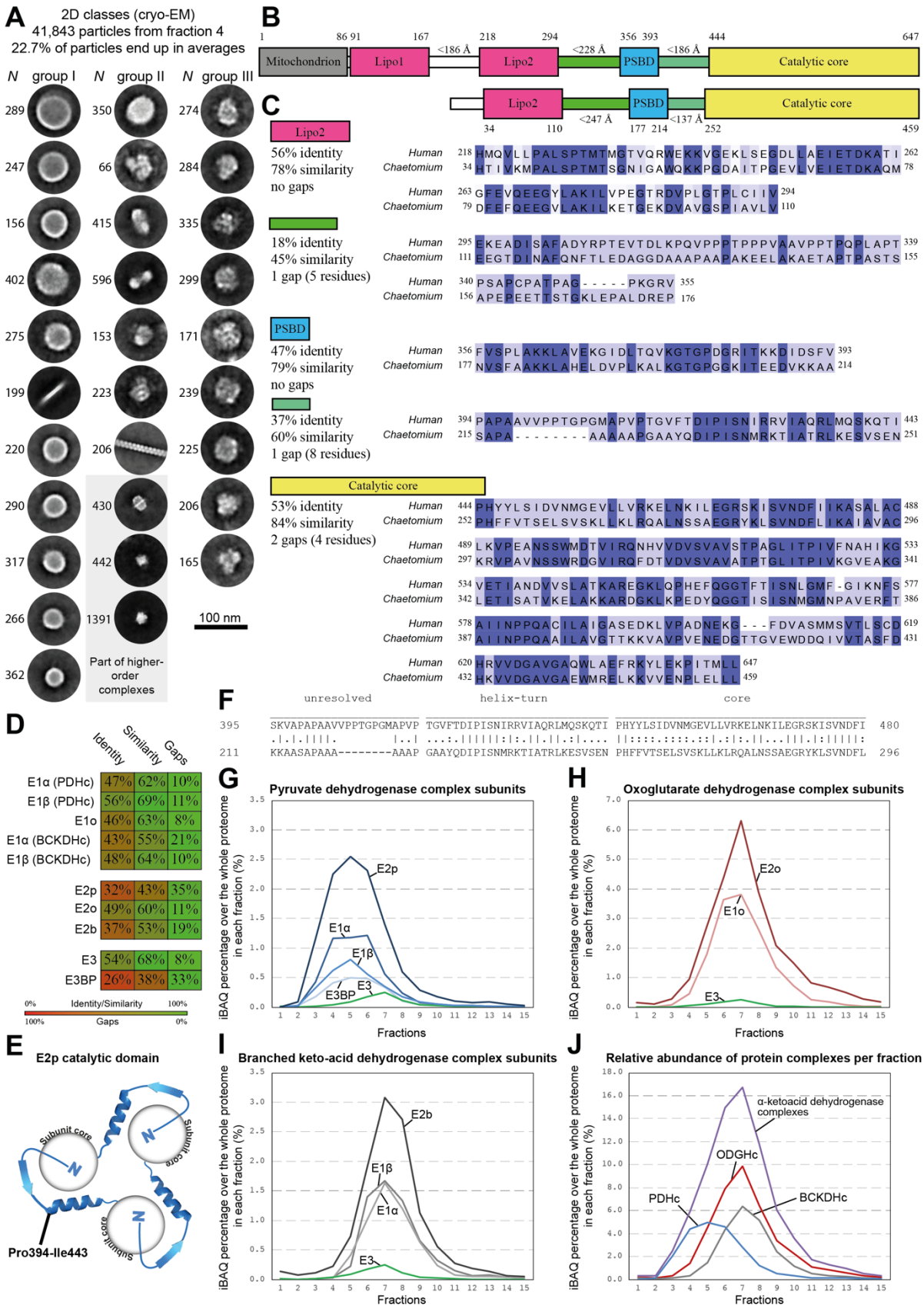


Figure S5. Fraction 4 cryo-EM class averages and α-keto acid dehydrogenase complex, their sequence characteristics and mass spectrometry elution profiles of component proteins. **A.** 2D class averages retrieved from cryo-EM data collected from fraction 4, where large biomolecular complexes are observed. 3 Groups of class averages are shown (Groups I-III) along with the number of particles in each class (*N*). Group I include membrane-containing particles, Group

II other complexes of striking morphology and Group III possible PDHc class averages. **B.** Comparison of human vs *C. thermophilum* E2 polypeptide sequence, highlighting the different domains; Lipo1 and Lipo2: Lipoyl domains; PSBD: peripheral subunit binding domain (binding the E1); Mitochondrion is the signal peptide for mitochondrial localization. **C.** Sequence alignment of different domains of the E2 protein from human and *C. thermophilum*. **D.** Percentage of sequence similarity, identity and gaps among calculated alignments of *C. thermophilum* and human α-keto acid dehydrogenase complex proteins. **E.** Sequence alignment of the peripheral sequence of the E2 catalytic domain and conservation of the sequence between human and *C. thermophilum*. **F.** Organization of the trimeric catalytic domain of the E2 protein, where the subunit catalytic core is peripherally surrounded by the structural element described in (E), that of a loop-helix-turn-loop domain. **G-J.** Co-elution of protein species identified by mass spectrometry [57] for all subunits of PDHc (G), OGDHc (H), BCKDHc (I) and their complexes (J), after summation of their respective protein abundances. Relative abundance is reported per fraction, which is the abundance of each of the proteins measured by their iBAQ score, over all identified proteins in each of the fractions (**Table S3** for details). (Source: Figure reproduced from [140])

