Structural Reconstruction of E. coli Ubi Metabolon Using an AlphaFold2-Based Computational Framework

Romain Launay, Sophie-Carole Chobert, Sophie S Abby, Fabien Pierrel, Isabelle André, Jérémy Esque

Structural Reconstruction of *E. coli* Ubi Metabolon Using an AlphaFold2-Based Computational Framework

Romain Launay, Sophie-Carole Chobert, Sophie S. Abby, Fabien Pierrel, Isabelle André,* and Jérémy Esque*

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Ubiquinone (UQ) is a redox polyisoprenoid lipid found in the membranes of bacteria and eukaryotes that has important roles, notably one in respiratory metabolism, which sustains cellular bioenergetics. In *Escherichia coli*, several steps of the UQ biosynthesis take place in the cytosol. To perform these reactions, a supramolecular assembly called Ubi metabolon is involved. This latter is composed of seven proteins (UbiE, UbiG, UbiF, UbiH, UbiI, UbiJ, and UbiK), and its structural organization is unknown as well as its protein stoichiometry. In this study, a computational framework has been designed to predict the structure of this macromolecular assembly. In several successive steps, we explored the possible protein interactions as well as the protein stoichiometry, to finally obtain a structural organization of the complex. The use of AlphaFold2-based methods combined with evolutionary information enabled us to predict several models whose quality and confidence were further analyzed using different metrics and scores. Our work led to the identification of a "core assembly" that will guide functional and structural characterization of the Ubi metabolon.
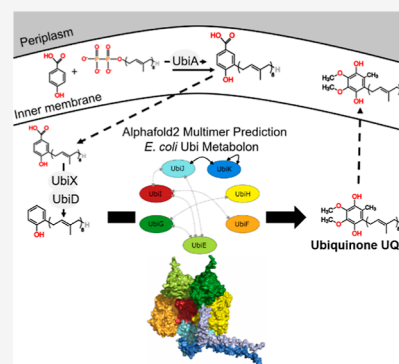
1. INTRODUCTION

Ubiquinone (UQ) is a cellular membrane component of bacteria and eukaryotes.[1,2] It is a redox prenol lipid, having several cellular roles, such as an electron shuttle in the respiratory chain or an antioxidant against various radicals. In view of its roles and its conservation, UQ can be considered as a key molecule for a wide variety of living cells.[3]

Redox property of UQ is related to its aromatic polar head, which can adopt three main redox states: (i) quinone form (oxidized state), (ii) quinol form (reduced state), and (iii) semiquinone form (an intermediate form comparable to a dehydrogenated radical).[4] The polar head is bound to a polyisoprenoid tail, which is hydrophobic.[2] The number of isoprenoid units is species-dependent[5] (from 4 to 14 isoprenyl,[6] with for example 6 for UQ from *Saccharomyces cerevisae* and 10 for that from *Homo sapiens*). In *Escherichia coli*, 8 units constitute the tail, leading to the name of UQ$_8$.

The prenylation of the hydrophobic tail to the polar head is catalyzed by the 4-hydroxybenzoate octaprenyltransferase UbiA, an integral membrane protein in *E. coli*.[7] The product is a precursor of UQ called 3-octaprenyl-4-hydroxybenzoate (OHB). Seven enzymatic modifications are performed on the quinone head of OHB to produce UQ$_8$. The first catalytic reaction is a decarboxylation to form the precursor called octaprenyl phenol (OPP), performed by UbiD (OHB carboxylyase) and UbiX (Flavin prenyltransferase). Then, a series of six enzymatic reactions (three methylations and three hydroxylations) are performed by a cytoplasmic complex

called Ubi Metabolon, a 1 MDa assembly, characterized in 2019 by Chehade et al.[8] A detailed schematic view of these reactions is shown in Figure 1.

Ubi metabolon is composed of seven different protein partners, which can be divided into three groups, (i) hydroxylases (UbiF, UbiH, and UbiI), (ii) methyltransferases (UbiE and UbiG), and (iii) structural subunits (UbiJ and UbiK). Ubi hydroxylases are Class A Flavin MonoOxygenase (FMO).[9] These enzymes use a flavin adenine dinucleotide (FAD) cofactor and molecular oxygen to perform hydroxylation reactions,[10] which correspond to the transfer of a hydroxyl group on specific positions of the headgroup. The molecular mechanism of class A FMO has been well-studied, notably for the reference enzyme named *p*-HydroxyBenzoate Hydroxylase (PHBH).[11] FAD is permanently bound and the reduction mechanism is ensured by a hydrid donor, which can be either reduced Nicotinamide Adenine Dinucleotide or reduced Nicotinamide Adenine Dinucleotide Phosphate (NAD(P)H) transitorily bound at the protein surface.[12] FAD is reduced thanks to conformational changes between two conformations, *in* (catalytic state) and *out*, which is the
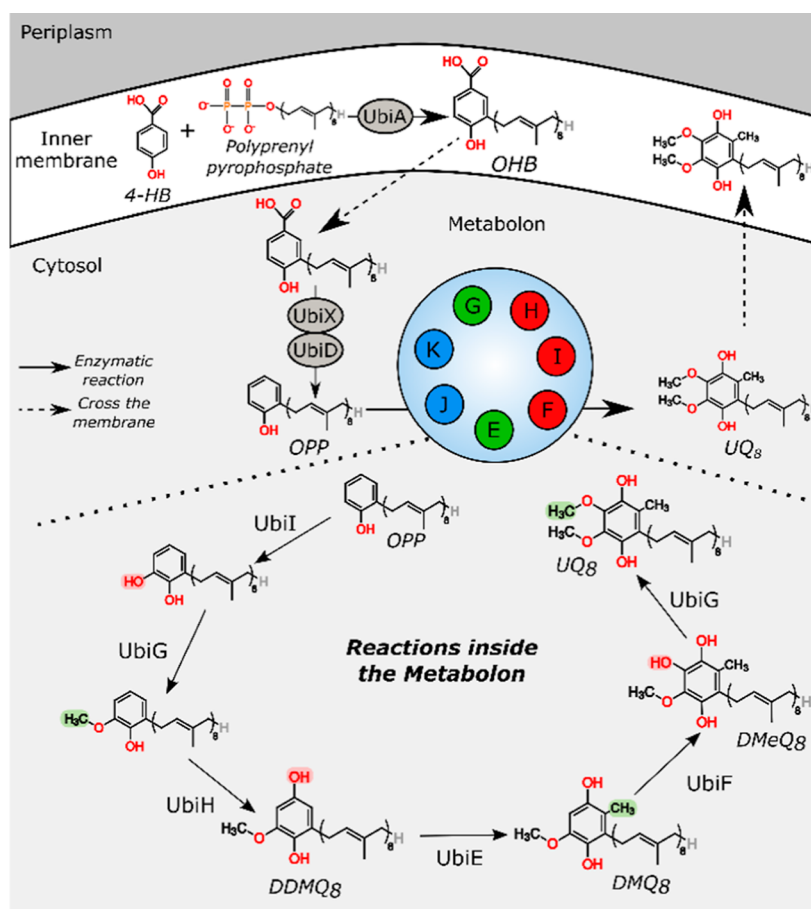
Figure 1. UQ biosynthesis pathway from UbiA enzymatic reaction. Molecules are represented in their reduced forms. Ubi metabolon enzymes (E−K) are colored according their function (in blue, structural proteins, in green, methyltransferases and in red, FMOs). Abbreviations: 4-HB: 4-hydroxybenzoic acid; OHB: octaprenyl-4-hydroxybenzoic acid; OPP: octaprenyl phenol; DDMQ$_8$: C2-demethyl-C6-demethoxy-UQ$_8$; DMQ$_8$: C6-demethoxy-UQ 8; DMeQ$_8$: 6-demethyl-UQ 8; UQ$_8$: Ubiquinone 8 (shown here Ubiquinol corresponding to the reduced form).

conformation enabling the interaction with NAD(P)H. The conformation and molecular oxygen nearby the substrate binding site (quinone head in the case of Ubi FMOs) are required for the hydroxylation reaction.[13]

Regarding the Ubi methyltransferases (UbiE and UbiG), they belong to Class I S-adenosyl-L methionine (SAM)-dependent methyltransferases.[14] In this enzyme family, the methyl group is provided by the SAM (or AdoMet) cofactor, and the reaction is called S_N2 methylation. A linear arrangement between the transferable methyl from SAM and the substrate is required for the chemical reaction. The N-ter region of these enzymes is rather flexible to enable substrate recognition and oligomerization in some representative structures.[15] The demethylated SAM cofactor form, named S-adenosyl-L homocysteine (SAH), corresponds to the coproduct of the reaction and needs to be recycled into SAM. To do so, SAH has to leave the active site, which requires conformational changes.[16] Moreover, it has been shown that UbiG needs to interact with specific lipids in order to release the SAH product cofactor and thus to get a productive system.[17] Finally, Ubi structural subunits (UbiJ and UbiK) are proteins without any enzymatic role but that are essential to the metabolon functioning. Indeed, the presence of an heterotrimer subcomplex (UbiJ−UbiK$_2$) has been shown experimentally to be crucial for the metabolon integrity with a hypothetical role of UbiJ SCP2 domain on the UQ$_8$

binding.[9,18] In our previous study, we have investigated the role of the heterotrimer (UbiJ−UbiK$_2$), proposing a molecular mechanism of UQ$_8$ release into the membrane.[19]

The proteins that compose the Ubi metabolon have been identified.[9] However, some questions are still open such as the three-dimensional (3D) structure of some proteins, their structural organization, and stoichiometry within the Ubi metabolon. In terms of structural information, X-ray structures have been determined for truncated forms of UbiJ[8] (PDB ids: 6H6O, 6H6N, and 6H6P corresponding only to SCP2 domain ~120 first residues) and UbiI[9] (PDB id: 4K22 with residues 35−45 missing, and also the 37 last C-ter residues). In contrast, X-ray structures[14] (PDB ids: 4KDC, 4KDR, and 5DPM) were solved for almost the entire UbiG protein, with only the first nine residues missing. Moreover, very limited information is available regarding interactions of Ubi partners, except for the complex UbiJ−UbiK$_2$ that has been evidenced experimentally[20] and computationally[19] as well as UbiG that was shown to interact with the membrane.[14,17]

It is worth noting that experimental determination of the structural assembly of the Ubi metabolon is quite challenging as it would require substantial quantities of a homogeneous purified biological sample, among other issues. To circumvent these limitations, computational methods appear to be an appealing alternative to predict the supramolecular assembly of

the metabolon and gain knowledge on the protein−protein interactions (PPIs) involved in the complex organization.

Emergence of artificial intelligence-based methods for the prediction of protein structures has led to major advances in the field of molecular modeling and unlocked new ways to better understand protein structures at molecular level. During the CASP14 challenge, the game changer AlphaFold2 (AF2)[21] has outperformed all other methods in the prediction of both monomeric and multimeric 3D protein structures.[22] The key to this success is mostly due to the use of (co)evolutionary information by the deep-learning block, called EvoFormer. Indeed, (co)evolutionary signals have been shown to be important for the accuracy of residue contact prediction,[23,24] even before the development of AF2[21] and AF2-multimer.[25] However, Evoformer block appears to detect and use the information in a more efficient way, as shown by the better results obtained by AF2 compared to other benchmarked methods during the last CASP15 challenge.[26]

More recently, other methods relying on the AF2 algorithm have been developed to predict multimeric structures, such as AF2Complex,[27] AF2-multimer (last release),[25] or MolPC.[28] This latter method, MolPC, combines AF2 with a Monte Carlo tree search in order to assemble molecular bricks and reconstruct the entire macromolecular complex. However, all these methods require a known stoichiometry of the assembly. As the stoichiometry of the different partners of the Ubi metabolon remains undefined, we decided to design our own methodological framework to reconstruct the structural assembly of the Ubi metabolon using a stepwise approach involving the use of different computational methods and tools.

In this article, we report a computational framework relying on the use of the AF2 method to tackle the challenging reconstruction and characterization of the Ubi metabolon at the molecular level. Our results combined with available experimental information enabled us to propose a 3D model of the minimal unit forming the Ubi metabolon and to shed some light on its mode of action, notably the importance of the stoichiometry of protein partners within the molecular assembly, as well as substrate channeling along the UQ$_8$ biosynthesis pathway.

2. MATERIALS AND METHODS

2.1. Sequence Data Set. 17514 complete genomes of the phylum *Proteobacteria* (recently renamed *Pseudomonadota*) available at the NCBI database were downloaded (last accessed in July 2022).[29] One genome for each species (from species taxonomy ID) was selected amounting to a total of 3568 genomes. Genomes without CDS annotation were annotated with Prokka (v1.14.5).[30]

We used HMM (Hidden Markov Model) profiles for UbiI, UbiH, UbiF, UbiE, UbiG, UbiJ, and UbiK published in Pelosi et al.,[31] Kazemzadeh et al.,[32] and Launay et al.[19] A total of 10 Ubi protein profiles were used, including 7 profiles for the genes of interest and 3 homologues, as well as 12 "decoy" profiles designed to increase the annotation specificity of some genes.

We performed a similarity search on the set of Proteobacteria genomes with the hmmsearch program (HMMER suite v3.3.2)[33] and selected the hits with a profile coverage (alignment length over sequence length) above 0.5 and an *i*-evalue below 0.001. The hit with the best *i*-evalue was selected when multiple profiles matched the same sequence.

From this approach, 3524 sequences of UbiE, 2208 sequences of UbiI, 878 sequences of UbiF, 2189 sequences of UbiJ, 2023 sequences of UbiH, 3497 sequences of UbiG, and 2974 sequences of UbiK were available. Due to a redundancy between the organisms and a diversity of amino acid sequence length, a filtering was done to keep only one sequence per species and to remove sequences having more than 150% of the corresponding *E. coli* sequence length. Finally, 3188 sequences of UbiE, 2129 sequences of UbiI, 878 sequences of UbiF, 2155 sequences of UbiJ, 2023 sequences of UbiH, 2923 sequences of UbiG, and 2974 sequences of UbiK were aligned using Muscle.[34]

2.2. Prediction of 3D Models. Local version of AF2 from GitHub (https://github.com/deepmind/alphafold, accessed on 7 April 2022) was used with default parameters to model monomeric (AF2 V2.1) and multimeric (AF2-multimer v2.2) structures.[21,25]

Multimeric structures were also predicted using Local-ColabFold,[35] an alternative version of AF2. The difference with AF2 is the sequence search to feed the neural networks. In this version, we can either use our customized multiple sequence alignment (MSA) or the automatic search with MMseqs,[36] which is 40−60-fold faster than AF2 search.

