Structure

CellPress
OPEN ACCESS

Perspective

The human touch: Utilizing AlphaFold 3 to analyze structures of endogenous metabolons

Toni K. Träger,[1,2] Christian Tüting,[1,2] and Panagiotis L. Kastritis[1,2,3,4,*]
[1]Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3, 06120 Halle/Saale, Germany
[2]Biozentrum, Martin Luther University Halle-Wittenberg, Weinbergweg 22, 06120 Halle/Saale, Germany
[3]Institute of Chemical Biology, National Hellenic Research Foundation, 11635 Athens, Greece
[4]Interdisciplinary Research Center HALOmem, Charles Tanford Protein Center, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3a, 06120 Halle/Saale, Germany
*Correspondence: panagiotis.kastritis@bct.uni-halle.de
https://doi.org/10.1016/j.str.2024.08.018

SUMMARY

Computational structural biology aims to accurately predict biomolecular complexes with AlphaFold 3 spearheading the field. However, challenges loom for structural analysis, especially when complex assemblies such as the pyruvate dehydrogenase complex (PDHc), which catalyzes the link reaction in cellular respiration, are studied. PDHc subcomplexes are challenging to predict, particularly interactions involving weaker, lower-affinity subcomplexes. Supervised modeling, i.e., integrative structural biology, will continue to play a role in fine-tuning this type of prediction (e.g., removing clashes, rebuilding loops/disordered regions, and redocking interfaces). 3D analysis of endogenous metabolic complexes continues to require, in addition to AI, precise and multi-faceted interrogation methods.

INTRODUCTION

Prediction of biomolecular complex structure and dynamics at atomic resolution, i.e., an accurate three-dimensional model of its function, is the holy grail of computational structural biology. During recent years, a revolution in this field is ongoing, with algorithms such as AlphaFold[1] and RoseTTAFold[2] spearheading the field, initially predicting only static—and often the most stable—conformation of the given target. Despite AlphaFold and RoseTTAFold not including the time component in predicting biomolecular structures, progress in interpreting predicted models in the context of conformational states is rapidly evolving[3] (reviewed elsewhere[4]). Applications of those algorithms, due to their accessible nature, are many,[5,6] and the scientific community has made progress in addressing previously thought intractable challenges; these include, but are not limited to, protein-protein interactions,[7] protein design,[8] conformational changes,[4] protein dynamics,[9] antibody-antigen interactions,[10] pathogenic SNP identification,[11] and, even, biomolecular stoichiometry.[12] Besides these, integration of experimental data with computationally predicted models, pioneered by HADDOCK,[13] has also seen major progress, with the release of entirely new software that combines AI tools with, e.g., cryoelectron microscopy (cryo-EM) data[14] and cross-linking data.[15]

The new release of AlphaFold 3[16] sparked another wave of excitement in the scientific community. This is not only because AlphaFold 3 predicts protein complexes more accurately than its predecessor (76.6% vs. fidelity 67.5% for AlphaFold 2.3 Multimer).[1,16] Note that already AlphaFold 2.3 Multimer outperformed traditional protein complex prediction algorithms.[7] AlphaFold 3 now has the important capability to accurately model biomolecular complexes that include, other macromolecules (DNA and RNA), residue modifications, lipids, cofactors, and ions. Currently, AlphaFold 3 is available as a webserver (https://golgi.sandbox.google.com/) with limited options to model cofactors and with no access to or sharing any computation of the model, except the output. This limitation hampers the systematic and coherent structural analysis of biomolecules. As a result, exploring the full predictive power of AlphaFold 3 is presently impossible due to these closed-source constraints. This black-box solution eventually confines scientific development in biomolecular structure analysis and also restricts our own analysis performed in this perspective to elucidate strengths and weaknesses of AlphaFold 3. Currently, its output *"… cannot be used in docking or screening tools or to train machine learning models or related technology for biomolecular structure prediction similar to AlphaFold Server"* (https://alphafoldserver.com/about). In addition, AlphaFold 3 introduced a new method for estimating the computational complexity of each prediction: the tokenization of amino acids, nucleotide bases, ligand atoms, ions, etc. Tokenization might impose boundaries on structure prediction, particularly for large-scale projects and laboratories with limited funding. However, we should note that supplementary material of the published manuscript openly and in detail reports the updated algorithmic structure as well as the datasets used for training/validation; in addition, the DeepMind team expects to release the executable—including (algorithm) model weights—by the end of 2024. Thinking of such details, it will be of no surprise if AlphaFold 3-inspired code is released soon with a performance equivalent, or even higher, than that

reported[16]; this was witnessed with AlphaFold 2-inspired software, such as ColabFold[17] and OpenFold.[18]

As structural biologists, we teach biochemistry in the University of Halle-Wittenberg and adapted course material to include AI tools to train biochemists.[19] Research-wise, we communicated AlphaFold 2 applications,[14,20] as well as AlphaFold 2 integration with other machine learning software to predict biomolecular interactions from protein complexes identified in native cell extracts.[21] We are currently interested to address the detailed structure-function connections underlying the pyruvate dehydrogenase complex[22]—a 10 megadalton wonder of a molecule which oxidizes pyruvate and generates acetyl-CoA and CO_2. This ancient catalytic reaction, present in all organisms[23]—and in the case of eukaryotes localized within the mitochondrial matrix—fuels the Krebs cycle[24] and has numerous other functions, e.g., moonlighting in the eukaryotic nucleus for histone acetylation[25] and is a central pathway during the Warburg effect.[26]