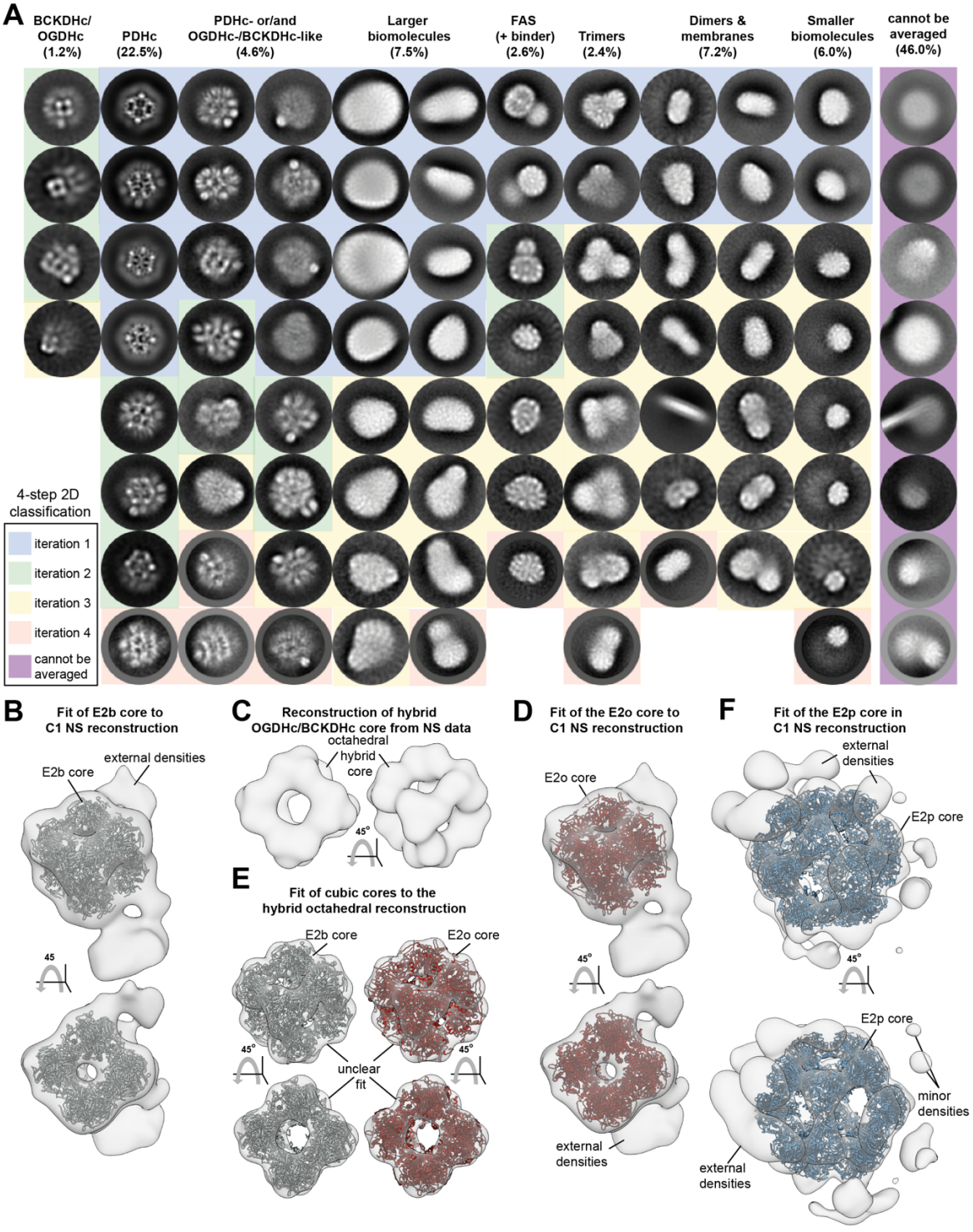


Figure S6. 2D class averages from negative staining data present in fraction 6 and fitting of derived *C. thermophilum* E2 models of α-keto acid dehydrogenase complexes into 3D negative staining reconstructions. **A.** 2D class averages from the large-scale acquisition of fraction 6. Four iterations of 2D classification are shown and clear structural signatures are presented in each iteration; 46.0 % of single-particles could not be averaged after the iterative classification. **B.** Fit of BCKDHc E2 core in the reconstructed C1 density of fraction 6; **C.** the cubic density of the hybrid OGDHc/BCKDHc reconstruction; **D.** Fit of OGDHc E2 core in the reconstructed C1 density of fraction 6; **E.** Fit of cubic cores to the hybrid octahedral reconstruction and satisfaction of both higher-order assemblies; **F.** Fit of PDHc E2 core in the reconstructed symmetrized density of fraction 6; Apparent densities for external subunits is highlighted. (Source: Figure reproduced from [140])

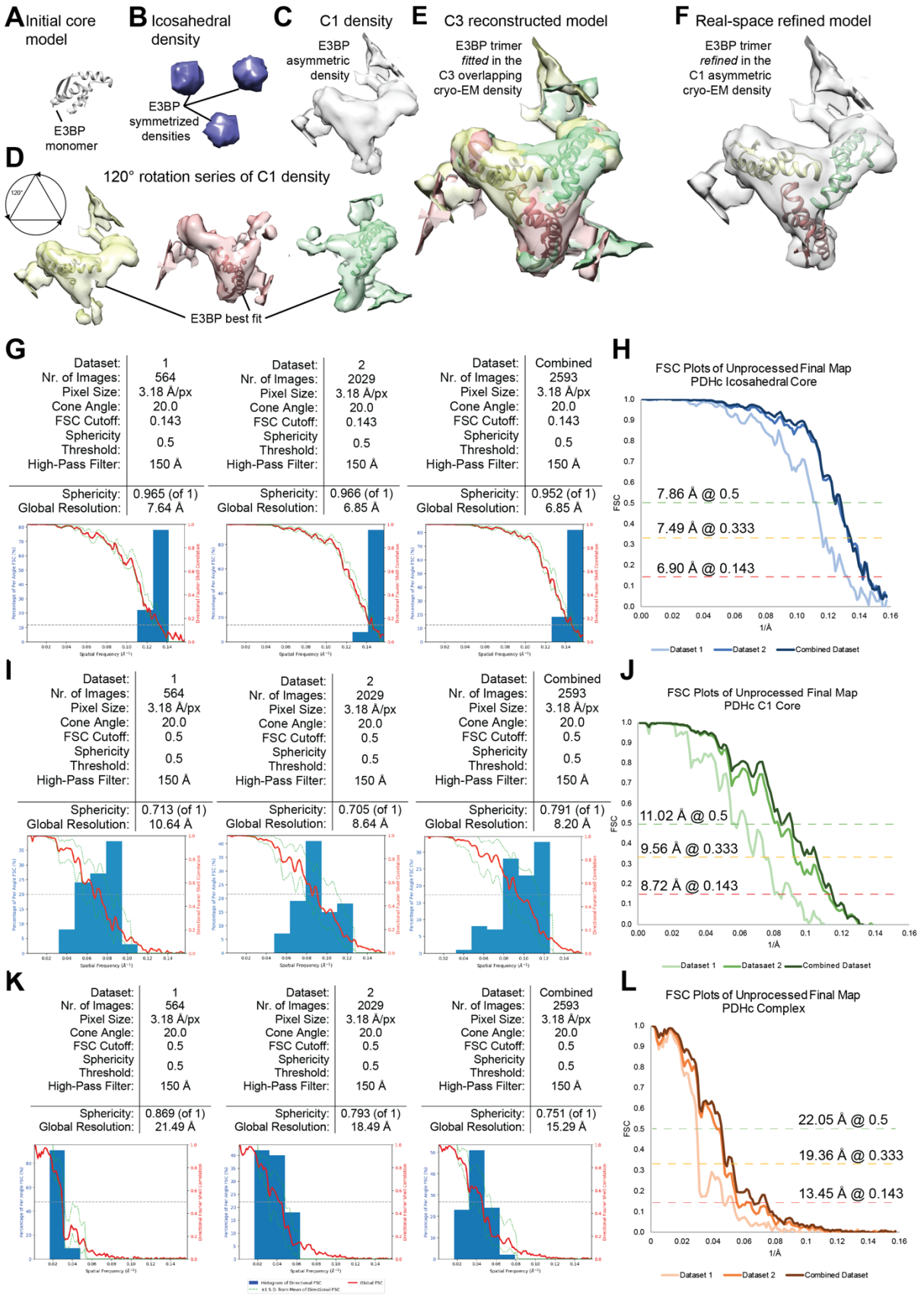


Figure S7. Method for modeling, fitting and reconstructing the E3BP trimer in the cryo-EM map and presentation of associated validation metrics for cryo-EM reconstructions reported in this work. **A.** The derived molecular model of the C-ter of ctE3BP based on structural homology. **B.** The densities inside the symmetrized native ctPDHc core are symmetrized; **C.** In the PDHc core asymmetric reconstruction (8.7 Å, FSC-0.143), a clear density appears exhibiting C3