Regarding monomer modeling with AF2, several MSAs were built as follows: (i) JackHHMER (HMMER 3.3.2[37]) and HHBlits (HHBlits 3.3.0[38]) were used to search for sequences from Uniref90 (v2020_01[39]), MGnify (v2018_12[40]), Uniclust (v2018_12[41]), and big fantastic database (BFD), (ii) hmmsearch (HMMER 3.3.2) was used to search for sequences having structural information from the PDB70 database.[42] For each targeted sequence, five models were generated.

Regarding multimer modeling using the AF2-multimer, the only difference is the utilization for sequence search of the UniProt database in addition to other databases and the generation of 25 models at the end.

Regarding multimer modeling with localcolabfold, MSAs were built on a distant server using MMseqs Targeted sequences are aligned against protein sequences from several databases: (i) Uniref30 (v2022[39]), (ii) a combination of BFD and MGnify, (iii) PDB70 database, and (iv) ColabFoldDB.[35] Pairs and unpairs sequences were used. Modeling of multimers was also performed using our MSAs from designed data set.

Localcolabfold parameters are the default ones, with three prediction recycles (which is the default recycling number in AF2 implementation). Five localcolabfold calculations with different random seeds were performed for each target, to generate in total 25 models.

All 3D models were relaxed with default parameters of AF2 using AMBER Force Field.

2.3. Software. TM-score values were computed from TM-align[43] for dimeric forms and MM-Align[44] for multimeric ones. Conservation analyses have been performed using Consurf.[45] Autodock Vina 1.2[46] has been used to perform docking assays, using default parameters. The analyses were performed using Python3 and library. Visualization of 3D structures was done using Pymol 2.5.[47]

3. RESULTS AND DISCUSSION

The computational framework is decomposed in successive steps described in Figure 2. The first step involved the prediction of the 3D structure of the proteins as monomers. Next, all the 3D structures of the pairs of proteins were predicted using three different MSAs. Protein pairs predicted
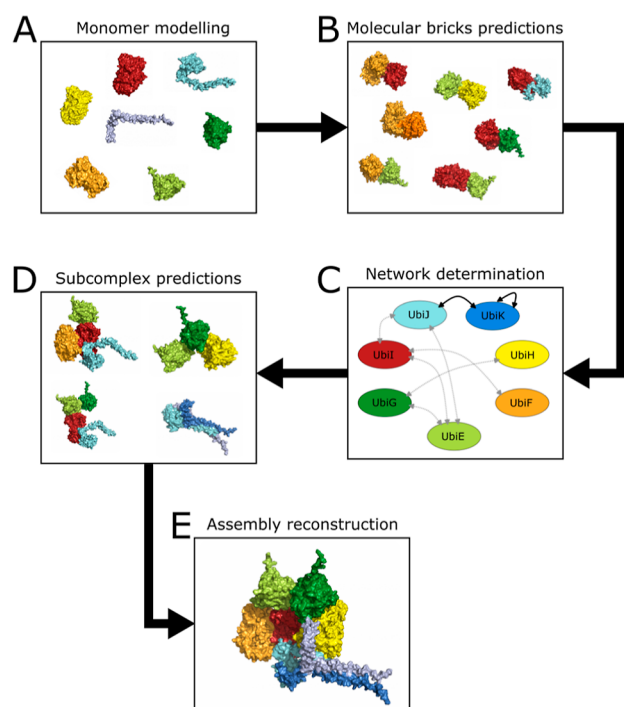
Figure 2. Schematic representation of the computational framework used to reconstruct the Ubi assembly, with the successive steps (A) monomer modeling, (B) molecular bricks predictions, (C) network determination, (D) subcomplex predictions, and (E) assembly reconstruction.

Table 1. Global plDDT and rmsd (Å) Range for the Enzymatic Subunits of Ubi Metabolon

	UbiE	UbiG	UbiF	UbiH	UbiI
plDDT	88.4-90.3	83.8-89.6	84.5-89.5	86.4-91.2	84.2-87.9
rmsd	0.5-0.8	0.2-1.2	0.2-1.2	0.2-0.4	0.1-0.8

3D models for each enzyme. Both results indicate that the global AF2 quality of each best model is reliable. To go further in the analysis, local plDDT scores were investigated, and the results are shown in Figure 3 only for the five best 3D models. As expected, all enzymes were modeled with a good confidence in view of the number of residues with a plDDT score higher than 70. An analysis was performed at the residue level, with the computation of the percentage of amino acid residues having a plDDT inferior to 70. This led to the following results, 2.4% for UbiE, 9.6% for UbiG, 6.1% for UbiF, 1.3% for UbiH, and 8.3% for UbiI. Regarding other residues, the (very-)low confidence could be related to the low number of sequences in the MSA at these positions. Moreover, low confidence regions predicted by AF2 are in general associated with flexible or disordered regions.[21]

In the case of the methyltransferases (UbiG and UbiE), the N-ter region shows low confidence values, suggesting a higher flexibility as often observed in terminal polypeptide regions (Figure 3A,B). Moreover, it was shown that the N-ter region in Class I methyltransferases could be involved in the recognition of the substrate or homodimerization process.[15] For UbiG, another region, between residues 161 and 191, was predicted with a low confidence. Interestingly, this region was identified to interact with lipids for SAH release, especially between residues 177 and 186 corresponding to a conserved region called CR III.[14] This region is flexible and can adopt "open"-"closed" conformations, explaining the low plDDT scores predicted by AF2.

In the case of hydroxylases (UbiI,H,F), analysis of all best 3D models revealed a low confidence for the C-ter region, which is predicted to adopt three kinked helices and a terminal disordered/flexible part (Figure 3C–E). Even if they adopt a similar 3D organization, a diversity in terms of sequence length is observed, ranging from residues 349 to 373 for UbiF, from residues 346 to 380 for UbiI, and from residues 358 to 364 for UbiH. A similar observation is found for other 3D models of each enzyme showing larger regions with a poor confidence in C-ter, revealing the uncertainty of this region. It is worth noting that X-ray structure of UbiI could be determined only after truncating the last 36 residues of the protein,[9] explained by a probable flexibility of the C-ter region.

3.1.2. Experimental Validation. As mentioned previously, some X-ray structures of Ubi enzymes or homologous sequences are available in the Protein DataBank,[42] either partially determined (missing residues) or almost completely solved (with only few missing residues in N-ter or C-ter regions). To assess the quality of Ubi 3D models, structural comparisons were performed against experimental data, i.e., crystallographic structures. rmsd and TM-score using the TM-align program[43] were computed and used as main metrics for further analyses. The structural differences are shown and discussed hereafter for each class of enzymes, (i) methyl-

with high confidence were then selected to construct a network of possible PPIs, which was used subsequently to predict subcomplexes. Finally, the complete assembly was reconstructed based on the different subcomplexes. This framework is detailed in the following.

3.1. Monomer Modeling. The first step of our framework aimed at constructing 3D models of monomeric forms of the seven protein partners involved in the Ubi metabolon. As the AF2-multimer will be used to predict monomeric structures, we predicted the different protein subunits using AF2 for validation. As UbiJ and UbiK structure predictions have already been described in a previous work,[19] we focused on the modeling of the five other proteins (UbiE, UbiF, UbiG, UbiH, and UbiI) involved in the metabolon.

3.1.1. AF2 Predictions. Prediction of monomeric structures was performed using default parameters of AF2 (see Material and Methods). The quality of the 3D models was analyzed using the predicted local-distance difference test (plDDT), an AF2 metric providing a confidence on the modeling of each amino acid. This metric, based on the model/structure comparison score lDDT,[48] ranges from 0 (very low confidence) to 100 (high confidence). Moreover, the global plDDT has been shown as a good estimator of the predicted structure quality. Based on this consideration, the global plDDT values were analyzed for five models of each enzyme (UbiI,H,F,E,G) and the corresponding value ranges are given in the Table 1. All global plDTT values are above 80, indicating a good confidence in the models. To investigate the conformational diversity of the 3D models, we computed for each enzyme the root mean square deviation (rmsd) values on Cα between the best structure (Figure 3) against other 3D models. All Cα-rmsd values are given in Table 1 and are lower than 1.25 Å, showing a convergence between the five predicted
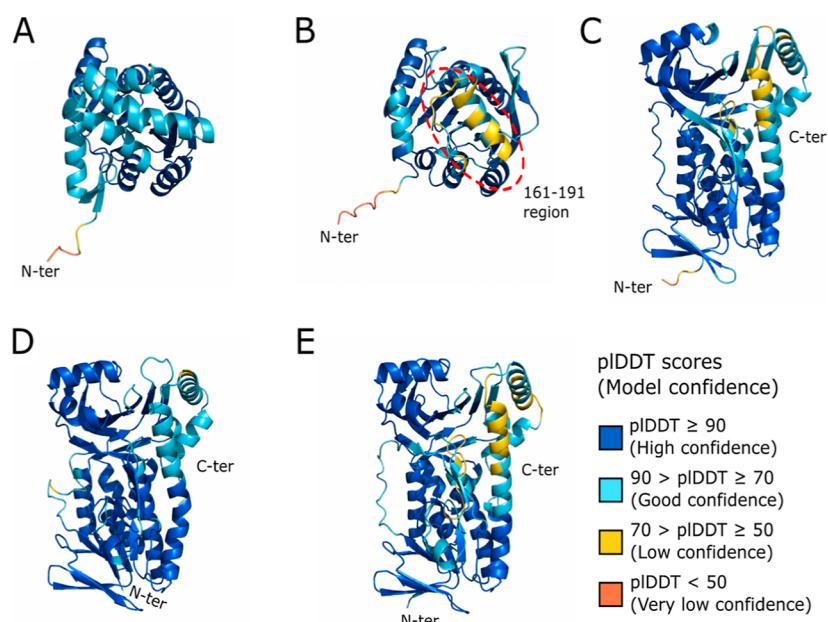
Figure 3. 3D structure of the best predicted model of (A) UbiE, (B) UbiG, (C) UbiF, (D) UbiH, (E) UbiI, colored according to their plDDT score and shown in cartoon.

transferases (UbiE and UbiG), (ii) hydroxylases (UbiI, UbiH and UbiF).

For UbiE, structures of a homologous protein (Coq5 of *Saccharomyces cerevisiae*,[49] 38% sequence identity) have been determined with a high resolution of 2.2 Å for 4OBX and 2.4 Å for 4OBW.[50] rmsd calculated on $C\alpha$ between the X-ray structures resulted in a value of 0.28 Å, revealing a very similar conformation between the structures. Comparison between the best AF2 model and 4OBX (highest resolution) provided a TM-score of 0.88 and a rmsd value of 1.59 Å. Main differences between both structures reside in the conformation of the loops (due to their different lengths) and in the orientation of some helices, which differ slightly. Nonetheless, this highlights the structural similarity between the X-ray structure and the AF2 model, emphasizing the high confidence on the quality of the 3D model.

For UbiG, three X-ray structures of the corresponding *E. coli* sequence were determined with a high resolution of 2.0 Å for 4KDR, 2.1 Å for 4KDC[14] and 5DPM. Interestingly, the holo form in complex with SAH (4KDR and 5DPM) and apo form (4KDC) were determined, and some conformational changes were highlighted in the presence of the cofactor SAH, as shown by rmsd value calculations, e.g., 0.16 Å between 4KDR and 5DPM (holo forms) versus 2.23 Å between 4KDR (holo form) and 4KDC (apo form). Structural superpositions of X-ray structures are displayed in Figure 4A, and the main result is the observation of three regions, which seem to be key for the dynamics of the cofactor entry/release: (i) N-ter residues from 1 to 36, (ii) residues from 160 to 193, and (iii) residues from 212 to 226. When comparing the best AF2 model with 4KDR (holo form) and 4KDC (apo form), rmsd values are found to be 1.43 and 2.37 Å, respectively. This result indicates that AF2 model is structurally closer to the holo form compared to the apo one. The structural differences were visually inspected by superimposition as shown in Figure 4B,C. The N-ter region was modeled in a conformation similar to 4KDR, which may influence the orientation of the two other regions (160−193 and 212−226). Of note, the region 160−193 corresponds to a
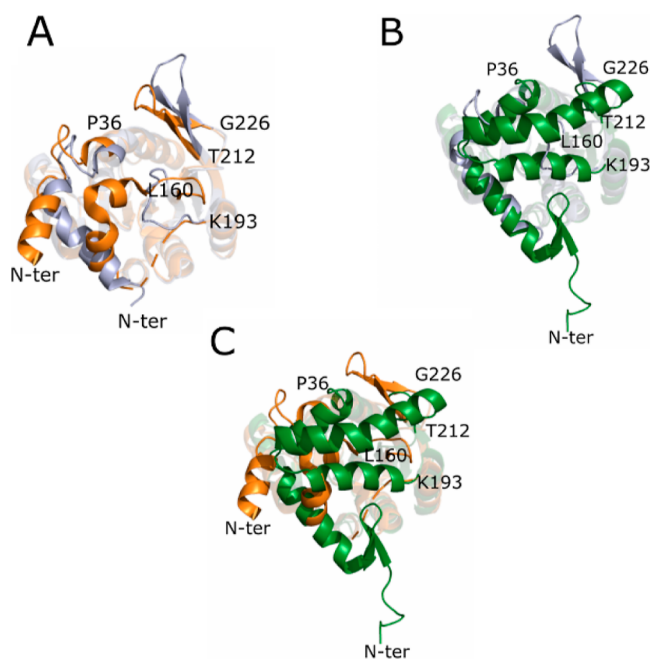

Figure 4. Cartoon representation of 3D structure of UbiG. (A) Superimposition of 4KDR (in gray-blue) and 4KDC (in orange) structures. (B) Superimposition of 4KDR and AF2 (in green) structures. (C) Superimposition of 4KDC and AF2 structures. Regions 1−36, 160−193, and 212−226 are highlighted with no transparency.

medium confidence region predicted by AF2 (with both good and low confidence plDDT scores), indicating the uncertainty of the position of some residues. This is consistent with the different conformations observed in X-ray structures. Interestingly, this region was identified to be involved in interaction with lipids to enable cofactor release, which highlights its functional role. This information is important for the modeling of the metabolon assembly as we hypothesize that this region should be accessible to ensure its function.