In essence, the eukaryotic core pyruvate dehydrogenase complex (PDHc) comprises approximately 176 polypeptide chains[27]; we will be focusing here on the fungal PDHc, but the overall architecture must be conserved across eukaryotes with slight modifications.[22] Moderate modifications can also exist,[28] but still, architecturally, PDHc must be comparable across the tree of life.[22,23] This is possibly because if a central reaction has been discovered in evolution, it is very hard to re-invent it or to subject it to faster evolutionary rates. A well-known example of such mechanism lies behind the genetic code, which has remained almost the same across most organisms, too.[29]

The PDHc reaction, dubbed as the *link reaction* in cellular respiration, is known[30]—within PDHc, there are multiple copies of three key enzymatic components: pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2), and lipoamide dehydrogenase (E3). PDHc facilitates the oxidative decarboxylation of pyruvate, yielding acetyl-CoA and CO_2. Mechanistically,[31] pyruvate and thiamine pyrophosphate (TPP or vitamin B1) bind to E1, leading to a CO_2 release by pyruvate decarboxylation and the formation of a hydroxyethyl derivative bound to TPP. Subsequently, the C2-body is transferred to an oxidized lipoamide, covalently bound at a lysine residue in the E2 lipoyl domains (LDs). After translocation to the E2 active site, the acetyl group is then transferred to E2-bound coenzyme A (CoA), generating acetyl-CoA. Finally, the flavoprotein dihydrolipoamide dehydrogenase (E3) catalyzes the re-oxidation of the now fully reduced LD to produce lipoamide-E2 and NADH. This transfer is achieved by a flexible arm which includes the LD and travels from one enzymatic site to another (E1 → E2 → E3). Additionally, in eukaryotes, a fourth subunit, E3BP, binds to the E2 core, anchors E3 near the core,[27,32] and also includes LDs, thus forming the entire metabolon.

The overall architecture of the complex is generally known, as electron microscopy images[33] and crystal structures of the individual domains[22] have been communicated since decades. With the "resolution revolution"[32] in cryo-EM, our laboratory[21,27,34] and others[28,32,35–37] have provided an entirely new view of this metabolon. Although we have made significant progress in understanding the structure-function relationships of PDHc, the complex architecture is still elusive and a plethora of questions remain regarding its structure, catalysis, composition, interactions, cellular localization, role of its flexible regions, regulations of its interfaces by kinases and phosphatases, among others. For starters, no model of the complete PDHc exists to start addressing such questions, although efforts were indeed made[38,39]; these models, although useful, neither harnessed the current structural knowledge for PDHc nor the power of modern AI tools.

Therefore, we sought to model the fungal PDHc using AlphaFold 3. The discussion here is based on AlphaFold 3 server models that we generate for the purpose of this perspective and are freely downloadable.[40] For the evaluation of the derived data, we will deliberately comment on and treat generated solutions in a similar way as performed during manual scoring submissions in the Bonvin group during a critical assessment of prediction of interactions (CAPRI) experiment,[41] mainly visualizing those and evaluating the models based on experience. We will also briefly comment on the predicted local distance difference test (pLDTT) and predicted aligned error (PAE) scores. In short, the pLDDT score indicates the confidence of a single residue, whereas the PAE score indicates the confidence in spatial placement (detailed explanation in the AlphaFold paper[16]). The pLDTT score is quite decent for high-confident regions (e.g., structured domains) but not for flexible regions, which is by now common knowledge and is used for visualization of AlphaFold models in the open-access AlphaFold database.[42] In the following text, we will also show that PAE scores decrease with increased complexity of the analyzed PDHc subcomplex.

AlphaFold 3 accurately models individual PDHc enzymatic components

Briefly, we have previously used modeling approaches based on known templates (termed as homology modeling) and data-driven docking (before AlphaFold 2 came out) to accurately model the *Chaetomium thermophilum* E1, E2, E3, and E3BP subcomplexes.[27] After the release of AlphaFold 2, we immediately predicted its E2 core structure (and corrected it with cryo-EM data[21]), as well as its E3[20] (for another keto-acid dehydrogenase complex because it is a shared subunit). We also predicted the binding of the lipoyl arm to the various PDHc components, and, due to the open-source nature of AlphaFold 2, we were able to tune the generated models to recapitulate the biochemically known distances of the lipoyl with the binding sites across enzymes involved in the keto acid dehydrogenase reaction.[20] Overall, we observed that AlphaFold 2 was ideal to predict highly stable/permanent biomolecular complexes, especially when homologs were structurally known. A prominent example was the generation of the AlphaFold-guided cryo-EM structure of *C. thermophilum* fatty acid synthase: AlphaFold 2 efficiently predicted its higher-order A_6B_6 state,[21] although it was never trained to account for protein-protein interaction principles. Tuning AlphaFold 2 was pivotal to generate models of PDHc enzymatic components, consistent with biochemical knowledge; such adjustments are impossible to perform with the current AlphaFold 3 release, consequently limiting our analysis described in this perspective article.