symmetry. **D-E.** Top fit of the E3BP model in the extracted inner density; C3 rotations of E3BP model superimpose well in the densities, recapitulating the experimentally-derived curvature. **F.** Real-space refinement of the model in the map generates a clash-free C3 symmetric dimer with a stable, novel interface. **G.** 3DFSC statistics of the symmetrized PDHc cores, reported sphericity values and derived FSC plots (see materials and methods). **H.** FSC plot calculated for the symmetrized PDHc core, corresponding to the different datasets. **I.** Same as panel (G), but for the asymmetric reconstruction of the PDHc core. **J.** Same as (H), but for the asymmetric reconstruction of the PDHc core. **K.** same as panel (G), but for the asymmetric reconstruction of the full PDHc. **L.** same as (H), but for the asymmetric reconstruction of the full PDHc and reported FSC is 0.5 (22.05 Å). Reported resolutions through the manuscript are those retrieved from RELION 3.0 [10]. (Source: Figure reproduced from [140])

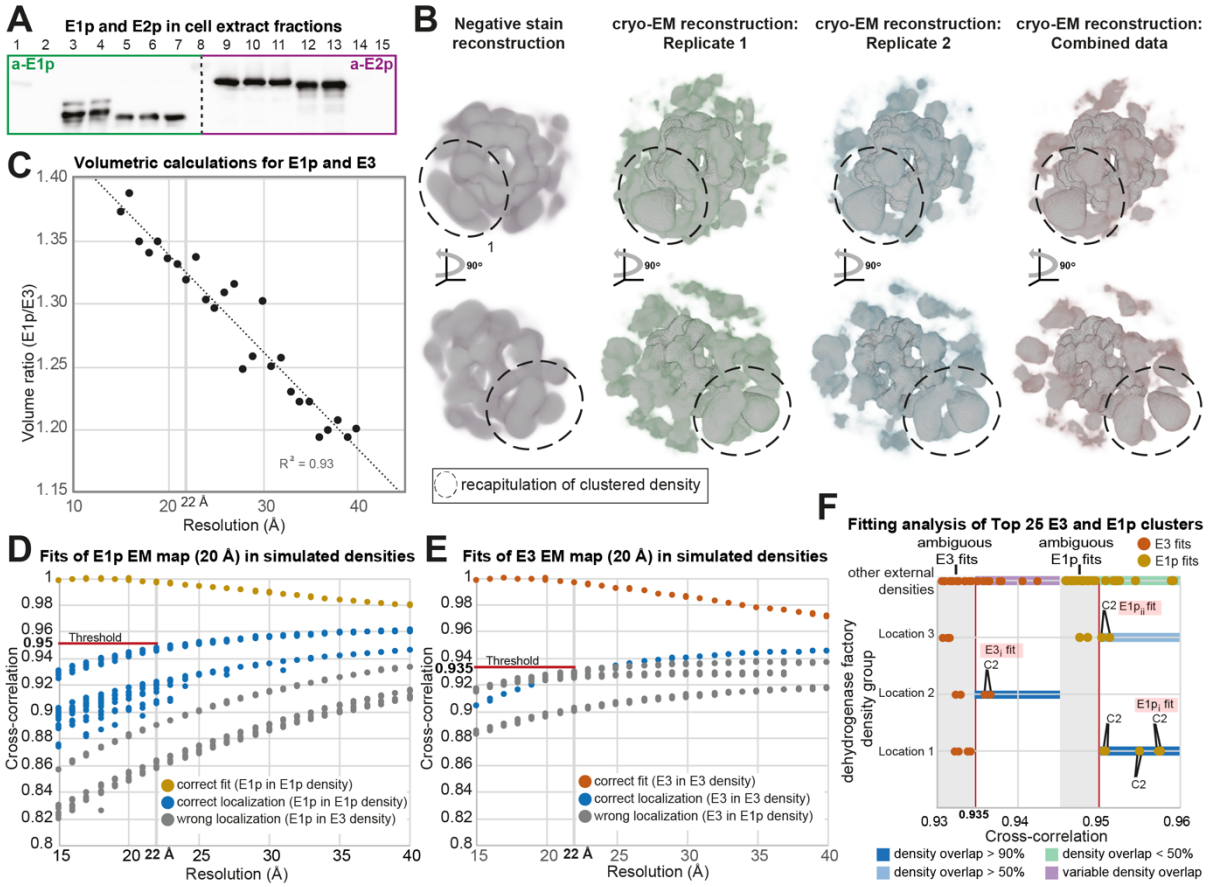


Figure S8. Structure-based analysis for the dehydrogenase factory model and single-particles of PDHc with higher-order binders, forming protein communities. **A.** Cross-reactivity of Abs against E1p and E2 proteins and constructs is not detected, allowing efficient semi-quantification of band intensity for stoichiometry analysis. Analysis of the immunodetection experiment estimates ~20 copies of E1p for each PDHc complex (see text for details). Lanes: 1: 20.0 ng E2p-His-Tag, 2: Fr 29 (NC), 3-4: 20.0 ng and 15.0 ng E1α-His-Tag, respectively, 5-7: Fr 6 (2 µL), 8: Marker, 9-11: Fr 6 (2 µL), 12-13: 15.0 ng and 20.0 ng E2p-His-Tag, respectively, 14: Fr 29 (Negative control, NC (2 µL), E2p and E1α are not present as shown by MS data), 15: 20.0 ng E1α-His-Tag. **B.** Validation using different reconstructions of the asymmetric complex of PDHc; from left to right: reconstruction from negative stain data, the first and the second cryo-dataset from different biological replicates and the final merged dataset; The density of the prominent cluster is recapitulated in all of those and is highlighted. **C.** Volumetric calculations of derived densities for E1p and E3 after low-pass filtering at different resolution thresholds (15-40 Å). E1p is consistently larger than E3 in any resolution and is more than 1.3 times larger than E3 at 22 Å resolution indicating that the 2 molecules can be distinguished. **D.** Fits of E1p low-pass filtered EM map (20 Å) used for unsupervised fitting in the PDHc reconstruction shown in **Figure 44**. When systematically low-pass filtering the E1p and the E3 densities at different resolution (15-40 Å), correct fits are still recapitulated at 22 Å, with a CC threshold of 0.95. correct localization denotes that E1p fits in its density but in an incorrect orientation. **E.** Same as (D), but applied to E3 and derived threshold at 22 Å for a correct fit is CC=0.935. **F.** Analysis of Top 25 E3 and E1p clusters shown in Fig. 7B. Considering the density thresholds for both E1p and E3 fits (CC=0.935 and CC=0.950, respectively), only one of the molecules can be fitted above thresholds in each Location. In addition, density overlaps for each fit above threshold are extensive and C2 symmetric fits are always recapitulated, indicating resolution information is sufficient for discrimination of fits. (Source: Figure reproduced from [140])