For UbiI, the X-ray structure of a truncated *E. coli* enzyme was solved with a high resolution of 2.0 Å (PDB id: 4K22[9]). Structural alignment between the best AF2 3D model and the X-ray structure of the truncated form provided a rmsd value of 1.69 Å. This indicates that the AF2 model is on average structurally close to the X-ray structure. The residues missing in the X-ray structure correspond to two key regions, (i) the flexible loop (residues from 35 to 49) and (ii) the C-ter (residues 364–400). Interestingly, these regions were solved in other 3D structures from homologous or related proteins. For example, the corresponding flexible loop of UbiI was determined in other Class A FMO, such as PHBH (PDB ids: 1D7L,[51] 1IUT,[52] 1PBE,[53] 1DOD,[54] 7ON9...[55]), PHHY (1FOH[56]), MHBH (2DKH[57]), PgaE (2QA1[58]), CabE (2QA2[58]), RebC (2R0C[59]), PhzS (2RGJ[60]), DHPH (2VOU[61]). This suggests that the presence of FAD could help to stabilize the loop, allowing, therefore, its structure determination. Indeed, with the exception of UbiI (4K22), all other X-ray structures were cocrystallized in the presence of FAD. When superimposing Class A FMO structures with that of PHBH (1PBE), we observed that the loop is packed against the FAD binding site, engaging in some interactions (Figure S1A). This closed conformation was not observed in the best AF2 model, which adopts an open conformation with only limited contacts with FAD (Figure 5A). This emphasizes one of the limits of AF2 in protein structure prediction as ligands are not considered in the 3D modeling.[62] Regarding the C-ter region, the X-ray structure of a homologous protein of UbiI, called VisC, from *P. atrosepticum* was solved with a high resolution of 2.5 Å (PDB id: 4N9X). The flexible loop was not

determined, and no FAD was cocrystallized. However, the C-ter region was structurally resolved and shares a sequence identity of 72.2% with UbiI using Needleman-Wunsch alignment algorithm from EMBOSS Needle.[63] As a reference, 4N9X was superimposed on the best AF2 model, highlighting some structural differences. Indeed, both 3D structures are composed of three α-helices connected by loops, but their lengths differ (Figure 5B). Helix boundaries are 328–357, 361–377, and 379–389 for UbiF, 326–355, 359–374, and 376–387 for UbiH, and 329–357, 362–377, and 381–390 for UbiI. Of note, the asymmetric unit of 4N9X is a homohexamer, whereas 4K22 (UbiI) is a homotetramer. For both X-ray structures, one unit of oligomerization involved the C-ter region. However, the crystallization of UbiI was made possible thanks to the truncation of the C-ter. Therefore, a potential bias of the crystal packing could be introduced in 4N9X. Moreover, UbiI or VisC are also expected to interact with other Ubi proteins, limiting the number of copies in the homo-oligomerization.

To go further in the analysis, the best AF2 model of UbiI was compared to the X-ray structure of PHBH (1PBE), showing a higher structural similarity than with 4N9X in the C-ter region [TM-score between UbiI (residues 331–400) and 1PBE (323–391) is 0.67 versus 0.54 between UbiI and 4N9X (331–399)]. Thus, several attempts (data not shown) were carried out to force AF2 to sample more extensively conformational diversity and identify a conformation closer to 4N9X, but all results were found nearly similar. In all cases, AF2 did not model this region with a high confidence. This could be due to the conformational diversity and the flexibility observed in the superimposition of X-ray structures (Figure S1B). Indeed, some of these X-ray structures are composed of three helices with different lengths (1PBE, 2QA1, 2QA2), whereas for others, this region was solved in a less structured conformation (loops) like in 1FOH, 2VOU, or was missing (like in 2DKH, 2R0C, 2RGJ). This indicates that modeling of the C-ter region is challenging and thus, all predicted PPIs involving this region have to be considered with caution.

Considering that UbiF and UbiH are homologous proteins to UbiI (respectively 37 and 30% of sequence identity)[32] and that they share structural similarity (TMscore of 0.95 and 0.94, rmsd of 1.54 and 1.75 Å, respectively for UbiF and UbiH in comparison with UbiI), the same remarks as for UbiI can be drawn.

To conclude, AF2 enabled us to obtain 3D models of the five enzymes (UbiE,G,I,H,F) with a good quality on average. However, some key regions were identified to be potentially flexible which can impact the confidence on the 3D model of the metabolon assembly.

3.2. Prediction of Molecular Bricks. In order to explore the protein interaction combinatory, pairwise modeling of the seven Ubi proteins was performed using the AF2-multimer.[25] This strategy aimed at identifying the most probable interfaces, which could be used subsequently to build a protein interaction network. To enhance our confidence on AF2 results, three approaches, each generating 25 models, were followed: (i) classical approach using AF2 MSA by default, (ii) localcolabfold approach using MSA from MMseqs search (MMseqs MSA), (iii) localcolabfold approach using our customized MSA from a curated data set of related sequences, the main difference being the initial information derived from the MSA. This will enable us to verify the convergence and the
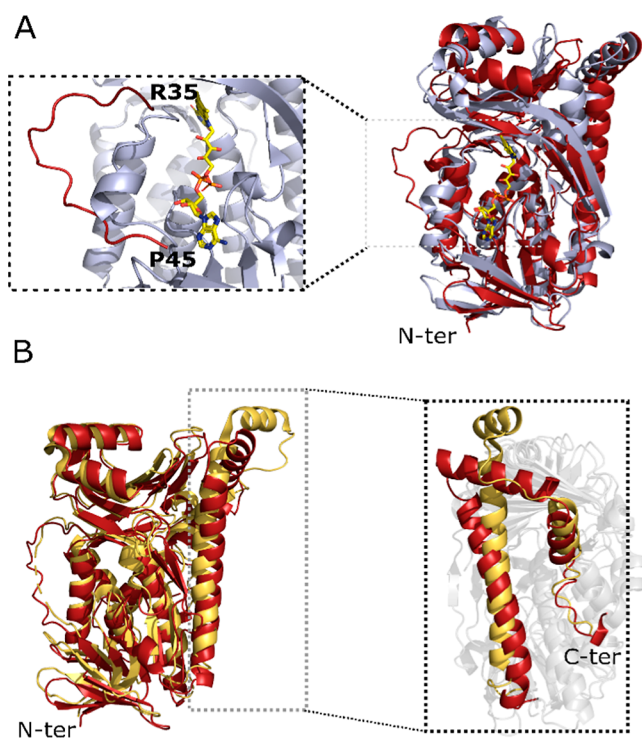

Figure 5. Structural comparison of FMOs. (A) PHBH structure superimposed with UbiI (in red), with a zoom on the FAD binding region of PHBH with the FAD binding loop of UbiI in red, in cartoon representation. (B) Cartoon representation of 3D structure of UbiI AF2 model (in red) and VisC structure (in yellow) with the highlighting of the C-ter region.

Table 2. AF2 Scores for Each Input MSA by Dimer, Gray Rows Are the Very High Confidence Dimers, Bold Ones Correspond to the High Confidence Dimers, and the Italic Ones Correspond to the Medium Confidence Dimers

Ubi Pairs/MSA	Alphafold2 MSA		MMSeqs MSA		Customized MSA	
	Mean/STD	Max	Mean/STD	Max	Mean/STD	Max
UbiE-UbiE	0.41 +/- 0.10	0.60	0.42 +/- 0.16	0.77	0.25 +/- 0.02	0.29
UbiE-UbiF	0.32 +/- 0.06	0.42	0.29 +/- 0.02	0.35	0.26 +/- 0.02	0.30
UbiE-UbiG	0.81 +/- 0.01	0.82	0.80 +/- 0.01	0.82	0.84 +/- 0.01	0.85
UbiE-UbiH	0.29 +/-0.08	0.54	0.25 +/-0.03	0.36	0.21 +/- 0.01	0.22
UbiE-UbiI	0.82 +/- 0.01	0.83	0.81 +/- 0.01	0.83	0.70 +/- 0.08	0.83
UbiE-UbiJ	0.77 +/- 0.02	0.80	0.80 +/- 0.01	0.83	0.77 +/- 0.04	0.82
UbiE-UbiK	0.65 +/- 0.01	0.66	0.64 +/- 0.02	0.69	**0.71 +/- 0.04**	0.80
UbiF-UbiF	0.35 +/- 0.10	0.69	0.41 +/- 0.13	*0.74*	0.49 +/- 0.20	*0.84*
UbiF-UbiG	0.26 +/- 0.01	0.29	0.26 +/- 0.01	0.28	0.23 +/- 0.02	0.27
UbiF-UbiH	0.26 +/- 0.04	0.35	0.25 +/- 0.04	0.36	0.25 +/- 0.02	0.29
UbiF-UbiI	**0.74 +/- 0.03**	0.79	0.53 +/- 0.13	*0.73*	**0.78 +/- 0.03**	0.84
UbiF-UbiJ	0.25 +/- 0.01	0.28	0.24 +/- 0.01	0.27	0.23 +/- 0.01	0.25
UbiF-UbiK	0.23 +/- 0.01	0.24	0.22 +/- 0.01	0.24	0.25 +/- 0.03	0.29
UbiG-UbiG	0.33 +/- 0.09	*0.71*	0.23 +/- 0.04	0.31	0.22 +/- 0.10	0.45
UbiG-UbiH	0.81 +/- 0.01	0.82	0.80 +/- 0.01	0.83	0.81 +/- 0.01	0.84
UbiG-UbiI	0.26 +/- 0.07	0.44	0.27 +/- 0.08	0.54	0.22 +/- 0.02	0.27
UbiG-UbiJ	0.21 +/- 0.01	0.24	0.22 +/- 0.02	0.25	0.19 +/- 0.01	0.28
UbiG-UbiK	0.24 +/- 0.02	0.28	0.22 +/- 0.03	0.33	0.24 +/- 0.11	0.53
UbiH-UbiH	0.37 +/- 0.10	0.59	0.34 +/- 0.08	0.57	0.29 +/-0.05	0.52
UbiH-UbiI	0.33 +/- 0.13	0.68	0.24 +/- 0.02	0.26	0.20 +/- 0.01	0.24
UbiH-UbiJ	0.33 +/- 0.06	0.47	0.35 +/- 0.04	0.43	0.67 +/- 0.06	*0.77*
UbiH-UbiK	0.30 +/- 0.02	0.33	0.26 +/- 0.02	0.30	0.27 +/- 0.03	0.33
UbiI-UbiI	0.34 +/- 0.06	0.44	0.45 +/- 0.16	*0.78*	0.41 +/- 0.14	0.69
UbiI-UbiJ	**0.73 +/- 0.03**	0.77	**0.76 +/- 0.04**	0.81	0.67 +/- 0.08	*0.79*
UbiI-UbiK	0.26 +/- 0.06	0.21	0.24 +/- 0.02	0.29	0.27 +/- 0.02	0.31
UbiJ-UbiJ	0.21 +/- 0.07	0.48	0.20 +/- 0.06	0.35	0.15 +/- 0.04	0.18
UbiJ-UbiK	0.37 +/- 0.04	0.45	0.37 +/- 0.04	0.45	0.42 +/- 0.02	0.44
UbiK-UbiK	0.67 +/- 0.02	0.69	0.67 +/- 0.04	*0.72*	0.43 +/- 0.04	0.53

reliability of results as AF2 is an MSA-based deep learning method, and the Evoformer block is crucial for the accuracy.

Taken together, these approaches enabled us to generate 2025 dimeric models (27 pairs × 25 AF2 models × 3 MSA approaches), including 525 homodimers and 1500 heterodimers. In order to discriminate the most probable interfaces, the AF2 score was used to filter solutions. This latter is a metric based on two scores, which are pTM and ipTM, that are indicators of the monomer and the interface quality, respectively. The final score, ranging from 0 to 1, is weighted according to the following formula: $0.8 \times$ ipTM $+ 0.2 \times$ pTM. Thus, the higher AF2 score, the better the confidence is. The results are summarized in Table 2, and all values are plotted for each pair in Figure S2. It is worth to note that the distribution of AF2 scores is important for convergence analysis and it helps for the confidence on the interface selection. The first criterion for filtering is based on a threshold value of 0.7, related to a sufficient score according to DockQ distribution related to AF2 score,[25] and a low standard deviation (<0.1). According to these conditions, two categories were defined: (i)

very high confidence when the criteria are observed for the three MSA approaches (gray rows in Table 2) and (ii) high confidence when at least one approach is concerned (bold rows in Table 2). As a result, four pairs with very high confidence (UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiG−UbiH) and three pairs with high confidence (UbiE−UbiK, UbiF−UbiI, and UbiI−UbiJ) were selected. Additionally, five pairs (UbiE−UbiE, UbiF−UbiF, UbiG−UbiG, UbiH−UbiJ, UbiI−UbiI, UbiK−UbiK) were also picked because at least one best model among the three approaches had an AF2 value higher than 0.7 (italic row in Table 2).

In total, 13 pairs were selected for further analyses. First, structural analyses were performed by computing TM-score with the MM-Align program.[44] TM-score is a metric, ranging from 0 to 1, for assessing the topological similarity of the protein structures and it is less sensitive than rmsd. Of note, a similar strategy has been already used in the literature to verify the structural convergence of the 3D models.[27,64,65] Indeed, it is not guaranteed that the convergence of AF2 score is associated with only one similar binding pose. Moreover, different AF2 scores could be related to similar binding poses, but displaying slight differences. TM-score values of all 3D models constructed for all pairs of proteins are summarized (mean, standard deviation, and maximal value) in Table 3, and the TM-score distribution is plotted in Figure S3. A threshold of 0.8 and a standard deviation lower than 0.1 have been used to identify similar interfaces among the different 3D models. Using these criteria, 8 pairs (UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiE−UbiK, UbiG−UbiH, UbiH−UbiJ, UbiI−UbiI, UbiI−UbiJ) were found out of all MSAs (gray rows in Table 3), indicating a strongly conserved binding mode. Regarding other pairs, two of them (UbiF−UbiI and UbiK−UbiK) showed a slight divergence for one MSA (bold rows in Table 3, MMSeqs MSA for UbiF−UbiI and our customized MSA for UbiK−UbiK), whereas TM-score values for UbiE−UbiE, UbiF−UbiF, and UbiG−UbiG were significantly lower than selection criteria. Results for the pair UbiK−UbiK can be explained by the disordered N-ter region (low plDDT score values), as shown and discussed in our previous work.[19] The lower confidence found when using our customized MSA could result from the stringency of the data set. To better understand results of other pairs, visual analysis was performed and compared to TM-score and AF2 metrics.