All these structures mentioned earlier have structural homologs of relatively high similarity determined by crystallography or cryo-EM; therefore, such modeling of orthologs, could be considered quite trivial by relatively experienced molecular

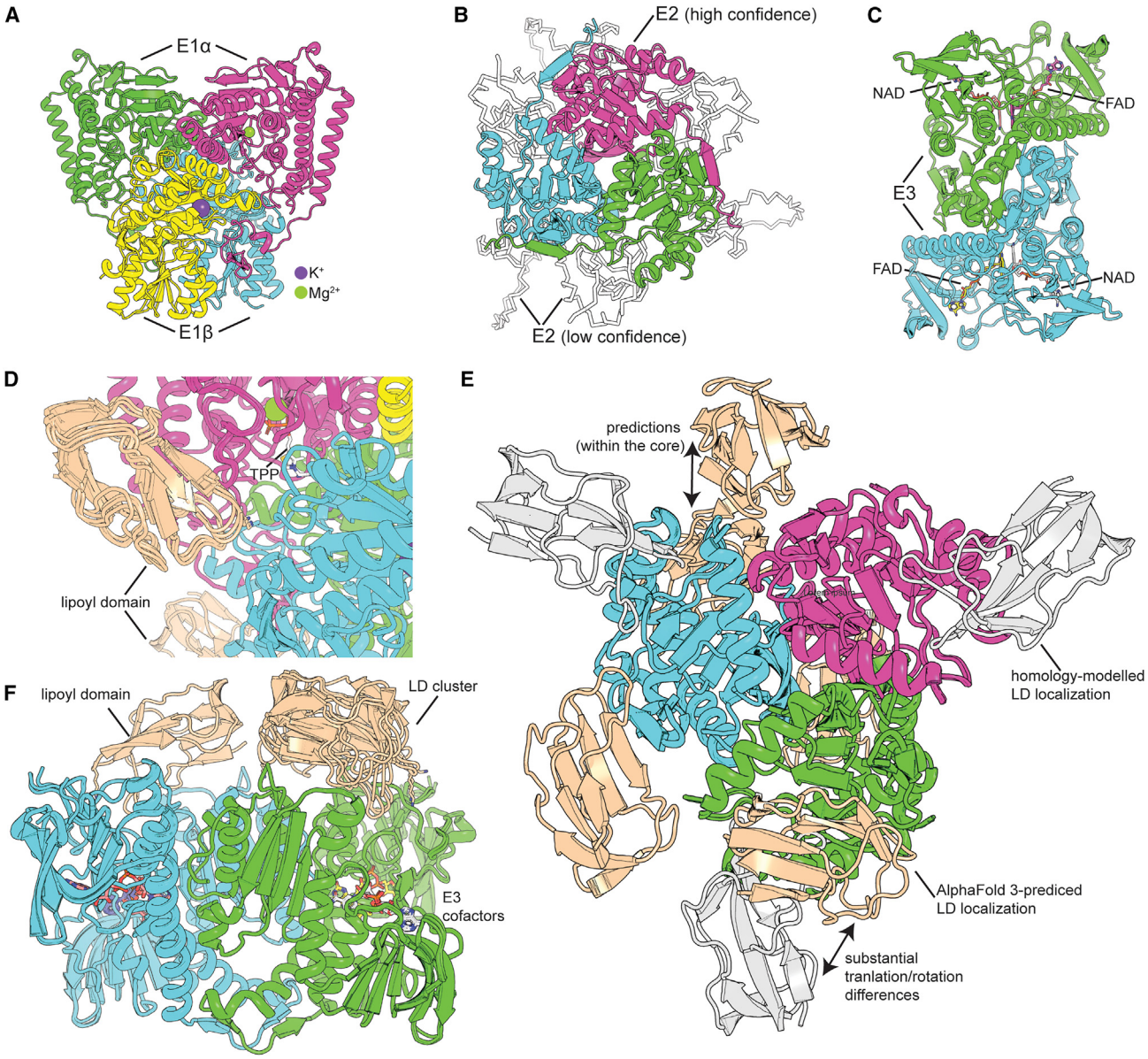

Figure 1. Predicted structures of PDHc components using Alphafold 3

(A–C) Predicted structure of pyruvate dehydrogenase complex (PDHc) components E1, E2, and E3 from *Chaetomium thermophilum*. (C) Enriched structure of PDHc components with ions and co-factors using AlphaFold 3.

(D–F) Predicted models of E1 (D), E2 (E), and E3 (F) in complex with the E2 lipoyl domain (LD).

modelers—but in our opinion, the importance of modeling must not be underestimated. Analyzing such protein complexes without selecting templates, manually editing sequence alignments, or predicting flexible loop regions to be removed from the alignments, has been liberating. Therefore, we always use AlphaFold 2 for most modeling tasks, carefully evaluating derived models with biochemical consistency. Although, still, homology models, especially for structure-based drug design can well be of higher accuracy than the equivalent AI-generated ones—this is because side-chain rotamers, waters, and ions in active sites or interfaces could be critical for hit identification.