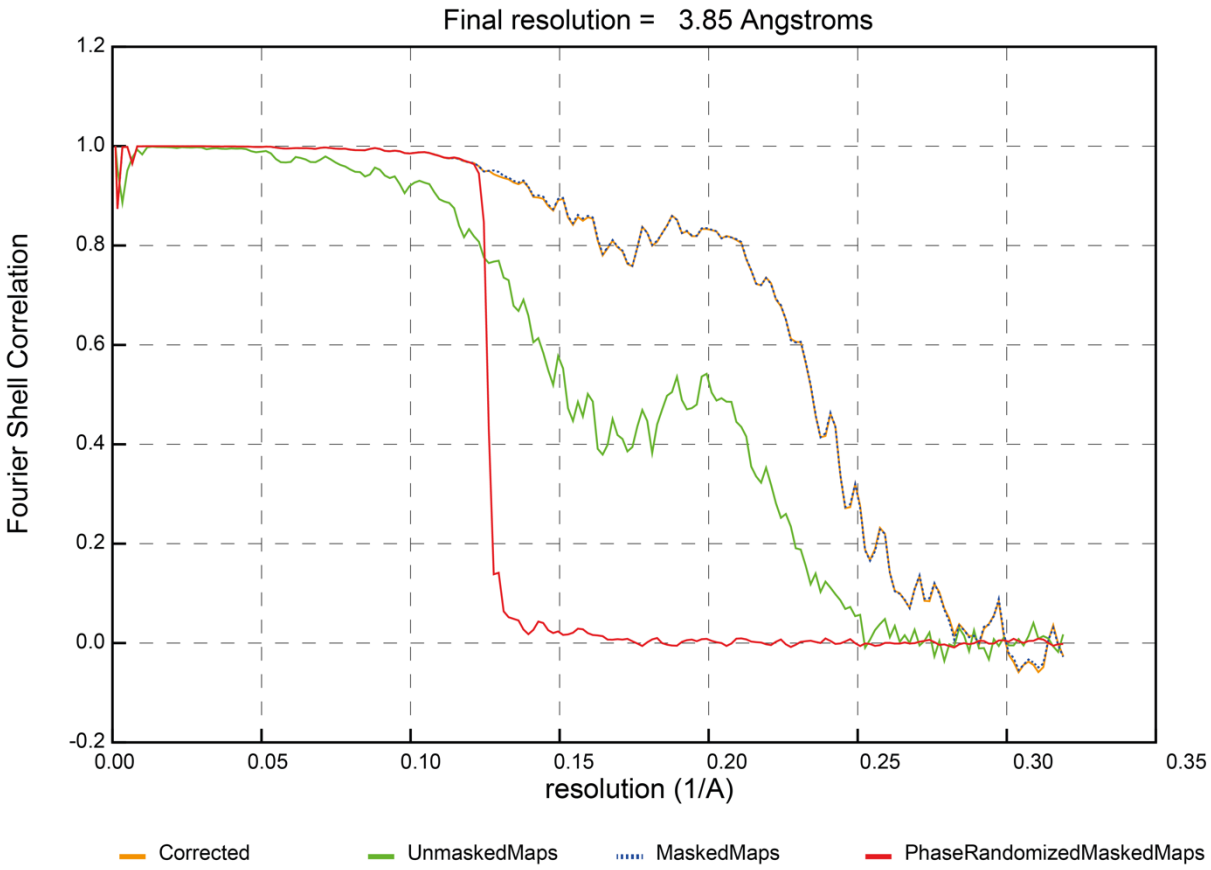


Figure S9. Fourier Shell Correlation of the icosahedrally-averaged cryoEM map corresponding to PDHc core. (Source: Figure reproduced from [183])

7.2 Supplementary tables

Table S1. Nobel prize awards related to high-resolution structural biology.

For biomolecular systems applications and fundamental developments. (Reproduced from [56])

Year	Category	Recipients	Significant contribution
Structure and Function of Biomolecular Processes			
2012	Chemistry	R.J. Lefkowitz & B.K. Kobilka	G-protein-coupled receptors
2009	Chemistry	V. Ramakrishnan, T.A. Steitz & A.E. Yonath	Ribosome
2006	Chemistry	R.D. Kornberg	RNA polymerase
2003	Chemistry (1/2)	R. MacKinnon	Ion channels
1997	Chemistry (1/2)	P.D. Boyer & J.E. Walker	ATP synthase
1988	Chemistry	J. Deisenhofer, R. Huber & H. Michel	Photosynthetic center
1972	Physiol. or Med.	G.M. Edelman & R.R. Porter	Antibodies
1972	Chemistry (1/2)	C.B. Anfinsen	Ribonuclease (Sequence/Structure)
1972	Chemistry (1/2)	S. Moore & W.H. Stein	Ribonuclease (Structure/Function)
1964	Chemistry	D.C. Hodgkin	Penicillin, vitamin B_{12} and Insulin
1962	Physiol. or Med.	F.H.C. Crick, J.D. Watson & M.H.F. Wilkins	DNA
1962	Chemistry	M.F. Perutz & J.C. Kendrew	Myoglobin, oxy-Hemoglobin
1958	Chemistry	F. Sanger	Insulin
1954	Chemistry	L.C. Pauling	Protein secondary structure
Method Development for High-Resolution Structural Biology			
2017	Chemistry	J. Dubochet, J. Frank & R. Henderson	Cryo-Electron Microscopy
2014	Chemistry	E. Betzig, S.W. Hell & W.E. Moerner	Super-Resolution Microscopy
2013	Chemistry	M. Karplus, M. Levitt & A. Warshel	Computational Structural Biology
2002	Chemistry (1/2)	J. B. Fenn & K. Tanaka	Mass Spectrometry
2002	Chemistry (1/2)	K. Wüthrich	Protein Nuclear Magnetic Resonance Spectroscopy
1994	Physics	B.N. Brockhouse & C.G. Shull	Neutron Scattering
1991	Chemistry	R.R. Ernst	High-Resolution Nuclear Magnetic Resonance Spectroscopy
1986	Physics (1/2)	E. Ruska	Electron Microscopy
1985	Chemistry	H.A. Hauptman & J. Karle	Direct Methods for X-ray Crystallography
1982	Chemistry	A. Klug	Electron Crystallography
1946	Chemistry (1/2)	J.B. Sumner	Enzyme Crystallisation
1946	Chemistry (1/2)	J.H. Northrop & W.M. Stanley	Biochemical Purification
1915	Physics	Sir W.H. Bragg & W.L. Bragg	X-ray Crystallographic Analysis
1914	Physics	M. von Laue	X-ray Diffraction of Crystals

Table S2. Data collection and analysis of negatively-stained micrographs from fractions 5-11 along with validation of derived 3D reconstructions. (Reproduced from [140])