The UbiF−UbiI pair is an interesting example showing a widespread distribution of AF2 score using MMSeqs MSA (Figure S2B), whereas a bimodal distribution is observed for TM-score (Figure S3B). The visual inspection of AF2 models showed two clusters of binding poses exemplified in Figure 6. The first one (Figure 6A) is the most represented, whereas the second one (Figure 6B) was observed only twice with lower AF2 score values of 0.26 and 0.27. Moreover, the first cluster was similar to the ones found for other MSAs (AF2 and customized MSAs) with higher AF2 scores. Altogether, the results point out the importance of MSAs in the calculation of the AF2 score. Indeed, for similar poses but using different MSAs, this example shows a heterogeneity in the AF2 scores. However, the UbiF−I dimer seems to be consistent due to the high confidence from two MSAs over three and should be considered for further analyses. Regarding the last three Ubi pairs, UbiE−E, UbiF−F, and UbiG−G, they showed no convergence on TM score for any MSA method, but at least one best model was found to have a high AF2 score (>0.7) (Table 2). To better understand these results, the best model

Table 3. TM-Scores Range for Each Input MSAs by Pair, Convergence between all the Pairs for a Used MSA in Gray, Pairs Found Convergence for at Least One MSA in Bold

| Ubi Pairs/MSA	AlphaFold2	MMSeqs	Customized
UbiE-UbiE	0.71 +/- 0.20	0.67 +/- 0.14	0.68 +/- 0.17
UbiE-UbiG	0.99 +/- 0.00	0.99 +/- 0.00	0.99 +/- 0.01
UbiE-UbiI	1.00 +/- 0.00	1.00 +/- 0.00	0.99 +/- 0.03
UbiE-UbiJ	0.88 +/- 0.03	0.87 +/- 0.04	0.92 +/- 0.04
UbiE-UbiK	0.92 +/- 0.05	0.93 +/- 0.05	0.98 +/- 0.01
UbiF-UbiF	0.76 +/- 0.14	0.72 +/- 0.15	0.90 +/- 0.10
UbiF-UbiI	**1.00 +/- 0.00**	0.92 +/- 0.16	**1.00 +/- 0.00**
UbiG-UbiG	0.86 +/- 0.14	0.71 +/- 0.14	0.53 +/- 0.13
UbiG-UbiH	1.00 +/- 0.00	1.00 +/- 0.00	0.99 +/- 0.00
UbiH-UbiJ	0.94 +/- 0.03	0.96 +/- 0.02	0.86 +/- 0.08
UbiI-UbiI	0.88 +/- 0.10	0.94 +/- 0.05	0.97 +/- 0.02
UbiI-UbiJ	0.95 +/- 0.02	0.94 +/- 0.02	0.94 +/- 0.02
UbiK-UbiK	**0.94 +/- 0.10**	**0.96 +/- 0.07**	0.65 +/- 0.12
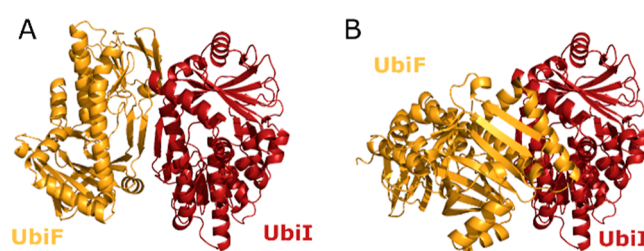
Figure 6. Cartoon representation of both predicted conformations of the dimer UbiF−UbiI, with UbiF in orange and UbiI in red. The representative models associated (A) to cluster 1 (AF2 score of 0.73) and (B) to cluster 2 (AF2 score of 0.27).

of each pair was visually inspected. For the pairs UbiF−F and UbiG−G, it results that AF2 predicted nonreliable 3D structures with strong overlaps between the two partners as shown in Figure S4. For this reason, these pairs were not further considered for the modeling of the full assembly. For UbiE, the best 3D model appeared structurally consistent (Figure S5); however, caution is required due to the nonconvergence of AF2 and TM score metrics.

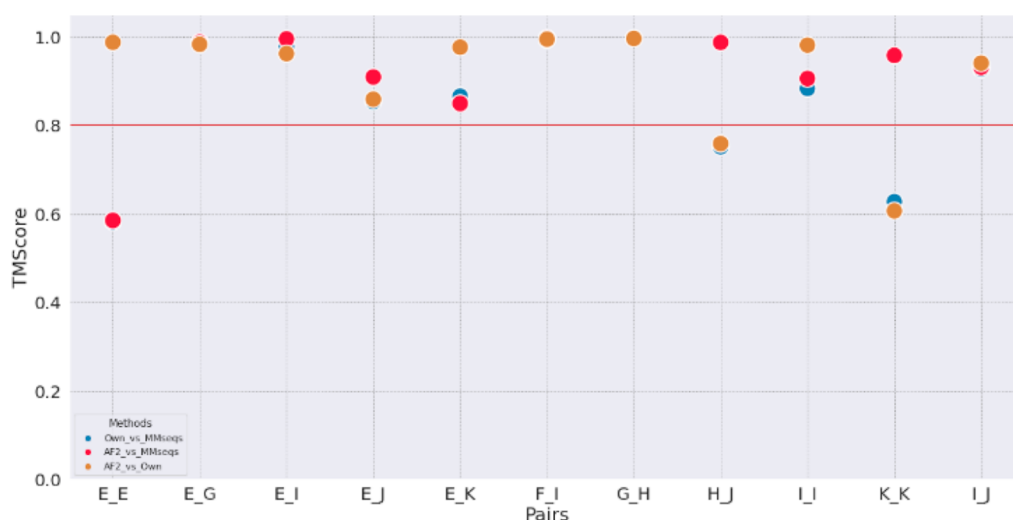Finally, 11 pairs (UbiE−UbiE, UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiE−UbiK, UbiF−UbiI, UbiG−UbiH, UbiH−

Figure 7. Distribution of TM-score between the best models for each input MSA. The blue circle represents the TM-score between the best model using our customized MSA and the one using MMSeqs MSA, the red circle represents the TM-score between the best models using AF2MSA and the one using MMSeqs MSA, and the orange circle represents the TM-score between the best models using AF2MSA and the one using our customized MSA.

UBiJ, UbiI−UbiI, UbiI−UbiJ, and UbiK−UbiK) were selected as potential biological interfaces. To validate them, a structural convergence analysis was performed based on the TM score values of the best 3D models from each MSA (Figure 7). The comparison of TM score values between best 3D models from each method shows that eight pairs (UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiE−UbiK, UbiF−UbiI, UbiG−UbiH, UbiI−UbiI, and UbiI−UbiJ) were predicted with a high structural similarity (TM score >0.8) independently of the MSA input. On the other hand, three other pairs (UbiE−UbiE, UbiH−UbiJ, and UbiK−UbiK) showed that one approach diverged (TM score <0.8) compared to two others, e.g., the best 3D model from our customized MSA for UbiK−UbiK and from MMseqs for UbiE−UbiE and UbiH−UbiJ. Regarding the pair UbiK−UbiK, the structural diversity resides in the N-ter (Figure S6A). For the pair UbiE−UbiE, the best model generated using MMseqs is a different binding pose (no common residue pairs) compared to the two others (Figure S6B), leading to an uncertainty of the pose, which depends on the MSA used as input. Finally, the last pair UbiH−UbiJ shows a structural diversity, especially in the orientation of the SCP2 domain of UbiJ, but the interface residues remain similar (between residues 145 to 180 for UbiJ). However, this region was shown to be flexible in our previous work,[19] and the weak interaction with UbiH could be due to transient interactions.

Next, eight pairs were chosen for further analyses, i.e., UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiE−UbiK, UbiF−UbiI, UbiG−UbiH, UbiI−UbiI, and UbiI−UbiJ. As UbiK−UbiK has been validated biochemically,[20] this pair will not be considered for the following investigation. To validate the seven other pairs, the average predicted Alignment Error (PAE) of interchain residue pairs (PAE interaction)[66] was computed for all 3D best models (Table 4). This metric was used to discriminate interfaces between two protein chains. Indeed, low PAE interaction values may indicate a good confidence on the interface. All PAE interaction values are given in Table 4. An arbitrary threshold of 5 Å was considered to filter the most probable interfaces. Interestingly, all but two pairs (UbiE−UbiK and UbiI−UbiI for AF2 MSA) showed PAE interaction values higher than 5 Å (gray rows in Table 4).

Table 4. Mean PAE for Each Input MSA Best Model, Gray Rows Highlight Pairs with Mean PAE Superior to 5 Å for One MSA, at Least

Ubi Pairs/MSA	AlphaFold2	MMSeqs	Customized
UbiE-UbiG	2.06	1.84	1.80
UbiE-UbiI	2.50	2.44	2.62
UbiE-UbiJ	4.50	4.62	2.86
UbiE-UbiK	10.08	2.60	4.07
UbiF-UbiI	3.22	4.97	2.58
UbiG-UbiH	2.22	2.43	2.20
UbiI-UbiI	14.18	3.71	4.88
UbiI-UbiJ	3.41	2.70	2.56

For the UbiE−UbiK and UbiI−UbiI pairs, the high PAE interaction value could be explained by low AF2 score values, 0.66 and 0.44, respectively. However, both metrics do not seem to be correlated. Indeed, two similar 3D models (TM score > 0.8) for the pair UbiE−UbiK show similar AF2 score values (0.66 for AF2 MSA and 0.69 for MMSeqs) but different PAE interaction values (10.08 for AF2 MSA and 2.60 for MMSeqs MSA). This example highlights the sensitivity of the metrics and indicates that the pairs UbiE−UbiK and UbiI−UbiI should be considered with caution.

To investigate potential oligomerization (trimer assembly) of these last eight pairs, two types of analyses were performed, (i) pairwise structural alignment and (ii) AF2 prediction using MMseqs (localcolabfold). The first strategy consisted of

aligning structurally two pairs having one common enzyme, UbiE−UbiG with UbiE−UbiI for example. This was done for all combinations, and steric clashes were only observed for the pair E−I when superposing with UbiE−UbiK or UbiE−UbiJ (see Figures S6 C,D). As discussed, the subunits UbiK and UbiJ are the most challenging to predict with other partners due to their likely flexibility and their transient interactions. Indeed, when modeling the interaction of UbiE with only the SCP2 domain of UbiJ (main interactions for the full system), all models from localcolabfold show high confidence scores ranging from 0.87 to 0.89. However, modeling the complex UbiE−UbiJ−UbiK$_2$ (more realistic system) led to low AF2 scores (between 0.48 and 0.58). This result highlights several points: (i) the low confidence on the interactions between UbiE and UbiJ (mostly on C-ter) or UbiK and UbiE, (ii) the influence of the high signal of UbiJ−UbiK$_2$ in the modeling of a larger complex, and (iii) only the pair UbiE−UbiI appears probable, with a potential transient interaction with the SCP2 domain of UbiJ.

It is worth noting that localcolabfold is a good compromise between computational efficiency and accuracy (intermediate size of data set) compared with the two other methods. For these reasons, it will be used as default for all further analyses. Therefore, in order to investigate other potential binding sites of interest to elucidate stoichiometry, AF2 models were generated by duplicating one of the subunits for each following pair, UbiE−UbiG, UbiE−UbiI, UbiE−UbiJ, UbiF−UbiI, UbiG−UbiH, and UbiI−UbiJ. Altogether, 3D models of 12 trimers (UbiE−UbiE−UbiG, UbiE−UbiG−UbiG, UbiE−UbiE−UbiI, UbiE−UbiI−UbiI, ...) were generated and their AF2 score values are given in Figure S7. Results show that all trimer models have AF2 score values lower than 0.7, indicating a low confidence. Therefore, at this stage, no potential nucleus could be found to start exploring stoichiometry. Moreover, all trimer models that include two copies of UbiI led to the loss of homodimer interactions. This suggests that the highest AF2 score of Ubi (0.78 for MMseqs in Table 2) could be an artifact and the pair UbiI−UbiI might not be possible in the presence of other partners.

To better understand the confidence on the five last pairs, UbiE−UbiG, UbiE−UbiI, UbiG−UbiH, UbiI−UbiJ, and UbiF−UbiI, visual inspection of the 3D models enabled us to localize the most challenging regions (low local plDDT) identified by AF2. Indeed, this analysis is important to avoid ambiguous regions to be involved in the selected interfaces. Based on the analysis of the monomeric structures, ambiguities were mainly observed in the C-ter for all hydroxylases (UbiF, UbiH, and UbiI) and the residues from 166 to 182 for UbiG. Regarding UbiJ and UbiK, as described in our previous work,[19] the most challenging regions are the residues from 129 to 165 for UbiJ and the N-ter (residues from 1 to 42) for UbiK. Given the known UbiJ−UbiK interaction, all interactions involving the coiled-coil region of one of the proteins individually should be taken with caution. For all pairs, with the exception of I-J, interfaces were found to contain these challenging regions (see Figure S8). Indeed, the low confidence region 129−165 of UbiJ was involved in the interaction with UbiI. Considering these results, four pairs, UbiE−UbiG, UbiE−UbiI, UbiG−UbiH, and UbiF−UbiI, appear to have been modeled with good confidence and will thus be considered as the core for the subassembly exploration. Other heterodimeric pairs such as UbiI−UbiJ and UbiE−UbiJ, but also potential homodimeric UbiI−UbiI and UbiE−UbiE, will be considered with caution

due to their ambiguity, especially UbiI−UbiI and UbiE−UbiE, which are not predicted in the context of trimers. Finally, all enzymatic subunits (UbiI,H,F and UbiG,E) of the metabolon appear to have at least one partner predicted with a good confidence, which is important for the reconstruction of a functional assembly.