AlphaFold 3 can efficiently generate models of the component enzymes that assemble the pyruvate dehydrogenase complex metabolon[22]: Briefly, E1 is an obligate heterotetramer, the E2 from eukaryotes is an obligate 60-mer with remarkable icosahedral symmetry, and the E3 is an obligate dimer. Recently, fungal E3BP has also been analyzed, and is a trimer,[27,32,34,35] residing within the E2 core in 4 distinct copies. Its structure in higher eukaryotes, e.g. humans, follows a substitution model, in which E2 and E3BP form a heterotrimeric core assembly with a stoichiometric ratio of 4:1.[43] AlphaFold 3 models of the fungal *C. thermophilum* E1, E2, and E3 are almost flawless (Figure S1). The models are also enriched with ions and co-factors when AlphaFold 3 is applied (Figures 1A–1C), are highly like those generated by AlphaFold 2, and are closer-to-expectations, after comparing to the derived homology models.[27] This, on its

own merit is outstanding;these PDHc components can be analyzed to understand the individual pyruvate oxidation reactions. Still, 3D modeling of TPP or (acetyl-)CoA is not currently available in the AlphaFold 3 server, but simple superposition of those ligands is entirely feasible, e.g., manually or with AlphaFill.[44]

Prediction of the binding the lipoyl domain binding modes is incomplete

However, to properly understand each of these reactions during pyruvate oxidation, the interactions of each of those enzymes with the LD are critical as it is the carrier of the lipoyl. We have re-modeled all these complexes with the bound LD from both E2 and E3BP sub-sequences and found out that AlphaFold 3 cannot effectively predict these complexes as indicated by the high error in the PAE heatmap (Figure S2). Due to its working mode, AlphaFold 3 will produce atomic models, and these are often in close proximity. We believe, that this is the consequence of limiting the Cartesian box in which AlphaFold 3 will generate the model. In the atomic model, the LD is often in close contact with its binder, but the aforementioned PAE-scores will tell, that this spatial placement is of low confidence. In fact, the top-ranking solution included the interaction of the LD to the E2 core in the atomic model, but this was not observed to be populated in the cryo-EM data.[21] This was the case in the highly related keto acid dehydrogenase complex that produces succinyl-CoA and CO_2 (the oxoglutarate dehydrogenase complex, OGDHc). Anyhow, this transient interaction, mediated by electrostatics,[20] has also been recently reported from the bacterial ortholog,[37] and therefore, it is not surprising that AlphaFold 2 could indeed frequently predict this binding mode.[20]

Furthermore, human supervision revealed that derived models were not highly accurate when PDHc was studied in solution. Cross-linking data supported the interaction of the LD with the different subunits to cover a broader interaction surface, i.e., a guiding surface,[20] and not a highly specific orientation of the LD to the E1, E2, and E3 binding sites. Therefore, we conclude from the AlphaFold 3 models of E1, E2, and E3 in complex with the LD that such transient interactions are still challenging to predict without supervision (Figures 1D–1F, Figure S2). LD is predicted to be in proximity to the active site of E1 (Figure 1D); it is also spread over the surface of the E2 but does not cover known binding modes (Figure 1E); and finally, it is not in reactive distance to the correctly modeled FAD and NAD of E3 (Figure 1F). Aforementioned results might improve if the lipoyl modification becomes available in future releases of AlphaFold 3. Additionally, due to the transient nature of the LD-E1/E2/E3 interactions, integration with biochemical knowledge, proteomics analysis, and molecular biophysics of non-covalent bonds will better address this structural question (as aiming for an AlphaFold model that can satisfy the reactive distance of the lipoyl is possible but not enough for describing transient binding in general). These LD-mediated transient interactions provide activity to the PDHc—if the interaction was too strong it could hamper its normal shuttling function. This makes it an interesting test case for prediction since evolutionary pressure has likely made it weaker than the type of interactions that AlphaFold is trained to detect. Another factor that contributes to this pressure is the integration of the LD to the polypeptide chains of E2 and E3BP, which increases the effective concentration and concomitantly the equilibrium constant for binding. Besides reaction states within PDHc, it is also of note that the simple modeling exercise of the structural protein, the full-length E3BP homotrimer, cannot derive the known trimeric interface (Figure S3).[35]

Prediction of inter-subcomplex interactions is challenging within PDHc

Until this point, PDHc subcomplexes and their interactions with the LD were considered for modeling. However, PDHc is a 10-MDa complex in eukaryotes, includes multiple copies of each of these enzymes, has 12 copies of a structural protein (E3BP), and its elaborate architecture is organized by flexible regions which puzzlingly bring these enzymes in proximity to eventually produce the oxidation products, acetyl-CoA and CO_2. It is known that the peripheral subunit-binding domain (PSBD) of the E2 binds the E1 heterotetramer[45]—the stoichiometry or saturation degree are both unknown, but here, for simplicity, we assume a 3:1 stoichiometry.[27] AlphaFold 3 placed E1 in the correct face of the E2 (Figure 2A). The atomic model is tethered by the three PSBD domains of E2 and the three LD domains are located at around the E1 tetramer (Figure 2B). As previously already mentioned, AlphaFold is bound by its own restriction—namely the prediction box size and the user's input. Taking the PAE scoring (Figure S4A) into account, AlphaFold 3 shows elevated confidence for two out of the three PSBDs (Figure S4A, red and orange box), as well as higher confidence for a single LD domain (Figure S4B, blue box).

The interface predicted is, taking only the atomic coordinates, unlikely in the actual metabolon; however, deconvoluting all results supplied by AlphaFold, atomic model, and PAE scores, the actual metabolon's interfaces can be recapitulated—with *the human touch*.