	Fraction 5	Fraction 6	Fraction 7	Fraction 8	Fraction 9	Fraction 10	Fraction 5	Fraction 6	Fraction 7	Fraction 8	Fraction 9	Fraction 10	Fraction 11
Data collection and processing	PDHc	PDHc	PDHc	PDHc	PDHc	PDHc	PDHc	Hybrid OGDHc/ BCKDHc	Hybrid OGDHc/ BCKDHc	Hybrid OGDHc/ BCKDHc	Hybrid OGDHc/ BCKDHc	Hybrid OGDHc/ BCKDHc	Hybrid OGDHc/ BCKDHc
Magnification	45000X	36000X	45000X	45000X	45000X	45000X	45000X	36000X	45000X	45000X	45000X	45000X	45000X
Voltage (kV)	200	200	200	200	200	200	200	200	200	200	200	200	200
Electron exposure (e–/Å^2)	30	30	30	30	30	30	30	30	30	30	30	30	30
Applied defocus range (μm)	-0.8 to -1.6	-1.2 to -2.0	-0.6 to -1.6	-0.8 to -2.0	-1.0 to -2.0	[-2.0]	-0.8 to -1.6	-1.2 to -2.0	-0.6 to -1.6	-0.8 to -2.0	-1.0 to -2.0	[-2.0]	-1.0 to -2.0
Calculated defocus range (μm)	-0.2 to -4.0	-0.4 to -4.2	-0.2 to -3.5	-0.2 to -2.5	-0.3 to -4.6	-0.3 to -3.7	-0.2 to -4.0	-0.4 to -4.2	-0.2 to -3.5	-0.2 to -2.5	-0.3 to -4.6	-0.3 to -3.7	-0.2 to -4.0
Pixel size (Å)	3.18	4.03	3.18	3.18	3.18	3.18	3.18	4.03	3.18	3.18	3.18	3.18	3.18
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1
Initial particle images (no.)	155,306	1,407,677	68,587	136,353	148,979	60,069	155,306	1,407,677	68,587	136,353	148,979	60,069	66,323
Final particle images (no.)	9,270	267,547	6,797	4,360	3,095	1,580	4,620	9,000	3,758	3,293	10,819	6,527	1,632
Map resolution (Å) FSC threshold	30.6 0.5	26.6 0.5	38.8 0.5	39.0 0.5	37.9 0.5	39.7 0.5	40.2 0.5	44.6 0.5	39.9 0.5	41.8 0.5	39.2 0.5	52.3 0.5	58.1 0.5

Table S3. Analysis of previously communicated MS data with a focus on α-keto acid dehydrogenase complexes (excel file in the attached CD) (Reproduced from [140])

Table S4. Quantification of clusters surrounding reference-free class averages of negatively-stained single particles and embedded snapshot of 2D class averages (excel file in the attached CD) (Reproduced from [140])

Table S5. HADDOCK energy statistics for derived interfaces of various models of related protein complexes of α-keto acid dehydrogenase complexes. (Reproduced from [140])

Enzymatic assembly	Interface	Organism	Structure	van der Waals energy (kcal·mol^{-1})	electrostatic energy (kcal·mol^{-1})	desolvation energy (kcal·mol^{-1})	buried surface area (Å^2)
			Paired t-test, p-value	*0.675 (n.s)*	*0.422 (n.s.)*	*0.9516 (n.s.)*	*0.8070 (n.s)*
E1p	α/β-Interface	*C. thermophilum*	Homology model	-521.9 +/-3.8	-1524.7 +/-61.2	-134.4 +/-9.4	12846.6 +/-141.3
		Homo sapiens	2OZL	-573.8 +/-11.0	-1446.2 +/-50.3	-147.4 +/-9.3	13593.7 +/-37.7
	αβ/αβ-Interface	*C. thermophilum*	Homology model	-135.0 +/-1.5	-115.6 +/-23.4	-56.8 +/-3.7	3572.3 +/-40.3
		Homo sapiens	2OZL	-144.5 +/-2.4	-296.8 +/-25.1	-36.2 +/-4.1	3439.6 +/-37.3
E2p core PDHc	Intra-trimeric	*C. thermophilum*	Homology model	-117.7 +/- 2.8	-364.1 +/- 21.9	-40.3 +/- 5.3	3234.6 +/- 34.3
		Homo sapiens	6CT0	-114.7 +/- 3.4	-246.5 +/- 21.9	-50.7 +/- 0.5	3046.8 +/- 54.4

(negative stain)	Inter-trimeric	*C. thermophilum*	Homology model	-48.7 +/- 3.4	-102.5 +/- 12.5	-29.4 +/- 3.8	1575.5 +/- 45.9
		Homo sapiens	6CT0	-43.6 +/- 1.2	-237.3 +/- 18.9	-31.6 +/- 4.9	1627.1 +/- 36.7
E2o core OGDHc (negative stain)	Intra-trimeric	*C. thermophilum*	Homology model	-64.8 +/-3.9	-32.2 +/-12.8	-15.7 +/-6.6	2051.6 +/-56.4
		Homo sapiens	6H05	-53.2 +/-1.8	-84.9 +/-22.4	-16.5 +/-6.4	1869.1 +/-72.9
	Inter-trimeric	*C. thermophilum*	Homology model	-31.3 +/-1.8	-66.1 +/-23.6	-14.9 +/-1.0	1121.7 +/-55.2
		Homo sapiens	6H05	-27.9 +/-1.7	-124.6 +/-16.0	-16.2 +/-2.7	1127.1 +/-59.0
E2b core BCKDHc (negative stain)	Intra-trimeric	*C. thermophilum*	Homology model	-124.9 +/- 1.8	-332.8 +/- 26.6	-42.5 +/- 4.4	3072.2 +/- 51.0
		Bos taurus	2II3	-115.4 +/- 3.6	-358.0 +/- 8.4	-11.3 +/- 1.9	2907.1 +/- 39.2
	Inter-trimeric	*C. thermophilum*	Homology model	-86.3 +/- 9.1	-224.8 +/- 52.4	4.4 +/- 6.5	2340.2 +/- 48.1
		Bos taurus	2II3	-74.8 +/- 2.5	-415.7 +/- 24.9	-6.3 +/- 5.6	2263.8 +/- 47.6
E3	Dimeric	*C. thermophilum*	Homology model	-280.8 +/-5.2	-873.4 +/-26.1	-1.0 +/-4.2	7614.9 +/-51.9
		Homo sapiens	6I4R	-289.6 +/-7.5	-739.0 +/-39.1	-11.5 +/-6.1	7779.7 +/-74.9
E2p core PDHc (cryo-EM)	Intra-trimeric	*C. thermophilum*	This publication	-62.4 +/- 1.1	-204.8 +/- 8.7	-103.5 +/- 3.9	1908.0 +/- 26.0
		Homo sapiens	6CT0*	-54.4 +/- 2.2	-133.3 +/- 4.5	-41.3 +/- 2.2	1633.8 +/- 38.7
	Inter-trimeric	*C. thermophilum*	This publication	-54.1 +/- 3.9	-70.5 +/- 11.3	-94.9 +/- 6.8	1781.5 +/- 49.6
		Homo sapiens	6CT0*	-45.5 +/- 1.4	-256.1 +/- 16.2	-32.7 +/- 4.4	1686.7 +/- 37.8