3.3. Prediction of Multimeric Assemblies. Pairwise analyses enabled us to scan all interfaces, without any knowledge on the complex stoichiometry or potential interactions. Thus, the next stage is to use this information in order to build a potential interaction network as depicted on Figure 8.
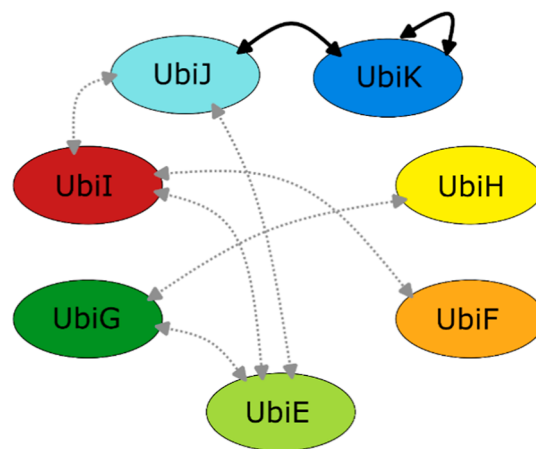

Figure 8. Network representation of the interactions between the different protein partners involved in the Ubi metabolon. Dotted lines represent predicted interactions and continuous lines show the interactions evidenced by experimental means.

This latter is key for a starting point to identify the molecular bricks and reconstruct the full and biologically functional assembly. Modeling based on each Ubi protein and its potential partners, from the interaction network depicted in Figure 8, was carried out in stepwise manner. The aim was to validate the stability of subcomplexes (maintain the identified interfaces, no binding site competition, ...) from confident pairs according to AF2. This analysis was performed on each Ubi subunit having more than one interaction, i.e., UbiE, UbiG, UbiI, UbiJ, and UbiK (see Figure 8). So, four subcomplexes (UbiE−UbiI−UbiJ−UbiG, UbiG−UbiE−UbiH, UbiI−UbiJ−UbiF−UbiE, and UbiJ−UbiK−UbiK−UbiE−UbiI) were modeled using only localcolabfold (MMSeqs MSA), to take advantage of its computational efficiency and accuracy as discussed earlier.

Regarding the UbiE central complex (UbiE−UbiI−UbiJ−UbiG), the best 3D model has an AF2 score value of 0.75, which is in the range of good confidence (Figure 9A). Interestingly, the tripartite UbiE−UbiG−UbiI was still modeled as interacting together. To verify the conservation of interfaces compared to the pairwise modeling, some analyses were performed. The superimposition of 3D models between the subcomplex (UbiE−UbiI−UbiG−UbiJ) and the pairs UbiE-UbiI (Figure 9B) or UbiE−UbiG (Figure 9C) showed a fairly good structural alignment. This observation was confirmed by metrics such as the TM-score and the interface area. Indeed, TM-score values were 1.00 and 0.99 for UbiE−UbiI and UbiE−UbiG against UbiE−UbiI−UbiG−UbiJ, respectively. Interface area values were very close when
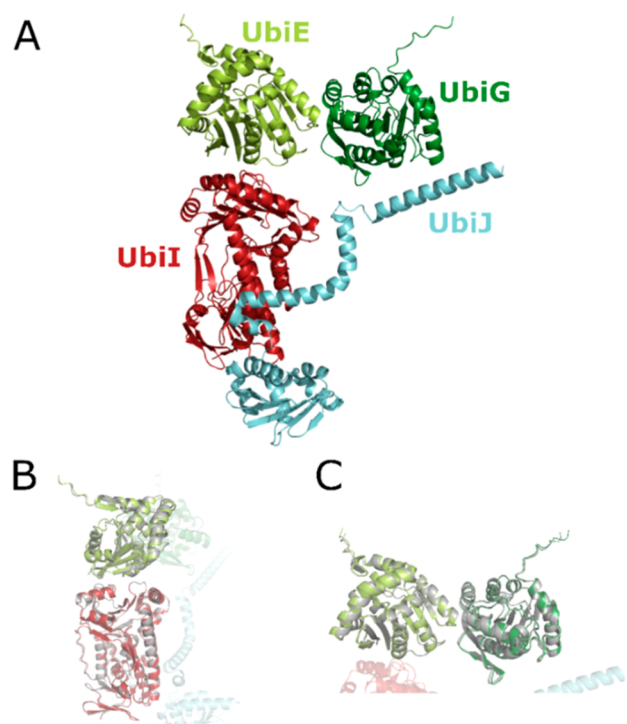
Figure 9. Cartoon representations of predicted 3D models of UbiE−UbiI−UbiJ−UbiG. (A) First conformation. (B) Zoom on the UbiI−UbiE interaction with the superimposition of the MMSeqs dimer (in gray). (C) Zoom on the UbiE−UbiG interaction with the superimposition of the MMSeqs dimer (in gray).

comparing UbiE−UbiI (667 Å^2) or UbiE−UbiG (528 Å^2) against the oligomer form, 660 and 522 Å^2, for UbiE−UbiI and UbiE−UbiG, respectively. These results indicate that the prediction of the interfaces UbiE−UbiI and UbiE−UbiG are stable individually or in complex. Regarding the interactions between UbiI and UbiJ, the same observation can be made with slight differences in the interface area (812 Å^2 for dimer form versus 868 Å^2 for oligomeric form), whereas the TM score value is 0.91. This slight difference could be explained by

the orientation of the region 129−165 of UbiJ, already described to be difficult to predict (low plDDT) due to its flexibility. Other 3D models were checked, among them, a 3D model involved UbiJ positioned between UbiE and UbiI as shown in Figure S9. This 3D model has a lower AF2 score (0.63) than the one with an interaction UbiE−UbiI; however, it remains difficult with AF2 to distinguish between the interactions UbiE−J and UbiE−I. This could be due to the potential role of UbiJ, which is thought to bind the lipophilic tail of UQ$_8$ biosynthetic intermediates and to present the headgroup to the five enzymes (UbiI,H,F and UbiE,G) for chemical modification.

The second modeled subcomplex corresponds to the UbiG center complex (UbiE−UbiG−UbiH). All 3D models were predicted with good confidence showing AF2 score values between 0.73 and 0.81. As previously, the conservation of interfaces in the subcomplex UbiE−UbiG−UbiH was investigated compared to pairs UbiE−UbiG and UbiE−UbiH. The best 3D model is displayed in Figure 10A, showing a central role of UbiG in the interaction with UbiE and UbiH. Structural superimposition between individual pairs (UbiE−UbiG Figure 10B or UbiE−UbiH Figure 10C) and the subcomplex UbiE−UbiG−UbiH was performed. The good structural alignment was validated by TM score values of 0.99 for both pairs UbiE−UbiG and UbiE−UbiH against UbiE−UbiG−UbiH. Slight differences were observed in the calculation of interface areas for the pair UbiE−UbiG, with 528 Å^2 in the dimeric form and 546 Å^2 in the oligomeric form, whereas the pair UbiG−UbiH showed larger differences, with 539 Å^2 in the dimer form against 643 Å^2 in the oligomeric one. This discrepancy could be explained by a different modeling and side chain packing in the two 3D models. However, both interfaces UbiE−UbiG and UbiG−UbiH are similar within dimers or oligomers, emphasizing the confidence on the interactions and the organization of the subcomplex UbiE−UbiG−UbiH.

The modeling of the third subcomplex, i.e., the UbiI central complex (UbiI−UbiJ−UbiF−UbiE), led to the generation of 3D models with good confidence values ranging from 0.71 to 0.78. The best 3D model was further analyzed and showed an unexpected interaction involving UbiE and UbiF (see Figure
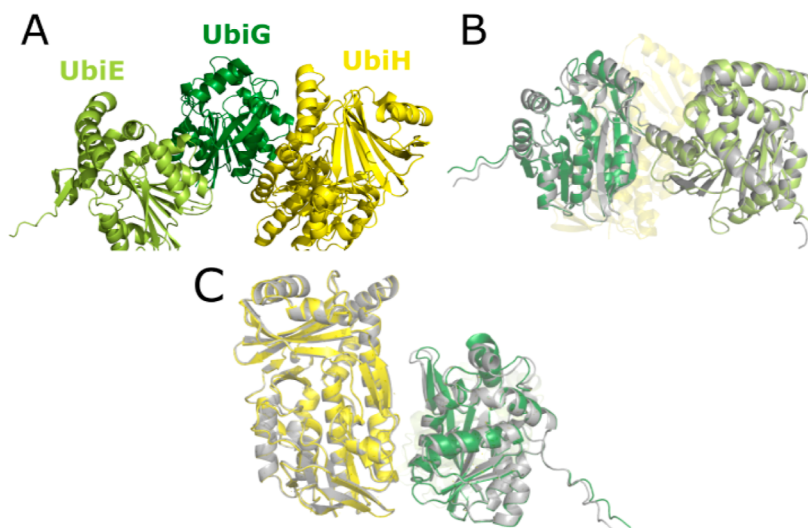

Figure 10. Cartoon representations of UbiE−UbiG−UbiH. (A) Global structure. (B) Zoom on UbiE−UbiG with the superimposition of the MMSeqs dimer (in gray). (C) Zoom on UbiG−UbiH with the superimposition of the MMSeqs dimer. UbiE is in light green, UbiG in forest green, and UbiH in yellow.
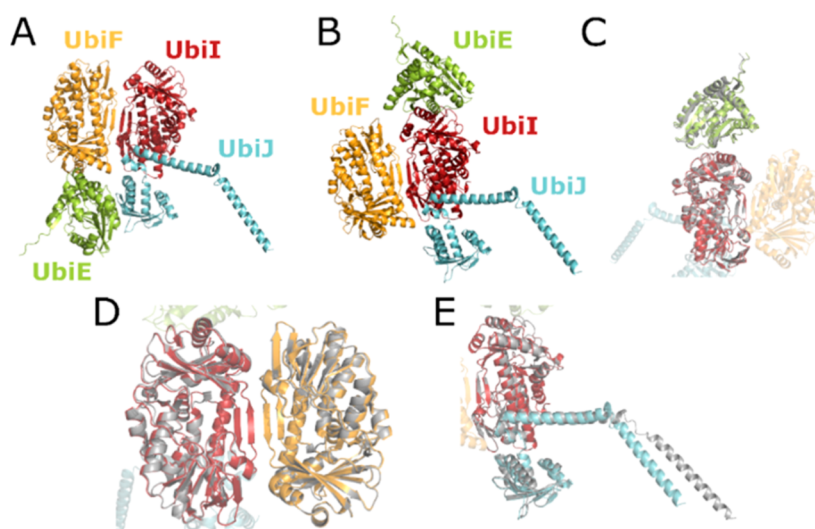
Figure 11. Cartoon representations of UbiI−UbiJ−UbiF−UbiE. (A) First conformation. (B) Second conformation. (C) Zoom on the UbiI−UbiE interaction with the superimposition of the MMSeqs dimer (in gray). (D) Zoom on the UbiI−UbiF interaction with the superimposition of the MMSeqs dimer (in gray). (E) Zoom on the UbiI−UbiJ interaction with the superimposition of the MMSeqs dimer (in gray). UbiE is in light green, UbiI in red, UbiF in orange, and UbiJ in cyan.
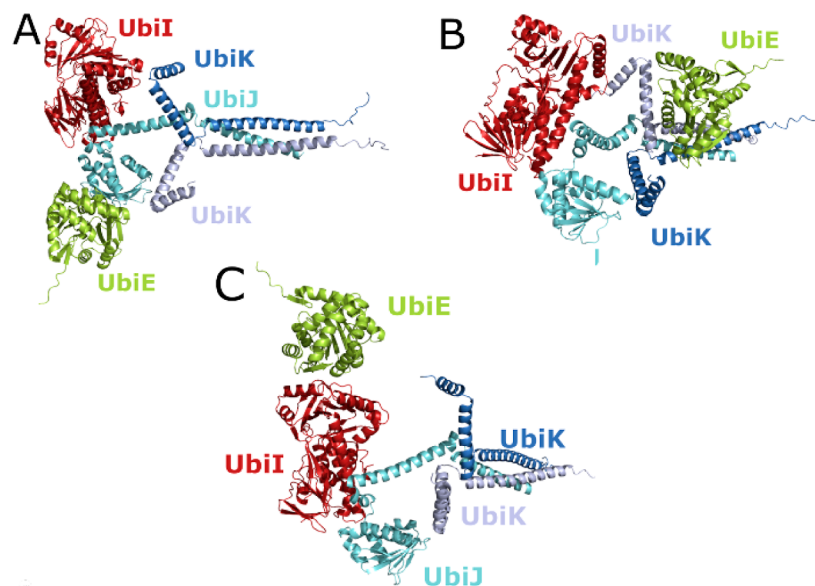

Figure 12. Cartoon representations of the 3D model predicted for UbiE−UbiI−UbiJ−UbiK$_2$. (A) Conformation of cluster 1 (best AF2 score 0.65). (B) Conformation of cluster 2 (best AF2 score 0.58). (C) Conformation of cluster 3 (best AF2 score 0.46). UbiI is in red, UbiE in light green, UbiJ in cyan, and UbiK in light and dark blue.

11A). Indeed, the pair UbiE−UbiF was not considered due to low AF2 scores for any MSA input (see Table 2). Among other predicted 3D models, 14 of them displayed an interaction between UbiE and UbiI (Figure 11B), whereas 10 of them were similar to the best 3D model (UbiE−UbiF interaction). Nevertheless, this latter (UbiE−UbiF) seems to be weak, with interface values varying from 335 to 407 Å^2, which might be driven by the presence of UbiJ which bridges the partners as shown in Figure 11A. In order to investigate these conformational differences, additional analyses were performed, (i) modeling the subcomplex without UbiJ (UbiI−UbiF−UbiE−UbiE) and (ii) modeling the subcomplex (UbiI−UbiJ−UbiF−UbiE) using our customized MSA. The first strategy should provide some insights into the role of UbiJ in the 3D model, whereas the second one should reduce the noise between UbiI

and UbiF. Modeling of the subcomplex UbiI−UbiF−UbiE−UbiE led to generation of 3D models with low AF2 confidence values ranging from 0.46 to 0.55. The analysis of UbiE−UbiF interaction showed a variation in the interface area values from 0 to 615 Å^2, indicating the uncertainty of AF2 in the prediction. Regarding the modeling of UbiI−UbiJ−UbiF−UbiE using our customized MSA as input, all 3D models showed a good AF2 confidence varying from 0.72 to 0.77. Interestingly, no interaction between UbiE and UbiF was found in all 3D models, but two clusters were identified: (i) one cluster containing UbiI−UbiE interaction and (ii) one cluster containing the UbiE−UbiJ interaction. This result is consistent with its potential role. These results highlight that the pair UbiE−UbiI is more probable compared to UbiE−UbiF. Thus, the 3D model of the subcomplex UbiI−UbiJ−
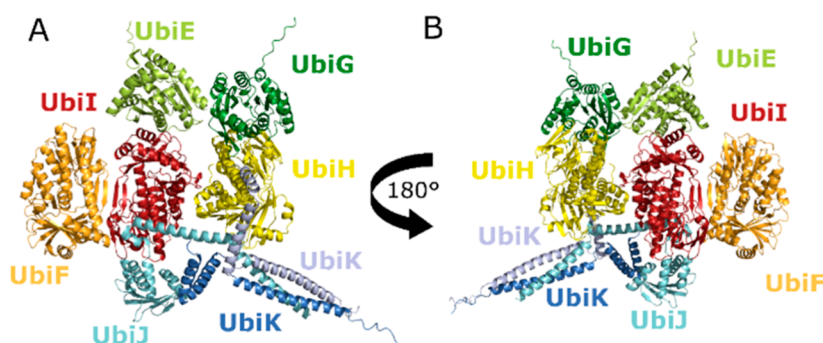
Figure 13. Cartoon representations of the core subunit, with UbiE in light green, UbiG in dark green, UbiF in orange, UbiH in yellow, UbiI in red, UbiJ in cyan, and UbiK in light and dark blue.