The two PSDBs with higher spatial confidence are C2 symmetric (Figure S4B), having an all-atom root-mean-square deviation (RMSD) of 0.419 Å. Due to experimental data, we know that the binding of a PSDB to E1 is mutually exclusive, resulting in a 1:1 PSDB:E1 stochiometry.[46] Due to this symmetry and AlphaFold's working mode ("predict everything, the user prompts"), both binding modes are saturated with similar PAE scores. Out of the three LD domains, only one showed elevated PAE scoring. Visualizing the atomic structure, this LD is in fact bound close to the region it is supposed to bind (Figure S3C), whereas the two others are bound off-target. Additionally, the α helix capping the E2 active site is found in endogenous cell extracts to acquire both unfolded and folded conformations,[34] whereas AlphaFold 3 always predicts this α helix in a folded conformation with high confidence both in per-residue (pLDDT) and spatial placement (PAE) scoring (Figure S1B). This observed secondary structure plasticity implies that the E2 binding site for the LD is additionally involved to localize the E1 relative to the E2—E1 can be much further than what is predicted currently in the AlphaFold 3 model if this α helix is disordered (e.g., >10 Å). AlphaFold 3 tends in this case to confidently predict the most stable state—a folded α helix with an additional β sheet,instead of a flexible linker, losing the metabolons' intrinsic flexibility, needed for efficient substrate shuttling. The predicted model also shows physically questionable entanglement (Figure 2C), a limitation the AlphaFold 3 team has explicitly acknowledged in their recent publication,[16] which may become progressively more apparent in 3D models of increased complexity.

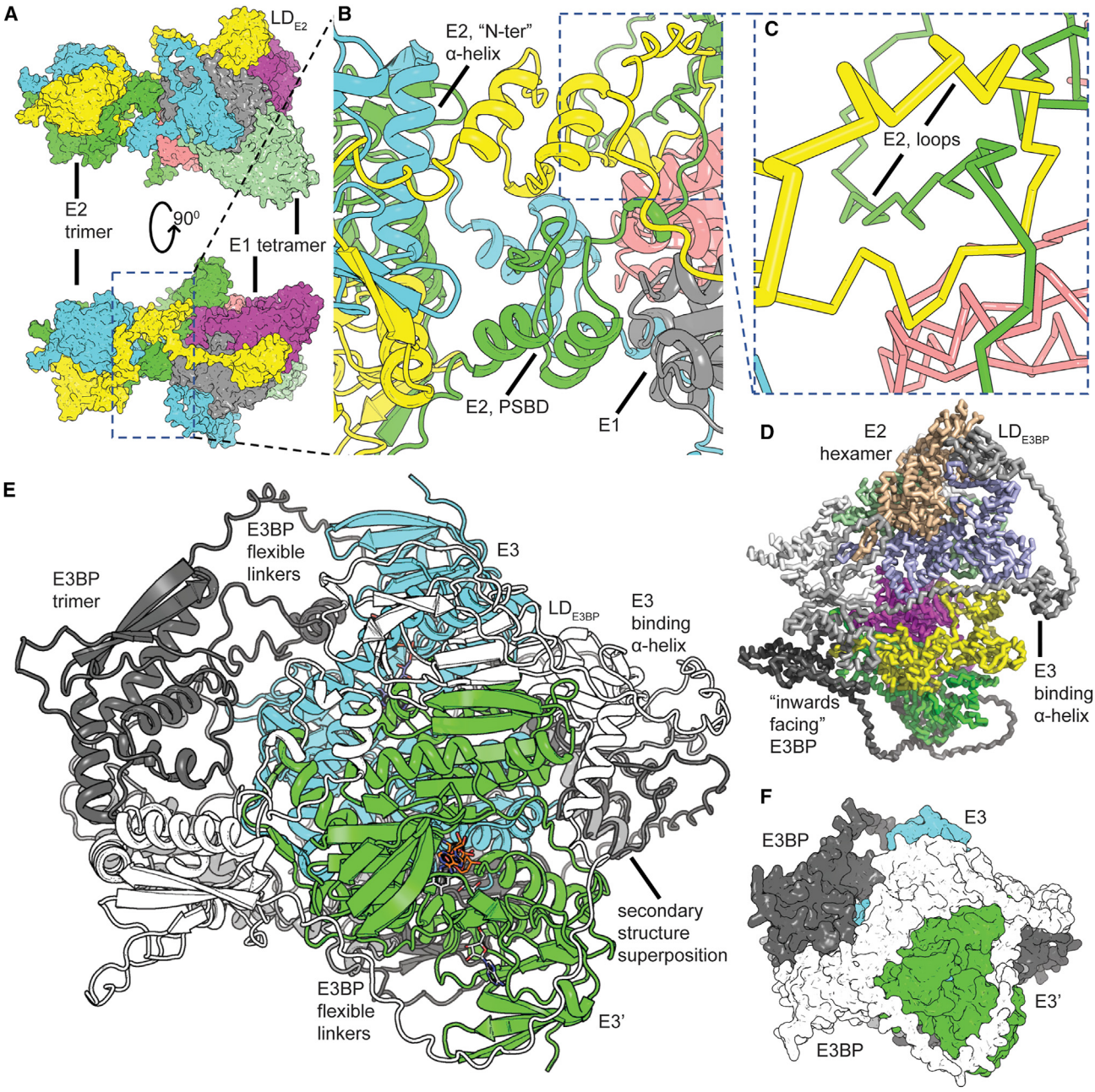

Figure 2. Challenges in predicting subcomplexes within the PDHc metabolon

(A) Predicted complex of E1 and E2 showing the extensive interface mediated by the PSBD.
(B) Zoomed view of the predicted interface between E1 and E2 mediated by 3 PSBDs.
(C) Prediction challenges related to forcibly organized flexible regions in proximity to the complex.
(D) Inaccurate prediction of the fungal E2-E3BP subcomplex.
(E) Accurate prediction of the E3BP trimer in the E3BP-E3 subcomplex which should not have an extensive interface.
(F) Concerns regarding false-positive predictions driven by surface complementarity in the E3BP-E3 subcomplex.