Table S6. Cryo-EM data collection for fraction 6. Analysis of cryo-EM data for icosahedral PDHc core, C1 PDHc core and C1 PDHc complex reconstructions and derived model statistics. (Reproduced from [140])

	Combined			Dataset 1			Dataset 2		
Data collection and processing									
Magnification				45000X			45000X		
Voltage (kV)				200			200		
Electron exposure (e–/Å^2)				30			30		
Defocus range (μm)				-0.8 to -2.0			-0.8 to -2.0		
Pixel size (Å)				3.18			3.18		
Symmetry imposed (model)	C1 (complex)	I (core)	C1 (core)	C1 (complex)	I (core)	C1 (core)	C1 (complex)	I (core)	C1 (core)
Initial particle images (no.)	-	-	-	99,434	99,434	99,434	105,817	105,817	105,817
Final particle images (no.)	-	-	-	9,795	9,765	9,795	19,721	19,721	19,721
Map resolution (Å)	22.06	6.90	8.73	34.53	7.56	12.03	24.06	7.10	8.82

FSC threshold	0.5	0.143	0.143	0.5	0.143	0.143	0.5	0.143	0.143
Refinement									
Initial model used (PDB code)		6ct0							
Model resolution (Å) FSC threshold		6.90 0.143							
Model resolution range (Å)		794.1 – 6.4							
Map sharpening *B* factor (Å^2)		-							
Model composition Non-hydrogen atoms Protein residues Ligands		94500 (monomer 1575) 12480 (monomer 208) 0							
B factors (Å^2) Protein Ligand		 103.25 -							
R.m.s. deviations Bond lengths (Å) Bond angles (°)		 0.40 0.62							
Clashscore		50.0							
Ramachandran favoured (%) allowed (%) outliers (%)		95.1 4.9 0.00							

Table S7. CryoEM data collection, refinement and validation statistics. (Reproduced from [183])

	#1 name (EMDB-13066) (PDB 7OTT)
Data collection and processing	
Magnification	92000 X
Voltage (kV)	200 kV
Electron exposure (e–/Å^2)	30
Defocus range (μm)	-0.5 to -1.5

Pixel size (Å)	1.5678
Symmetry imposed	I
Initial particle images (no.)	296779
Final particle images (no.)	10249
Map resolution (Å)	3.84
FSC threshold	0.143
Map resolution range (Å)	3.5-5.0
Refinement	
Initial model used (PDB code)	7BGJ
Model resolution (Å)	3.85
FSC threshold	0.143
Model resolution range (Å)	4.039-3.9
Map sharpening *B* factor (Å^2)	-94.5985
Model composition	
Non-hydrogen atoms	1583
Protein residues	209
Ligands	n/a
B factors (Å^2)	
Protein	22.55
Ligand	n/a
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.72
Validation	
MolProbity score	1.75
Clashscore	11.80
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	97.10
Allowed (%)	2.90
Disallowed (%)	0.00

Table S8. Key resources table with deposited data of the current study (Reproduced from [140])

Deposited data	**Source**	**Identifier**
Cryo-EM structure of human E2p core	[145]	PDB: 6CT0
Proteomics data	[57]	PXD006660
Crosslinking data	[57]	PXD006626
Crystallographic structure of human E1p	[178]	PDB: 2OZL

Cryo-EM structure of human E2o core	10.2210/pdb6H05/pdb	PDB: 6H05
Crystallographic structure of cattle E2b core	[177]	PDB: 2II3
Crystallographic structure of E3	[146]	PDB: 6I4R
Crystallographic structure of *A. vinelandii* cubic E2p core	[187]	PDB: 1EAA
Model of *C. thermophilum* PDHc E2p core from cryo-EM	This study	PDB: 7BGJ
Cryo-EM structure of native PDHc E2p core (icosahedrally-averaged and asymmetric) and native, asymmetric PDHc	This study	EMDB: EMD-12181
Negative stain reconstructions of PDHc core, fraction 6 and asymmetric reconstructions from fractions 5-10	This study	(D_1292113340)*
Negative stain reconstructions of hybrid BCKDc/OGDHc core, fraction 6 and asymmetric reconstructions from fractions 5-11	This study	(D_1292113352)*
Model of *C. thermophilum* PDHc E2p core from negative stain	This study	(https://data.sbgrid.org/817)
Model of *C. thermophilum* E1p PDHc	This study	(https://data.sbgrid.org/817)
Model of *C. thermophilum* E2o OGDHc from negative stain	This study	(https://data.sbgrid.org/817)
Model of *C. thermophilum* E2b BCKDHc from negative stain	This study	(https://data.sbgrid.org/817)
Model of *C. thermophilum* E3BP trimer from cryo-EM	This study	(https://data.sbgrid.org/817)
Model of *C. thermophilum* E3	This study	(https://data.sbgrid.org/817)

Acknowledgments

It’s 3:39 a.m. 13.11.2021, I’m thinking all of these moments…I am thinking how to start and how to write the next few lines. My main concern is if I will manage to fit in this text all of these thoughts and my feelings for all those people and since I am not that good in writing (even if I do it with my hands that are meant to be my strength as an experimentalist), I will write expressing my heart.

I want to start from my colleagues! I had the luck/honour to be one of the first members of the Lab and thus see it growing and growing for the last almost 4 years. Since our project coordinator Dr. Niesbach-Kloesgen introduced me to the laboratory establishment just from scratch, together with Marie and Rositta, and later on Fabienne! Thanks to this team and to Ulla being always there for my repetitive questions. What can I say for Dr. Meister? Really whatever if I say will not be enough, inspiring calm and respectful scientist, with rare presenting skills and a scientist that I am proud working with! Yiannis or Ioannis as he was introduced in the lab! I will keep the ball down as I might be accused of favouritism as he is a Greek guy but I cannot lie that he was my right hand. Always positive and willing to help even if this had to do with "garbage science". Project-wise he was the person most close as we shared the love for our fungus and the cell extracts protocols! Thanks man! Lisa, the "tinitiny" most delicate hands in our lab, always helpful, whatever she was asked, it was a matter of time to see it being ticked in the check list! Also, a girl with always positive way of thinking and a nice sense of humour! Then came Dmitry, the first Postdoc of our group! My lunch buddy and a friend that taught me a lot of secrets from his long experience in the field! A kind person, but also a powerful scientist that no protein (even disordered ones) can withstand his ability and persistence to reveal the highest possible resolution with his magic image analysis tools! Then came Christian, "a very promising Postdoc" Panos told me! He was absolutely right; this guy is a machine of production ready to code for you whatever you might ask! He was there to discuss with me when something was going wrong with the experiments and always had an idea to the troubleshooting (even when server 18 and 19 had to be restarted running out of memory)! There comes the moment that my other-half, in Lab terms, arrives! "Besyar besyar khosh amadid Dr. Hamdi! (بسیار بسیار خوش آمدید!)" I said and he smiled, since then we spent together more time than I did with anyone else in the underground floor of Biozentrum, preparing samples, optimising conditions, discussing scientific and non-scientific issues, but for