UbiE−UbiF containing the UbiI−UbiE interaction (Figure 11B) was considered as more probable. Moreover, it corresponds to the second-best score with an AF2 value of 0.76 (results from localcolabfold using MMseqs). As for previous subcomplexes, the conservation of the interfaces was analyzed (Figure 11C−E). TM-score values were found to be 0.99 for UbiE−UbiI, 1.00 for UbiI−UbiF and 0.93 for UbiI−UbiJ. Regarding the interface area values, the results showed similar interfaces between dimeric and oligomeric forms, i.e., 633 Å^2 in oligomers against 667 Å^2 in dimers for UbiE−I, 677 Å^2 in oligomers against 669 Å^2 in dimers for UbiI−UbiF, and 815 Å^2 in oligomers against 812 Å^2 in dimers for UbiI−UbiJ.

The last subcomplex investigated corresponds to the UbiJ center complex (UbiJ−UbiK−UbiK−UbiE−UbiI). Strong signals are found by AF2 to create interactions between UbiE−UbiJ and/or UbiI−UbiJ. Moreover, adding two copies of UbiK can add some noise. Therefore, it is not surprising to find some heterogeneity in the 3D models. After clustering, three types of subcomplexes modeled by AF2 were observed (Figure 12). The first one corresponds to a scenario already viewed in the subcomplex UbiE−UbiI−UbiJ−UbiG, i.e., UbiJ is interacting with both UbiE and UbiI (Figure 12A). The second one corresponds to a new interaction involving UbiK with UbiE (Figure 12B). The last one shows UbiI interacting with UbiE on one side and UbiJ on the other side (Figure 12C). For all 3D models, the AF2 score values were relatively low (between 0.46 and 0.65). This result could be explained by the noise of the heterotrimer UbiJ−UbiK$_2$. Indeed, modeling the subcomplex without UbiK led to higher scores (from 0.58 to 0.77), and only clusters 1 and 3 were modeled.

These results highlight the complexity for AF2 to model correctly the interactions with UbiJ. Nevertheless, all 3D models including UbiK showed the heterotrimer UbiJ−UbiK$_2$, which is promising given its key role in the structure.[19,20] To conclude, the most probable 3D model for this subcomplex seems to be cluster 3 (Figure 12C), which is consistent in terms of organization with the subcomplex UbiE−UbiI−UbiJ−UbiG.

Based on these results, four potential bricks (subcomplexes) were identified as possible parts of the full assembly, i.e., UbiE−UbiI−UbiJ−UbiG (1), UbiG−UbiH−UbiE (2), UbiI−UbiJ−UbiF−UbiE (3), and UbiJ−UbiE−UbiI−UbiK$_2$ (4). Interestingly, all these molecular bricks share at least one Ubi subunit, which is useful to connect each brick by superposition of the common subunit. Even if the heterotrimer UbiJ−UbiK$_2$ is the only subcomplex evidenced experimentally, addition of these subunits to the molecular assembly generally decreased the AF2 score. Therefore, to connect the different subcomplexes, incremental AF2 modeling was performed in the presence or not of UbiJ to evaluate the influence of the structural subunit. The targeted complexes are thus the following: UbiE−UbiI−UbiG−UbiH(-UbiJ), UbiE−UbiI−UbiG−UbiF(-UbiJ), UbiE−UbiG−UbiH−UbiI−UbiF(-UbiJ). The first combination allows to assemble the subcomplex 1 with 2, the second one to assemble the subcomplex 1 with 3, and the last one assembles all subcomplexes. Interestingly, all combinations showed good AF2 score values for the best 3D models (>0.7), with AF2 score values ranging from 0.72 to 0.80 for the UbiE−UbiI−UbiG−UbiH subunit, from 0.78 to 0.83 for the UbiE−UbiI−UbiG−UbiH−UbiJ subunit, from 0.62 to 0.74 for the E−I−G−F subunit, from 0.62 to 0.78 for the UbiE−UbiI−UbiG−UbiF−UbiJ subunit, from 0.61 to 0.75 for the UbiE−UbiI−UbiG−UbiH−UbiF subunit, and from 0.79 to 0.83 for the UbiE−UbiI−UbiG−UbiH−UbiF−UbiJ subunit. These results give good confidence on the assemblies. We noticed that UbiF decreased the global AF2 score while UbiJ improved it slightly, especially for the full assembly. The UbiF effect could be explained by the pair UbiI−UbiF, which has lower AF2 scores, (see Table 2), especially when using MMSeqs as input. Regarding UbiJ, the pair UbiI−UbiJ predicted with a high AF2 score, could be responsible for the increase. Based on this observation, a more realistic assembly, called "core assembly", has been modeled, taking into account the heterotrimer UbiJ−UbiK$_2$. Interestingly, AF2 enabled us to predict 3D models of the complete assembly with good confidence (AF2 scores varying from 0.78 to 0.81). The best 3D model is displayed in Figure 13. It is worth noting that the interfaces are conserved among all 3D models of the assembly and also compared with dimer predictions. To better understand these strong signals, the interface conservation at residue level was investigated using Consurf[45] and manually curated MSA as input (see Materials and Methods). The list of conserved residues of each protein chain for each interface is given in Table S1. Surprisingly, only interfaces involving UbiJ showed the highest number of conserved residues, indicating that predicted interfaces of the core subunits seem to be related to (co)-evolutionary signals.

To conclude, this 3D model could be of great interest to better understand the interactions involved between Ubi subunits. In the next sections, we will aim at analyzing in more detail the molecular assembly and its biological/functional role in regard to available experimental data.

3.4. Molecular Insight into the Core Subunit. The final 3D model displayed in Figure 13 provides a molecular insight into the Ubi assembly including one copy of each Ubi subunit. The theoretical molecular weight (MW) of this complex
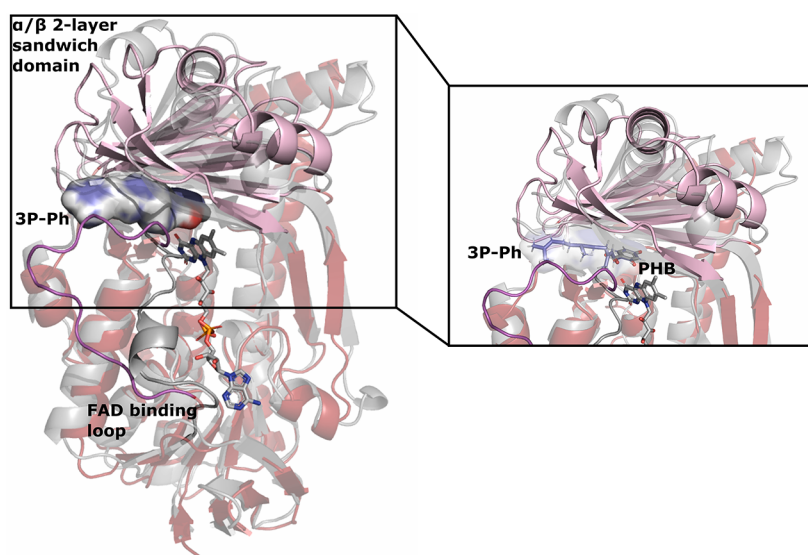
Figure 14. Position and accessibility of 3P-Ph in Ubi FMOs. UbiI (red cartoon) was used as reference for Ubi FMOs, whereas PHBH (PDB Id: 1PBE) was used as reference for class A FMO. FAD, shown in sticks, from 1PBE was placed by superimposition of 1PBE on the best model of UbiI. On the left, the FAD binding loop and the α/β 2-layer sandwich domain are shown in violet and pink, respectively in UbiI. 3P-Ph is shown on the surface to have an insight into the volume of the active site channel. On the right, a zoom of the active site shows the perfectsuperimposition of the aromatic ring (blue) of 3P-Ph on the ring of PHBH (gray).
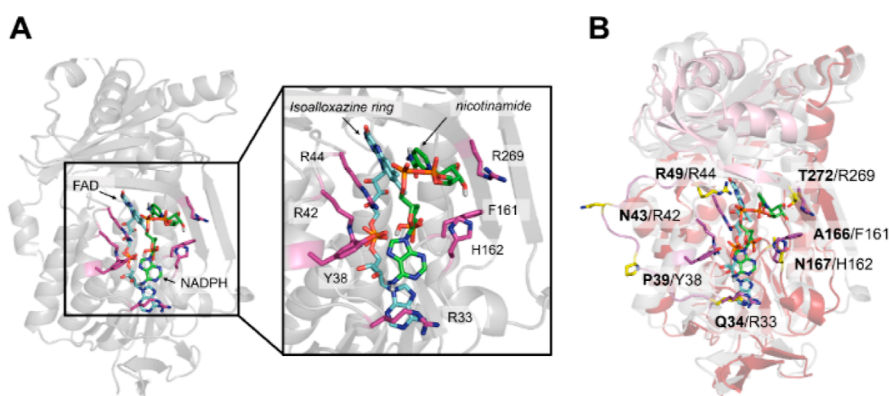

Figure 15. NADPH binding. (A) NADPH docking onto PHBH. The structure of PHBH (PDB Id: 1D0D) is shown in gray cartoon and key residues highlighted in magenta sticks. FAD and NADPH are displayed in cyan and green sticks, respectively. A zoom of NADPH binding is shown on the right. (B) Structural superimposition between PHBH (gray cartoon) and the best AF2 model of UbiI (red-pink-magenta cartoon). The key residues for NADPH binding are shown in yellow (UbiI) and magenta (PHBH) sticks. Bold labels refer to the corresponding residues in UbiI.

subunit would thus be around 230 kDa (UbiE: 28 kDa, UbiG: 27 kDa, UbiF: 43 kDa, UbiH: 42 kDa, UbiI: 44 kDa, UbiJ: 22 kDa, 11 kDa), which differs from the 1 MDa proposed by Chehade et al.[8] The most probable explanation would be the presence of several copies of this complex, which we call the Ubi metabolon core subunit. Indeed, this latter is thought to be the minimal functional assembly for UQ biosynthesis because it includes at least one copy of the essential Ubi subunits (UbiI,H,F,E,G,K,J).

To verify if our core subunit is compatible with function, we performed several analyses that aimed mostly at assessing the accessibility of ligands (cofactors, UQ precursors, ...) for each enzyme. The delivery of the UQ precursors by UbiJ will be discussed but could not be fully investigated due to dynamics and conformational rearrangements that are likely to be involved.

According to the molecular mechanism of class A FMOs described earlier, the three hydroxylases UbiI, H, and F should interact dynamically with three partners corresponding to

FAD, UQ precursors (Figure 1) and NAD(P)H. Only the FAD cofactor is thought to be permanently bound. This latter could be structurally placed by superimposing a reference structure of class A FMOs, which is PHBH (PDB ID: 1PBE) (Figure S10). Indeed, the fold between PHBH and Ubi FMOs (I, H, and F) is very similar, with TM-score values of 0.82, 0.86, and 0.85, respectively. rmsd values of aligned residues vary between 2.6 and 3.2 Å, especially due to the open conformation of the FAD binding loop described previously (Figure S1). Interestingly, PHBH was cocrystallized with its substrate named PHB (P-Hydroxybenzoic acid), which was used to place the aromatic ring in the active site of the Ubi FMOs (Figure 14). 3 farnesyl-phenol (3P-Ph) was used as an example for a visual identification and a length estimation of the potential active site channel. This analysis led to orient the isoprenyl chain of 3P-Ph between the FAD binding loop (residues from E33 to R49 for UbiI) and the CATH domain alpha/beta 2-layer Sandwich (EC 3.30.9.10) composed of residues G80-Y101 and H185−P270 (numbering for UbiI).

This result was confirmed by a molecular docking study of 3-octaprenyl-phenol (OPP) into a relaxed structure of UbiI with FAD using Autodock Vina[46] (Figure S10). Interestingly, one of the binding poses showed an interaction compatible with the molecular reaction, i.e., proximity of C5 from OPP and the carbon (C4X) from FAD. Finally, NADPH binding was the most challenging task due to the lack of structural information. Indeed, only one X-ray structure of PHBH (PDB ID: 1K0J) was solved in complex with NADPH; however, the adenosine group of NADPH was found in interaction with the isoalloxazine ring of FAD, instead of the nicotinamide group of NADPH. Moreover, FAD was solved in *in* conformation, while the *out* conformation seems to be required for NAPDH binding.[13] To tackle this issue, a molecular docking study of NADPH, using Autodock vina, was first performed in the PHBH structure (PDB Id: 1D0D). This latter was cocrystallized with FAD in *out* conformation. The best binding pose (Vina score −9.1 kcal/mol) showed consistent interactions with the molecular mechanism and experimental data (Figure 15A). Indeed, the nicotinamide group of NADPH is in interaction with the isoalloxazine ring of FAD, whereas the adenosine diphosphate moiety of NADPH is bound to residues (R33, Y38, R42, R44, F161, H162, R269) that were proven to be important for NADPH binding in mutagenesis studies.[67,68] To identify corresponding residues in UbiI, structural superimposition was performed and is shown in Figure 15B. It seems that these latter residues are not conserved between PHBH and UbiI, and more generally in Ubi FMOs (UbiI,H and F). This result is consistent with sequence analyses in the literature.[12] Finally, due to the sequence proximity between Ubi FMOs, all these results may be generalized to UbiH and UbiF.