The fungal E2-E3BP subcomplex[35] is not predicted correctly; Here, 6 E2s and 3 E3BPs were used as input to generate the complex, similar to the resolved structure[35] deposited in the Protein Data Bank (PDB, www.pdb.org). Apart from the fact that the E2 multimer has substantial limitations in being generated as a face or an edge-vertex-edge configuration, the E3BP is not predicted to trimerize and is only sometimes localized inwards pseudosymmetrically (Figure 2D). This challenge in predicting this interface can stem from various issues, some of them being (1) absence of this interface from the training data (with consequences related to generalizability); (2) predicting a sub-stoichiometric core complex instead of the 60-mer/12-mer stoichiometry (however, this complex is too large for the server to predict); and (3) presence of the LD domain in the E3BP sequence. This

structural domain presumably forces the E2-E3BP complex to adopt conformations that are a consequence of the ambiguous binding (according to PAE) of the E3BP LD to the E2.

Interestingly, in the E3BP-E3 subcomplex, the E3BP trimeric nature is now revealed (Figure 2E) while in the previous example (E2-E3BP), its trimeric form was not captured (Figure 2D). This shows that the presence of another complex may influence predicted oligomerization state. Furthermore, by visualizing the E3BP-E3 subcomplexes, one with experience in biomolecular docking directly contemplates that these solutions require manual intervention: They appear to be driven by extensive surface complementarity principles, something that has haunted protein-protein docking algorithms in retrieving false positives during scoring for decades. Taking the PAE heatmap into account, AlphaFold 3 seems not to overestimate shape complementary, as a very unlikely spatial interaction is reported. Instead, the structure reveals an intrinsic challenge of the AlphaFold 3 generative intelligence workflow: the input is most likely treated to result in a globular protein structure, and its prediction box size is a function of the number of amino acids. Whereas this makes absolute sense for monomeric proteins and also is most resource effective; this will always produce "globularized" dense protein-protein interactions in multidomain proteins. This is clearly visible in the arbitrary shape complementarity of the E3BP trimer and the E3 dimer (Figure 2F).

Conclusion

Overall, stable PDHc enzymatic subcomplexes with AlphaFold 3 are near-perfect predictions. However, increased complexity of the system is underlined by predictions with highly complementary structural interfaces and the positioning of flexible linkers in direct proximity and periphery of the generated solutions. Especially for the structural component, the E3BP, the trimeric interface can sometimes be predicted but only in the presence of another subcomplex. Implications of this observation indicate that a binding partner can be relevant for accurate protein complex modeling. Including this information during AlphaFold 3 modeling is currently impossible, which again makes using AlphaFold 2, despite its limitations, preferable. Clearly, at least for the PDHc, which has been showcased herein, such higher-order assemblies—and the provided orientation of the flexible linkers, are consistently against current biochemical evidence. It is essential for any user to consult the PAE scores in order to evaluate generated complexes;atomic models from AlphaFold come with quality criteria and must always be considered. The absence of high PAE scores shows that PDHc has a highly intricate architecture, which AlphaFold 3 could only partially recapitulate; Therefore, *de novo* modeling a eukaryotic mini-PDHc (e.g., inspired by the integrative model of the bacterial one[28]) is not currently realistic. It is evident from our preview here that the more composite a biological system becomes, the tougher would be for AI to approximate the unsupervised modeling of its architecture. However, we cannot exclude that by utilizing open-access tools developed for AlphaFold 2 predictions, as mentioned in the introduction, many more insights can be derived for the structure and function of PDHc addressing the aforementioned challenges in analyzing intrinsic flexibility, dynamics, multimerization, and protein interactions. Such analysis goes beyond merely analyzing AlphaFold 3 predicted models and would constitute integrative work that tests the limits of what is achievable for endogenous metabolon modeling in general. Here, we should also point out that traditional docking and molecular dynamics simulation algorithms are also likely to capture the complex dynamics within PDHc—methods which our group has previously applied to this family of complexes.[20,27,34] Finally, PDHc and related keto acid dehydrogenase complexes embed nearly all categories of biomolecular interactions (e.g., enzyme-ligand and enzyme-enzyme complexes; homo- or hetero-multimeric interfaces; permanent and transient complexes; interactions of (dis)ordered regions and their role in regulating enzymatic motions). Such abundance of different structural interactions can serve as an inspiration to other structural biologists and how they can approach their own systems of interest.

To make things short, to solve the structure of the PDHc (or any other endogenous, native metabolon in that matter), and therefore, to understand cellular respiration in a holistic manner, more than a cluster of GPUs running advanced AI software is needed. However, AlphaFold 3 represents a leap forward in structural biology, particularly for elucidating the intricate architecture of smaller or/and higher affinity protein complexes. As the technology continues to develop and integrate with established biochemical and biophysical methods, AlphaFold 3 holds immense promise for unraveling the complexities of larger assemblies and transient interactions within the crowded cellular environment. This collaborative approach, combining the power of AI with human expertise, will unlock a new era of innovation in structural biology.