all of these moments this smile and positive energy was still there! He taught me all I know about microscopy and not only! He shared with me the load of work that I had before his recruitment and was always there for me in the good but also in my bad moments! I am thankful to this person for inspiring me and for his immense help! The lab was growing and a guy from Kaiserslautern was about to join, he would be working with nanodiscs and his name was Kevin! It took a few weeks I think to declare him as my young brother, a "bit" stubborn but he has all the life in front of him (being one of the Lab babies) and he will grow into a good scientist! Except a scientific colleague he turned to be my German version spending his valuable time helping me build some proper sentences "auf Deutsch" or translating for me! Also, a biking buddy since I moved to my new neighbourhood two years ago! Thanks for all man! Then was Marija, I didn't have the chance to interact with her that much but I admire her as a scientist and thank her as well! Last group member was Jaydeep! I came to know him mostly at his very beginning in the Lab as we collaborated on the machine learning review that we were co-authors! A great scientist with a nice sense of humour and ready to meet any deadline no matter how strict! I want also to thank all the students Johannes, Felix, Anna for the nice environment in the lab and of course all the scientists that I have collaborated during these years, where I want to name a few great ones including Prof. Dr. Stubbs, Dr. König, Dr. Golbick, Prof. Dr. Paschke, Dr. Otrin, Dr. Hause, Dr. Rubner and Dr. Köck for collaborating and being always kind and helpful with me! A big thank to the people who trained/introduced me at my very first steps in the field, them being Dr. Gavin, Dr. Beck, Dr. Li, Dr. Allegretti, Dr. Zimmerli, Prof. Dr. Mosalaganti, Dr. Sadian, Dr. O'Reilly and Prof. Dr. Rappsilber. I would also like to thank all HALOmem members for the nice environment during our retreats.

I don't want to miss the chance to thank all of my friends from Greece, my previous supervisors and especially Dr. Thanos for introducing me in the scientific world and Dr. Banos to whom I owe the love for science! Except from being my scientific inspiration he is a true friend always there for me!

And now comes the moment that I have to thank the person who gave me the opportunity to write all the above lines! My Professor but also my friend! This man believed in me and I owe him everything that was done here! He trusted me like few people in my life and I replied with hard working and dedication to his plans! An exceptional scientist, an inspiring person with a capacity that has not been revealed even to the half! Thanks for everything from the bottom of my heart Panos! At this point

I have to thank also Marta, the other half of Panos, a very good friend of mine and of my family and a person that I admire!

But this story starts long time ago, 1st of June 1984 and two people are really happy, their new-born baby just came to this world! Since then, their life is dedicated to the happiness of their two children and their families! I feel so blessed being your son and thanks cannot express my feelings so I will just not stop doing my best to keep you proud and happy, Leonida and Katerina! I would like also to thank my brother Christos and his family Ioanna and Leonidas who are always on my side and silently keep an eye on my steps, giving me the balance when needed! Love you!

And now my family, my life, everything! The people who share with me all of my moments, the good and the bad ones, my wife Nancy and our two heroes Leon and Thodoris, people that I breath for and give meaning to my life! Your smiles, your presence on my side and how thankful I feel of you being in my life-trip are values and emotions I cannot express using just a keyboard…this thesis is dedicated to you!

A Big Thank to all of you and to the ones that I might forget!

5:31 a.m., 13.11.2021

P.S. I would like to thank all the members of the Kastritis Group since many of them where not in the Lab when this text was written!

Resume

Personal data

Name:	Fotios L. Kyrilis
Date of birth:	1st June, 1984
Place of birth:	Agrinio, Greece
Marital status:	Married

Education

2018-present **PhD Candidate**
Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Germany

2011-2013 **MSc Degree**
Kapodistrian University of Athens, Medical School, "Molecular Medicine"
Graduation Grade 9/10

2002-2009 **BSc Degree**
Kapodistrian University of Athens, School of sciences, Department of Biology, Graduation Grade 6.3/10

1999-2002 **Public High School of Agrinio, Greece**
Graduation Grade 18.4/20

Professional Experience

2018-present **Staff Scientist/Laboratory Officer in Charge:** Kastritis Laboratory for Biomolecular Research, Martin Luther University, Halle-Wittenberg, Germany

2013-2018 **Molecular Biologist-Laboratory Manager in Charge:** Center of Molecular Analysis and Research (ISO 15189:2012 certified), Agrinio, Greece

2003-2004 **Technical Assistant:** National Chemistry Lab, Subject: Maintenance of lab equipment of analytical Chemistry, University of Athens, Greece

Teaching experience: Practical Courses for Negative Stain & cryoEM sample preparation (Bachelor and Master Students), Molecular Biology Techniques (Bachelor and Master Students), Biology, Physics and Chemistry (High School Students) in Greece and Germany

Research Training

Apr-May 2018 **Visiting Researcher:** Training on Cell Extracts protein purification methods and cryoEM, Dr. Anne-Claude Gavin Lab, EMBL Heidelberg, Germany

May 2018 **Visiting Researcher:** Training on mass spectrometry methods, Dr. Juni Rappsilber Lab, Technical University of Berlin, Germany

2011-2013 **Research Employee:** "The role of histone variant macroH2A in differentiation and development", Dr. Dimitris Thanos Lab, Institute of Biomedical Research, Molecular Biology Laboratory, Basic Sciences, Academy of Athens, Greece

Sep 2011-Feb 2012 **Rotation MSc Project:** "Generation of stable cell lines in the study of *in vivo* lung cancer models - Retroviral technology in the study of human adenocarcinoma", Institute of Biomedical Research, Molecular Biology Laboratory, Basic Sciences, Academy of Athens, Greece

2003-2004 **Diploma Student:** "Molecular and genetic characterization of *At5g14520* gene in *Arabidopsis thaliana* during development and differentiation" (Diploma thesis), Prof. Dr. Kosmas Haralampidis Lab, University of Athens, Greece

Presentations

28 Oct 2021 "**HALOmem Status Seminar,** Halle (Saale), Germany (Virtual Seminar): "Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts" (Oral Presentation)

15-17 Jul 2021 "**Annual HALOmem Retreat Meeting,** Wittenberg, Germany: "Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts" (Oral Presentation)

16-17 May 2019 "**Annual HALOmem Retreat Meeting,** Goslar, Germany: Methods to extract metabolons for high-resolution cryo-EM: updates" (Oral Presentation)

19 Mar 2019	"**HALOmem Status Seminar,** Halle (Saale), Germany: Biochemical methods to extract metabolons for high-resolution cryo-electron microscopy" (Oral Presentation)
24-26 May 2018	"**Annual HALOmem Retreat Meeting,** Rathen-Dresden, Germany: Biochemical methods to extract metabolons for high-resolution cryo-electron microscopy" (Oral Presentation)
17 Dec 2013	"**Noise Plus Kick-Off Meeting,** IMBB Ioannina, Greece: The role of macroH2A in cellular reprogramming and development" (Oral Presentation), Ioannina
5 Dec 2013	"**Graduate Program Lectures,** BRFAA, Athens, Greece: The role of macroH2A in cellular reprogramming" (Oral Presentation), Athens, BRFAA