UbiE and UbiG belong to SAM-dependent methyl-transferases that require the entry of the SAM cofactor and the release of the SAH cofactor during the catalytic cycle.[16] The molecular mechanism corresponds to the transfer of the methyl group of SAM on a specific carbon atom (UbiE) or a specific oxygen atoms (UbiG) of the UQ biosynthetic intermediates. For both enzymes, SAH/SAM were cocrystallized in X-ray structures, such as 4KDR[14] and 4OBW,[50] corresponding to *E. coli* UbiG sequence and homologous sequence of *E. coli* UbiE, respectively. To investigate the structural basis for the binding of corresponding UQ precursors, docking and tunnel search assays were performed on both enzymes.

For UbiG, two main paths from docking poses were observed to allow access of the substrate DMeQ8 to the active site, spanning under the key loop including residues from G164 to K190 (Figure 16A). Interestingly, residues from 177 to 186 have been experimentally proven to interact with lipids and participate in the release of the SAH cofactor. Moreover, this region seems to be flexible due to the missing residues in X-ray structures (PDB IDs: 4KDR and 4KDC). Therefore, conformational changes are expected in this region. Finally, tunnel search was investigated using MOLE Web server,[69] which provided several solutions and the most functional ones are shown in Figure 16A. Tunnels 1 and 2 are compatible with the paths identified by the docking approach, and tunnel 2 looks like more favorable in view of docking energies, i.e., ∼−7 versus ∼−4 kcal/mol for the binding poses in the tunnel 1.

Regarding UbiE, the identification of the binding pocket has been done through a comparative analysis with its *S. cerevisae* homologue Coq5. Indeed, both sequences share 38% of
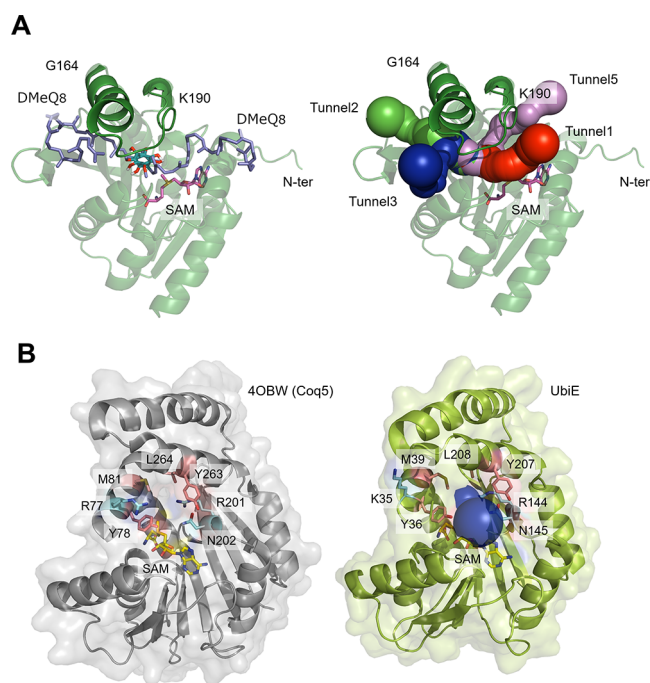

Figure 16. UQ precursor binding in Ubi methyltransferases. (A) Putative binding mode for DMeQ8 into UbiG. The enzyme is shown in forest green cartoon, whereas SAM cofactor and DMeQ8 are displayed in magenta and violet sticks, respectively. Binding poses (on the left) from Vina docking are compared to tunnels found by MOLE Web server (https://mol.upol.cz/)[69] (on the right). The docking poses were selected according to their energy and the positioning of the quinone head close to the methyl group of SAM. (B) Accessible pocket of UbiE for binding DDMQ8. The binding pocket in UbiE (limon cartoon and transparent surface on the right) was compared to that of its Coq5 homologue (gray cartoon and transparent surface on the left). The highlighted residues correspond to the pocket described in the literature for 4OBW.[50] The pocket was confirmed in UbiE using MOLE Web server and is shown in transparent blue balls (on the right).

identity and are structurally similar in terms of fold (TM-score of 0.88), and the docking of a substrate-analogue (DDMQ2H2) in Coq5 has already been described in the literature.[50] In this work, a list of key residues was suggested for interacting with the quinone head, corresponding to R77, Y78, M81, R201, N202, Y263, and L264. Except R77, all these residues were conserved in the MSA, which included the UbiE sequence. R77 was found to be replaced by a Lysine (K) or a Serine (S). In UbiE, the corresponding position is K35. All these key residues were mapped on the 3D structure of 4OBW for Coq5 and the best AF2 model for UbiE, and both structures showed the same topology of the binding pocket (Figure 16B). By analogy, we can assume that the UQ precursor (DDMQ8) should bind to this region. It is important to note that this pocket is accessible only with a displacement of the N-ter region. Indeed, to identify a binding pocket in UbiE comparable to that of 4OBW, the MOLE Web server[69] analysis was performed after removing the first 31 residues; otherwise, the site is hidden by the N-ter region of the AF2 model (Figure S11).

All these results led to the identification of possible binding pockets and their accessibility for all enzymes. Additional analyses, including the dynamics of the system, should be considered for some of them, which is out of the scope of this

study. To get a molecular insight into these key regions, the accessible residues were mapped onto the core-subunit model and are shown in Figure S12. For each site, a list of residues bordering the entry of the pocket is provided. The result is that all entries of potential binding pockets are accessible; however, some conformational rearrangements and flexibility of UbiJ−UbiK$_2$ are expected, in the case of UbiJ taking in charge each UQ intermediate.

Finally, the last point to discuss is the MW discrepancy between the size of the core assembly (230 kDa) and the estimated size of the Ubi metabolon (1 MDa) by Chehade et al.[8] To explain such a difference, one hypothesis could be the multiplicity of the core assembly to reach the correct MW value. If we admit that the different copies interact together, some secondary oligomerization sites have to be investigated. Even if this question was addressed during the pairwise search, the presence of other partners could help AF2 in a better prediction. To tackle this question, AF2 modeling using MMseqs alignment and AF2 score was performed on the core subunit plus one of Ubi subunit. The minimal and maximal AF2 score values are provided in Table 5 for each system.

Table 5. AF2 Scores for Each Added Protein to the EFGHI Subunit Using MMSeqs MSA as Input

Added Protein	EFGHI subunit
UbiE	0.35 - 0.65
UbiF	0.53 - 0.61
UbiG	0.39 - 0.66
UbiH	0.51 - 0.60
UbiI	0.47 - 0.57

These results show that adding any Ubi subunit to the core subunit decreases the confidence on the prediction (reference range of UbiE−UbiF−UbiG−UbiH−UbiI is 0.61−0.75). Moreover, all 3D models showed no interaction between the additional partner and the stable core-subunit. From these results, we investigated some potential homodimeric states from X-ray structures to ensure that we had explored all possibilities. Indeed, we know that the structures of Coq5 and some class A FMO enzymes such as PHBH were solved in homodimeric form. Surprisingly, AF2 was not able or not confident on these pairs, i.e., 0.36 on average for UbiE−UbiE (max value 0.77) and 0.4 on average for UbiI−UbiI (max value 0.81), as shown in Table 2. Regarding the confidence for UbiI, we explain the result by the difficulty to model correctly the C-ter region, which is predicted to be the dimerization region. For UbiE, the interface of Coq5 seems to be conserved inside the enzyme class including UbiE,[50] and so we do not have explanation on the results from AF2 compared to X-ray. To check these potential homodimeric forms as starting points for the extension of the core-subunit, we superimposed first the homodimeric form of PHBH (PDB Id: 1PBE) for Ubi FMOs or the X-ray homodimer of Coq5 (PDB Id: 4OBW) for UbiE, and then, we superimposed the second copy of the core subunit on the other chain of the homodimer. All assays led to severe clashes even if considering some conformational changes, especially for Ubi FMOs (Figure S13). Extending the core subunit from a homodimer UbiE−UbiE led to only clashes between the two copies of UbiG as shown in Figure 17. Altogether, we could not identify satisfying extensions from stable and/or strong interfaces coming from secondary binding sites.

To conclude, the molecular weight of the Ubi metabolon (1 MDa) may result from the association of additional proteins through weak interactions not detectable by AF2 or mediated by the environment, the solvent for example. Of note, the lipoprotein Blc was found to associate with the purified Ubi metabolon, although a functional role for Blc in UQ synthesis was refuted.[8] Moreover, we were limited to static analyses, and dynamics should be considering conformational changes such
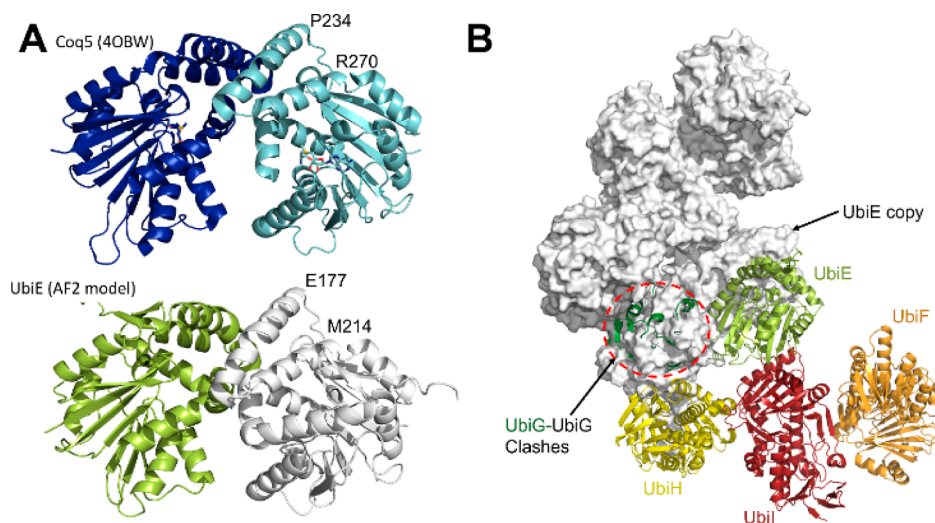

Figure 17. Potential extension from putative UbiE dimer. (A) Use of Coq5 dimer, homolog of UbiE, as a template for putative dimeric form of UbiE from AF2. The residues labeled indicate the region of the two helices participating to the interface. (B) Structural view of the two copies of the core subunit from UbiE dimerization. The steric clash between the two copies of UbiG (one in white surface, the other one in forest green cartoon) is displayed in dashed red circle. Each enzyme is labeled and colored as previously.

as some key loops, but such analyses are challenging without a clear idea of the complete environment. As such, it is worth considering that lipids may be part of the 1 MDa Ubi metabolon as part of UbiG was previously shown to interact with the membrane.[17] In addition, lipids are likely present in complex Q, the eukaryotic counterpart of the Ubi metabolon. Indeed, the structure of the subpart of complex Q composed of Coq7−Coq9 was recently solved with bound lipids.[70]

4. CONCLUSIONS

This computational work aimed to predict a key and very large protein complex in *E. coli*, the Ubi metabolon. This protein assembly is responsible for the biosynthesis of UQ, a molecule essential for the bioenergetics of many species. Nowadays, no molecular insight into the organization of this complex has been determined. Only few interactions were experimentally proven such as UbiJ−UbiK₂. Thanks to the development of AF2 and its multimeric version, we were able to propose a 3D model including one copy of each Ubi subunit (Ubi-G,E,I,H,F,J,K), which we call core assembly. The confidence and the functional potential of the predicted core assembly were validated by a combinatorial approach using AF2 and the investigation of the accessibility of key binding pockets.

Our work provides a tentative stoichiometry for the Ubi proteins within the Ubi metabolon and will guide future experimental validations of this entity, which could be approached by Cryo-EM, native mass spectrometry, or cross-linking.

■ ASSOCIATED CONTENT

Data Availability Statement

All input protein sequences used for our customized MSAs and the resulting final models (AF2 monomers and AF2 metabolon) are available in the following URL: http://wwwperso.insa-toulouse.fr/~esque/metabolon_ubi/.

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c00304.

> Structural information containing additional structural analysis and analyses of 3D models, plDDT and TM-score value distribution, and conserved residues (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Isabelle André − *Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France;* ⓞ orcid.org/0000-0001-6280-4109; Email: isabelle.andre@insa-toulouse.fr

Jérémy Esque − *Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France;* ⓞ orcid.org/0000-0002-5786-848X; Email: esque@insa-toulouse.fr

Authors

Romain Launay − *Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France;* ⓞ orcid.org/0000-0002-9260-8296

Sophie-Carole Chobert − *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France*

Sophie S. Abby − *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France*

Fabien Pierrel − *Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France;* ⓞ orcid.org/0000-0003-2083-3002

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.4c00304

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ ABBREVIATIONS

UQ, ubiquinone; AF2, AlphaFold2; FMO, a flavin mono-Oxygenase; SAH, S-adenosyl-L homocysteine; SAM, S-adénosyl-L-méthionine; MSA, multiple sequence alignment

■ REFERENCES

(1) Abby, S. S.; Kazemzadeh, K.; Vragniau, C.; Pelosi, L.; Pierrel, F. Advances in Bacterial Pathways for the Biosynthesis of Ubiquinone. *Biochim. Biophys. Acta Bioenerg.* **2020**, *1861*, 148259.

(2) Stefely, J. A.; Pagliarini, D. J. Biochemistry of Mitochondrial Coenzyme Q Biosynthesis. *Trends Biochem. Sci.* **2017**, *42*, 824−843.

(3) Bentinger, M.; Tekle, M.; Dallner, G. Coenzyme Q − Biosynthesis and Functions. *Biochim. Biophys. Res. Commun.* **2010**, *396*, 74−79.

(4) Nagata, T. Design of Synthetic Molecular Units Including Quinones towards the Construction of Artificial Photosynthesis. *Res. Chem. Intermed.* **2014**, *40*, 3183−3198.

(5) Meganathan, R. *Biosynthesis of Menaquinone (Vitamin K2) and Ubiquinone (Coenzyme Q): A Perspective on Enzymatic Mechanisms*; Elsevier, 2001; Vol. 61, pp 173−218.

(6) Kawamukai, M. Biosynthesis of Coenzyme Q in Eukaryotes. *Biosci., Biotechnol., Biochem.* **2016**, *80*, 23−33.