ACKNOWLEDGMENTS

We would like to thank all the Kastritis lab members for fruitful discussions. This work was supported by the European Union through funding of the Horizon Europe ERA Chair "hot4cryo" project number 101086665 (to P.L.K.), the Federal Ministry of Education and Research (BMBF, ZIK program) (grant nos. 03Z22HN23, 03Z22HI2, and 03COV04 to P.L.K.), the European Regional Development Funds (EFRE) for Saxony-Anhalt (grant no. ZS/2016/04/78115 to P.L.K.), the Deutsche Forschungsgemeinschaft (project number 391498659, RTG 2467), and the Martin Luther University Halle-Wittenberg.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2024.08.018.

REFERENCES

1. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

2. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. https://doi.org/10.1126/science.abj8754.

3. Wallner, B. (2023). AFsample: improving multimer prediction with AlphaFold using massive sampling. Bioinformatics *39*, btad573. https://doi.org/10.1093/bioinformatics/btad573.

4. Sala, D., Del Alamo, D., McHaourab, H.S., and Meiler, J. (2022). Modeling of protein conformational changes with Rosetta guided by limited

experimental data. Structure *30*, 1157–1168.e3. https://doi.org/10.1016/j.str.2022.04.013.

5. Yang, Z., Zeng, X., Zhao, Y., and Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct. Targeted Ther. *8*, 115. https://doi.org/10.1038/s41392-023-01381-z.

6. Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., et al. (2023). Scientific discovery in the age of artificial intelligence. Nature *620*, 47–60. https://doi.org/10.1038/s41586-023-06221-2.

7. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. *13*, 1265. https://doi.org/10.1038/s41467-022-28865-w.

8. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature *620*, 1089–1100. https://doi.org/10.1038/s41586-023-06415-8.

9. Ma, P., Li, D.W., and Brüschweiler, R. (2023). Predicting protein flexibility with AlphaFold. Proteins *91*, 847–855. https://doi.org/10.1002/prot.26471.

10. Marco, G., Constantin, S., Daniel, C., Nikita, D., Charlotte, M.D., and Alexandre, M.J.J.B. (2023). Towards the accurate modelling of antibody-antigen complexes from sequence using machine learning and information-driven docking. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.17.567543.

11. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. https://doi.org/10.1126/science.adg7492.

12. Shor, B., and Schneidman-Duhovny, D. (2024). CombFold: predicting structures of large protein assemblies using a combinatorial assembly algorithm and AlphaFold2. Nat. Methods *21*, 477–487. https://doi.org/10.1038/s41592-024-02174-0.

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

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

15. Stahl, K., Graziadei, A., Dau, T., Brock, O., and Rappsilber, J. (2023). Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. Nat. Biotechnol. *41*, 1810–1819. https://doi.org/10.1038/s41587-023-01704-z.

16. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

17. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

18. Ahdritz, G., Bouatta, N., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T.J., Berenberg, D., Fisk, I., Zanichelli, N., and Zhang, B. (2022). OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Preprint at bioRxiv. https://doi.org/10.1101/2022.2011. 2020.517210.

19. Kastritis, P.L., Bonvin, A.M., Karaca, E., and Schneidman-Duhovny, D. (2022). Impact of AlphaFold on teaching and training in life sciences. Scope and Vision of AlphaFold. https://www.ebi.ac.uk/training/events/scope-and-vision-alphafold/.10.6019/TOL.AlphaFoldTeaching-w.2022.00001.1.

20. Skalidis, I., Kyrilis, F.L., Tüting, C., Hamdi, F., Träger, T.K., Belapure, J., Hause, G., Fratini, M., O'Reilly, F.J., Heilmann, I., et al. (2023). Structural analysis of an endogenous 4-megadalton succinyl-CoA-generating metabolon. Commun. Biol. *6*, 552. https://doi.org/10.1038/s42003-023-04885-0.

21. Skalidis, I., Kyrilis, F.L., Tuting, C., Hamdi, F., Chojnowski, G., and Kastritis, P.L. (2022). Cryo-EM and artificial intelligence visualize endogenous protein community members. Structure *30*, 575–589.e576. https://doi.org/10.1016/j.str.2022.01.001.

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

23. Stacpoole, P.W., and McCall, C.E. (2023). The pyruvate dehydrogenase complex: Life's essential, vulnerable and druggable energy homeostat. Mitochondrion *70*, 59–102. https://doi.org/10.1016/j.mito.2023.02.007.

24. Reed, L.J. (1969). Pyruvate Dehydrogenase Complex. In Current Topics in Cellular Regulation, B.L. Horecker and E.R. Stadtman, eds. (Academic Press), pp. 233–251. https://doi.org/10.1016/B978-0-12-152801-0.50014-6.

25. Sutendra, G., Kinnaird, A., Dromparis, P., Paulin, R., Stenson, T.H., Haromy, A., Hashimoto, K., Zhang, N., Flaim, E., and Michelakis, E.D. (2014). A nuclear pyruvate dehydrogenase complex is important for the generation of acetyl-CoA and histone acetylation. Cell *158*, 84–97. https://doi.org/10.1016/j.cell.2014.04.046.

26. Stacpoole, P.W. (2012). The pyruvate dehydrogenase complex as a therapeutic target for age-related diseases. Aging Cell *11*, 371–377. https://doi.org/10.1111/j.1474-9726.2012.00805.x.

27. Kyrilis, F.L., Semchonok, D.A., Skalidis, I., Tüting, C., Hamdi, F., O'Reilly, F.J., Rappsilber, J., and Kastritis, P.L. (2021). Integrative structure of a 10-megadalton eukaryotic pyruvate dehydrogenase complex from native cell extracts. Cell Rep. *34*, 108727. https://doi.org/10.1016/j.celrep.2021.108727.