Participation in Conferences & Seminars

7-8 Dec 2021	**"NWO CHAINS, The Dutch Chemistry Conference**: Structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex" (Poster & Oral Flash Presentation) Veldhoven, Netherlands (Virtual Seminar)
29 Nov-1 Dec 2021	**"EMBO virtual workshop: Recent advances in structural biology of membrane proteins"** Heidelberg (Virtual Conference), Germany
20-22 Oct 2021	**"EMBL Conference: Bringing Molecular Structure to Life: 50 Years of PDB:** Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts" (Poster & Oral Flash Presentation) Heidelberg (Virtual Conference), Germany
14-15 Nov 2019	"**HALOmem International Meeting**: Integrative biology of native cell extracts: a new era for structural characterization of life processes FL Kyrilis, A Meister, I Skalidis, L Schmidt, K Janson, DA Semchonok, C Tüting, F Hamdi, PL Kastritis (Poster) Halle (Saale), Germany
8 Nov 2018	"**HALOmem Status Seminar:** Isolating metabolons for high-resolution cryo-EM imaging" (Poster) Halle (Saale), Germany
14-16 Dec 2012	"**Integer Conference**: Gene Regulation, From DNA Sequence to Nuclear Structure: The role of histone variant macroH2A in cell fate decision" (Poster) Athens, Eugenides Foundation
9-11 Dec 2011	**"62nd National Conference of Biochemistry and Molecular Biology"** Athens, Greece
17-19 May 2007	**"29th Scientific Conference of the Hellenic Society of Biological Sciences"** Kavala, Greece
18-20 May 2006	**"28th Scientific Conference of the Hellenic Society of Biological Sciences"** Ioannina, Greece
31 Mar-2 Apr 2005	**"1st National Conference in Biotechnology and Food Technology"** Athens, Greece

Prizes & Awards

Oct 2021	**Poster Prize Winner:** "EMBL Conference: Bringing Molecular Structure to Life: 50 Years of PDB"
Jul 2019	**Cover for the journal:** "Biological Chemistry"

Publications

First-author publications (Material included in these publications have been used in the current thesis)

- **Kyrilis FL***, Semchonok DA*, Skalidis I, Tüting C, Hamdi F, O'Reilly FJ, Rappsilber J, Kastritis PL. Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. Cell Rep. **2021**, 34(6):108727. doi:10.1016/j.celrep.2021.108727. [I.F.: 9.42]
- Tüting C*, **Kyrilis FL***, Müller J, Sorokina M, Skalidis I, Hamdi F, Sadian Y, Kastritis PL. Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. Nat Commun. **2021,** 12(1):6933. doi:10.1038/s41467-021-27287-4. [I.F.: 14.92]
- **Kyrilis FL,** Belapure J, Kastritis PL. Detecting Protein Communities in Native Cell Extracts by Machine Learning: A Structural Biologist's Perspective. Frontiers in Molecular Biosciences. **2021**, 8. 10.3389/fmolb.2021.660542. [I.F.: 4.62]
- **Kyrilis FL**, Meister A, Kastritis PL. Integrative biology of native cell extracts: a new era for structural characterization of life processes. Biol Chem. **2019**, 400(7):831-846. doi:10.1515/hsz-2018-0445. [I.F.: 3.92]

Second-author publications

- Skalidis I, **Kyrilis FL**, Tüting C, Hamdi F, Chojnowski G, Kastritis PL. Cryo-EM and artificial intelligence visualize endogenous protein community members. Structure. **2022**, S0969-2126(22)00001-6. doi:10.1016/j.str.2022.01.001. Epub ahead of print. [I.F.: 5.01]

Co-author publications (Method developments during this thesis have been used in the following publications)

- Li F, Harvey RD, Modicano P, Hamdi F, **Kyrilis F**, Müller S, Gruhle K, Kastritis P, Drescher S, Dailey LA. Investigating bolalipids as solubilizing agents for poorly soluble drugs: Effects of alkyl chain length on solubilization and cytotoxicity. Colloids Surf B Biointerfaces. **2022**, 212:112369. doi:10.1016/j.colsurfb.2022.112369. Epub ahead of print. [I.F.: 5.27]
- Janson K, Zierath J, **Kyrilis FL**, Semchonok DA, Hamdi F, Skalidis I, Kopf AH, Das M, Kolar C, Rasche M, Vargas C, Keller S, Kastritis PL, Meister A. Solubilization of artificial mitochondrial membranes by amphiphilic copolymers of different charge. Biochim Biophys Acta Biomembr. **2021**, 1863(12):183725. doi:10.1016/j.bbamem.2021.183725. [I.F.: 3.41]
- Otrin L, Witkowska A, Marušič N, Zhao Z, Lira RB, **Kyrilis FL**, Hamdi F, Ivanov I, Lipowsky R, Kastritis PL, Dimova R, Sundmacher K, Jahn R, Vidaković-Koch T. En route to dynamic life processes by SNARE-mediated fusion of polymer and hybrid membranes. Nat Commun. **2021**, 12(1):4972. doi:10.1038/s41467-021-25294-z. [I.F.: 14.92]
- Modicano P, Neumann PR, Schüller M, Holthof J, **Kyrilis FL**, Hamdi F, Kastritis PL, Mäder K, Ann Dailey L. Enhanced optical imaging properties of lipid nanocapsules as vehicles for fluorescent conjugated polymers. Eur J Pharm Biopharm. **2020**, 154:297-308. doi:10.1016/j.ejpb.2020.07.017. [I.F.: 5.57]
- Marušič N, Otrin L, Zhao Z, Lira RB, **Kyrilis FL**, Hamdi F, Kastritis PL, Vidaković-Koch T, Ivanov I, Sundmacher K, Dimova R. Constructing artificial respiratory chain in polymer compartments: Insights into the interplay between bo_3 oxidase and the membrane. Proc Natl Acad Sci U S A. **2020**, 117(26):15006-15017. doi:10.1073/pnas.1919306117. [I.F.: 11.21]
- Hamdi F, Tüting C, Semchonok DA, Visscher KM, **Kyrilis FL**, Meister A, Skalidis I, Schmidt L, Parthier C, Stubbs MT, Kastritis PL. 2.7 Å cryo-EM structure of vitrified M. musculus H-chain apoferritin from a compact 200 keV cryo-microscope. PLoS One. **2020**, 15(5):e0232540. doi:10.1371/journal.pone.0232540. [I.F.: 3.24]

Declaration

Declaration

I hereby declare that I have written this thesis independently and without external assistance other than the mentioned sources and aids being cited in this dissertation. Therefore, any extracts of external works used literally or figuratively in the present thesis are outlined and cited accordingly. I also declare that I have not applied this thesis at any other college or university in order to obtain an academic degree.

Declaration

Erklärung Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe. Ich habe keine anderen als die hier angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inthaltlich entnommenen Stellen als solche kenntlich gemacht habe. Ferner erkläre ich, dass ich mich mit dieser Arbeit an keiner anderen Hochschule oder Universität um die Erlangung eines akademischen Grades beworben habe.

Halle (Saale), Fotios L. Kyrilis