28. Meinhold, S., Zdanowicz, R., Giese, C., and Glockshuber, R. (2024). Dimerization of a 5-kDa domain defines the architecture of the 5-MDa gammaproteobacterial pyruvate dehydrogenase complex. Sci. Adv. *10*, eadj6358. https://doi.org/10.1126/sciadv.adj6358.

29. Crick, F.H., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. Nature *192*, 1227–1232. https://doi.org/10.1038/1921227a0.

30. Patel, M.S., and Korotchkina, L.G. (2003). The biochemistry of the pyruvate dehydrogenase complex. Biochem. Mol. Biol. Educ. *31*, 5–15. https://doi.org/10.1002/bmb.2003.494031010156.

31. Prajapati, S., Rabe von Pappenheim, F., and Tittmann, K. (2022). Frontiers in the enzymology of thiamin diphosphate-dependent enzymes. Curr. Opin. Struct. Biol. *76*, 102441. https://doi.org/10.1016/j.sbi.2022.102441.

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

33. Stoops, J.K., Cheng, R.H., Yazdi, M.A., Maeng, C.Y., Schroeter, J.P., Klueppelberg, U., Kolodziej, S.J., Baker, T.S., and Reed, L.J. (1997). On the unique structural organization of the Saccharomyces cerevisiae pyruvate dehydrogenase complex. J. Biol. Chem. *272*, 5757–5764. https://doi.org/10.1074/jbc.272.9.5757.

34. Tuting, C., Kyrilis, F.L., Muller, J., Sorokina, M., Skalidis, I., Hamdi, F., Sadian, Y., and Kastritis, P.L. (2021). Cryo-EM snapshots of a native lysate provide structural insights into a metabolon-embedded transacetylase reaction. Nat. Commun. *12*, 6933. https://doi.org/10.1038/s41467-021-27287-4.

35. Forsberg, B.O. (2023). The structure and evolutionary diversity of the fungal E3-binding protein. Commun. Biol. *6*, 480. https://doi.org/10.1038/s42003-023-04854-7.

36. Liu, S., Xia, X., Zhen, J., Li, Z., and Zhou, Z.H. (2022). Structures and comparison of endogenous 2-oxoglutarate and pyruvate dehydrogenase complexes from bovine kidney. Cell Discov. *8*, 126. https://doi.org/10.1038/s41421-022-00487-y.

37. Skerlova, J., Berndtsson, J., Nolte, H., Ott, M., and Stenmark, P. (2021). Structure of the native pyruvate dehydrogenase complex reveals the mechanism of substrate insertion. Nat. Commun. *12*, 5277. https://doi.org/10.1038/s41467-021-25570-y.

38. Hezaveh, S., Zeng, A.P., and Jandt, U. (2016). Human Pyruvate Dehydrogenase Complex E2 and E3BP Core Subunits: New Models and Insights from Molecular Dynamics Simulations. J. Phys. Chem. B *120*, 4399–4409. https://doi.org/10.1021/acs.jpcb.6b02698.

39. Hezaveh, S., Zeng, A.P., and Jandt, U. (2018). Full Enzyme Complex Simulation: Interactions in Human Pyruvate Dehydrogenase Complex. J. Chem. Inf. Model. *58*, 362–369. https://doi.org/10.1021/acs.jcim.7b00557.

40. Kastritis, P.L. (2024). AlphaFold 3 models for all subcomplexes embedded in the Chaetomium thermophilum pyruvate dehydrogenase complex (Zenodo), p. &squf. Kastritis Laboratory for Biomolecular Research, Zenodo. https://doi.org/10.5281/zenodo.11198823. https://drive.google.com/file/d/1hc379FKOCDx_eZOANnXpK3ichMnkXdHv/forreviewingpurposes.

41. Lensink, M.F., Brysbaert, G., Mauri, T., Nadzirin, N., Velankar, S., Chaleil, R.A.G., Clarence, T., Bates, P.A., Kong, R., Liu, B., et al. (2021). Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment. Proteins *89*, 1800–1823. https://doi.org/10.1002/prot.26222.

42. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

43. Zdanowicz, R., Afanasyev, P., Pruška, A., Harrison, J.A., Giese, C., Boehringer, D., Leitner, A., Zenobi, R., and Glockshuber, R. (2024). Stoichiometry and architecture of the human pyruvate dehydrogenase complex. Sci Adv. https://doi.org/10.1126/sciadv.adn4582.

44. Hekkelman, M.L., de Vries, I., Joosten, R.P., and Perrakis, A. (2023). AlphaFill: enriching AlphaFold models with ligands and cofactors. Nat. Methods *20*, 205–213. https://doi.org/10.1038/s41592-022-01685-y.

45. Fries, M., Jung, H.I., and Perham, R.N. (2003). Reaction mechanism of the heterotetrameric (alpha2beta2) E1 component of 2-oxo acid dehydrogenase multienzyme complexes. Biochemistry *42*, 6996–7002. https://doi.org/10.1021/bi027397z.

46. Frank, R.A.W., Titman, C.M., Pratap, J.V., Luisi, B.F., and Perham, R.N. (2004). A molecular switch and proton wire synchronize the active sites in thiamine enzymes. Science *306*, 872–876. https://doi.org/10.1126/science.1101030.