(7) Lee, P. C.; Salomon, C.; Mijts, B.; Schmidt-Dannert, C. Biosynthesis of Ubiquinone Compounds with Conjugated Prenyl Side Chains. *Appl. Environ. Microbiol.* **2008**, *74*, 6908−6917.

(8) Chehade, M. H.; Pelosi, L.; Fyfe, C. D.; Loiseau, L.; Rascalou, B.; Brugière, S.; Kazemzadeh, K.; Vo, C.-D.-T.; Ciccone, L.; Aussel, L.; Couté, Y.; Fontecave, M.; Barras, F.; Lombard, M.; Pierrel, F. A Soluble Metabolon Synthesizes the Isoprenoid Lipid Ubiquinone. *Cell Chem. Biol.* **2019**, *26*, 482−492.e7.

(9) Chehade, M. H.; Loiseau, L.; Lombard, M.; Pecqueur, L.; Ismail, A.; Smadja, M.; Golinelli-Pimpaneau, B.; Mellot-Draznieks, C.; Hamelin, O.; Aussel, L.; Kieffer-Jaquinod, S.; Labessan, N.; Barras, F.; Fontecave, M.; Pierrel, F. ubiI, a New Gene in Escherichia Coli Coenzyme Q Biosynthesis, Is Involved in Aerobic C5-Hydroxylation. *J. Biol. Chem.* **2013**, *288*, 20085−20092.

(10) Crozier-Reabe, K.; Moran, G. Form Follows Function: Structural and Catalytic Variation in the Class A Flavoprotein Monooxygenases. *Indian J. Manag. Sci.* **2012**, *13*, 15601−15639.

(11) Reis, R. A. G.; Li, H.; Johnson, M.; Sobrado, P. New Frontiers in Flavin-Dependent Monooxygenases. *Arch. Biochem. Biophys.* 2021, 699, 108765.

(12) Westphal, A. H.; Tischler, D.; Heinke, F.; Hofmann, S.; Gröning, J. A. D.; Labudde, D.; van Berkel, W. J. H. Pyridine Nucleotide Coenzyme Specificity of P-Hydroxybenzoate Hydroxylase and Related Flavoprotein Monooxygenases. *Front. Microbiol.* 2018, 9, 3050.

(13) Meneely, K. M.; Lamb, A. L. Biochemical Characterization of a Flavin Adenine Dinculeotide-Dependent Monooxygenase, Ornithine Hydroxylase from Pseudomonas Aeruginosa, Suggests a Novel Reaction Mechanism. *Biochemistry* 2007, 46, 11930−11937.

(14) Zhu, Y.; Wu, B.; Zhang, X.; Fan, X.; Niu, L.; Li, X.; Wang, J.; Teng, M. Structural and Biochemical Studies Reveal UbiG/Coq3 as a Class of Novel Membrane-Binding Proteins. *Biochem. J.* 2015, 470, 105−114.

(15) Sun, Q.; Huang, M.; Wei, Y. Diversity of the Reaction Mechanisms of SAM-Dependent Enzymes. *Acta Pharm. Sin. B* 2021, 11, 632−650.

(16) Micklefield, J. Streamlined Recycling of S-Adenosylmethionine. *Nat. Catal.* 2019, 2, 644−645.

(17) Zhu, Y.; Jiang, X.; Wang, C.; Liu, Y.; Fan, X.; Zhang, L.; Niu, L.; Teng, M.; Li, X. Structural Insights into the Methyl Donor Recognition Model of a Novel Membrane-Binding Protein UbiG. *Sci. Rep.* 2016, 6, 23147.

(18) Tsui, H. S.; Clarke, C. F. Ubiquinone Biosynthetic Complexes in Prokaryotes and Eukaryotes. *Cell Chem. Biol.* 2019, 26, 465−467.

(19) Launay, R.; Teppa, E.; Martins, C.; Abby, S. S.; Pierrel, F.; André, I.; Esque, J. Towards Molecular Understanding of the Functional Role of UbiJ-UbiK2 Complex in Ubiquinone Biosynthesis by Multiscale Molecular Modelling Studies. *Int. J. Mol. Sci.* 2022, 23, 10323.

(20) Loiseau, L.; Fyfe, C.; Aussel, L.; Chehade, M. H.; Hernández, S. B.; Faivre, B.; Hamdane, D.; Mellot-Draznieks, C.; Rascalou, B.; Pelosi, L.; Velours, C.; Cornu, D.; Lombard, M.; Casadesús, J.; Pierrel, F.; Fontecave, M.; Barras, F. The UbiK Protein Is an Accessory Factor Necessary for Bacterial Ubiquinone (UQ) Biosynthesis and Forms a Complex with the UQ Biogenesis Factor UbiJ. *J. Biol. Chem.* 2017, 292, 11937−11950.

(21) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 2021, 596, 583−589.

(22) Pereira, J.; Simpkin, A. J.; Hartmann, M. D.; Rigden, D. J.; Keegan, R. M.; Lupas, A. N. High-accuracy protein structure prediction in CASP14. *Proteins* 2021, 89, 1687−1699.

(23) Lensink, M. F.; Nadzirin, N.; Velankar, S.; Wodak, S. J. Modeling Protein-protein, Protein-peptide, and Protein-oligosaccharide Complexes: CAPRI 7th Edition. *Proteins* 2020, 88, 916−938.

(24) Zhang, H.; Bei, Z.; Xi, W.; Hao, M.; Ju, Z.; Saravanan, K. M.; Zhang, H.; Guo, N.; Wei, Y. Evaluation of Residue-Residue Contact Prediction Methods: From Retrospective to Prospective. *PLoS Comput. Biol.* 2021, 17, No. e1009027.

(25) Evans, R.; O'Neill, M.; Pritzel, A.; Antropova, N.; Senior, A.; Green, T.; Žídek, A.; Bates, R.; Blackwell, S.; Yim, J.; Ronneberger, O.; Bodenstein, S.; Zielinski, M.; Bridgland, A.; Potapenko, A.; Cowie, A.; Tunyasuvunakool, K.; Jain, R.; Clancy, E.; Kohli, P.; Jumper, J.; Hassabis, D. Protein Complex Prediction with AlphaFold-Multimer. *BioRxiv* 2021.

(26) Elofsson, A. Progress at Protein Structure Prediction, as Seen in CASP15. *Curr. Opin. Struct. Biol.* 2023, 80, 102594.

(27) Gao, M.; Nakajima An, D.; Parks, J. M.; Skolnick, J. AF2Complex Predicts Direct Physical Interactions in Multimeric Proteins with Deep Learning. *Nat. Commun.* 2022, 13 (1), 1744.

(28) Bryant, P.; Pozzati, G.; Zhu, W.; Shenoy, A.; Kundrotas, P.; Elofsson, A. Predicting the Structure of Large Protein Complexes Using AlphaFold and Monte Carlo Tree Search. *Nat. Commun.* 2022, 13 (1), 6028.

(29) Sayers, E. W.; Bolton, E. E.; Brister, J. R.; Canese, K.; Chan, J.; Comeau, D. C.; Connor, R.; Funk, K.; Kelly, C.; Kim, S.; Madej, T.; Marchler-Bauer, A.; Lanczycki, C.; Lathrop, S.; Lu, Z.; Thibaud-Nissen, F.; Murphy, T.; Phan, L.; Skripchenko, Y.; Tse, T.; Wang, J.; Williams, R.; Trawick, B. W.; Pruitt, K. D.; Sherry, S. T. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2022, 50 (D1), D20−D26.

(30) Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* 2014, 30 (14), 2068−2069.

(31) Pelosi, L.; Vo, C.-D.-T.; Abby, S. S.; Loiseau, L.; Rascalou, B.; Hajj Chehade, M.; Faivre, B.; Goussé, M.; Chenal, C.; Touati, N.; Binet, L.; Cornu, D.; Fyfe, C. D.; Fontecave, M.; Barras, F.; Lombard, M.; Pierrel, F. Ubiquinone Biosynthesis over the Entire O2 Range: Characterization of a Conserved O2-Independent Pathway. *mBio* 2019, 10 (4), 21.

(32) Kazemzadeh, K.; Pelosi, L.; Chenal, C.; Chobert, S.-C.; Hajj Chehade, M.; Jullien, M.; Flandrin, L.; Schmitt, W.; He, Q.; Bouvet, E.; Jarzynka, M.; Varoquaux, N.; Junier, I.; Pierrel, F.; Abby, S. S. Diversification of Ubiquinone Biosynthesis via Gene Duplications, Transfers, Losses, and Parallel Evolution. *Mol. Biol. Evol.* 2023, 40, msad219.

(33) Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* 2011, 7 (10), No. e1002195.

(34) Edgar, R. C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* 2004, 32 (5), 1792−1797.

(35) Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* 2022, 19 (6), 679−682.

(36) Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* 2017, 35 (11), 1026−1028.

(37) Johnson, L. S.; Eddy, S. R.; Portugaly, E. Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. *BMC Bioinf.* 2010, 11, 431.

(38) Remmert, M.; Biegert, A.; Hauser, A.; Söding, J. HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment. *Nat. Methods* 2012, 9, 173−175.

(39) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 2015, 31, 926−932.

(40) Mitchell, A. L.; Almeida, A.; Beracochea, M.; Boland, M.; Burgin, J.; Cochrane, G.; Crusoe, M. R.; Kale, V.; Potter, S. C.; Richardson, L. J.; Sakharova, E.; Scheremetjew, M.; Korobeynikov, A.; Shlemov, A.; Kunyavskaya, O.; Lapidus, A.; Finn, R. D. MGnify: The Microbiome Analysis Resource in 2020. *Nucleic Acids Res.* 2019, 48, D570.

(41) Mirdita, M.; von den Driesch, L.; Galiez, C.; Martin, M. J.; Söding, J.; Steinegger, M. Uniclust Databases of Clustered and Deeply Annotated Protein Sequences and Alignments. *Nucleic Acids Res.* 2017, 45, D170−D176.

(42) Burley, S. K.; Berman, H. M.; Bhikadiya, C.; Bi, C.; Chen, L.; Costanzo, L. D.; Christie, C.; Duarte, J. M.; Dutta, S.; Feng, Z.; Ghosh, S.; Goodsell, D. S.; Green, R. K.; Guranovic, V.; Guzenko, D.; Hudson, B. P.; Liang, Y.; Lowe, R.; Peisach, E.; Periskova, I.; Randle, C.; Rose, A.; Sekharan, M.; Shao, C.; Tao, Y. P.; Valasatava, Y.; Voigt, M.; Westbrook, J.; Young, J.; Zardecki, C.; Zhuravleva, M.; Kurisu, G.; Nakamura, H.; Kengaku, Y.; Cho, H.; Sato, J.; Kim, J. Y.; Ikegawa, Y.; Nakagawa, A.; Yamashita, R.; Kudou, T.; Bekker, G. J.; Suzuki, H.; Iwata, T.; Yokochi, M.; Kobayashi, N.; Fujiwara, T.; Velankar, S.; Kleywegt, G. J.; Anyango, S.; Armstrong, D. R.; Berrisford, J. M.; Conroy, M. J.; Dana, J. M.; Deshpande, M.; Gane, P.; Gáborová, R.; Gupta, D.; Gutmanas, A.; Koča, J.; Mak, L.; Mir, S.; Mukhopadhyay, A.; Nadzirin, N.; Nair, S.; Patwardhan, A.; Paysan-Lafosse, T.; Pravda,

L.; Salih, O.; Sehnal, D.; Varadi, M.; Vařeková, R.; Markley, J. L.; Hoch, J. C.; Romero, P. R.; Baskaran, K.; Maziuk, D.; Ulrich, E. L.; Wedell, J. R.; Yao, H.; Livny, M.; Ioannidis, Y. E.; Ioannidis, Y. E. Protein Data Bank: The Single Global Archive for 3D Macromolecular Structure Data. *Nucleic Acids Res.* **2019**, *47*, D520−D528.

(43) Zhang, Y. TM-Align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33*, 2302−2309.

(44) Mukherjee, S.; Zhang, Y. MM-Align: A Quick Algorithm for Aligning Multiple-Chain Protein Complex Structures Using Iterative Dynamic Programming. *Nucleic Acids Res.* **2009**, *37*, No. e83.

(45) Glaser, F.; Pupko, T.; Paz, I.; Bell, R. E.; Bechor-Shental, D.; Martz, E.; Ben-Tal, N. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. *Bioinformatics* **2003**, *19*, 163−164.

(46) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891−3898.

(47) Bhagavan, N. Three-Dimensional Structure of Proteins. *Med. Biochem.* **2002**, *24*, 51−65.

(48) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. lDDT: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Bioinformatics* **2013**, *29*, 2722−2728.

(49) Nguyen, T. P. T.; Casarin, A.; Desbats, M. A.; Doimo, M.; Trevisson, E.; Santos-Ocaña, C.; Navas, P.; Clarke, C. F.; Salviati, L. Molecular Characterization of the Human COQ5 C-Methyltransferase in Coenzyme Q10 Biosynthesis. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2014**, *1841*, 1628−1638.

(50) Dai, Y.-N.; Zhou, K.; Cao, D.-D.; Jiang, Y.-L.; Meng, F.; Chi, C.-B.; Ren, Y.-M.; Chen, Y.; Zhou, C.-Z. Crystal Structures and Catalytic Mechanism of the C -Methyltransferase Coq5 Provide Insights into a Key Step of the Yeast Coenzyme Q Synthesis Pathway. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *70*, 2085−2092.

(51) Ortiz-Maldonado, M.; Gatti, D.; Ballou, D. P.; Massey, V. Structure−Function Correlations of the Reaction of Reduced Nicotinamide Analogues with P-Hydroxybenzoate Hydroxylase Substituted with a Series of 8-Substituted Flavins. *Biochemistry* **1999**, *38*, 16636−16647.

(52) Gatti, D. L.; Entsch, B.; Ballou, D. P.; Ludwig, M. L. pH-Dependent Structural Changes in the Active Site of p-Hydroxybenzoate Hydroxylase Point to the Importance of Proton and Water Movements during Catalysis. *Biochemistry* **1996**, *35*, 567−578.

(53) Schreuder, H. A.; Prick, P. A. J.; Wierenga, R. K.; Vriend, G.; Wilson, K. S.; Hol, W. G. J.; Drenth, J. Crystal Structure of the P-Hydroxybenzoate Hydroxylase-Substrate Complex Refined at 1.9 Å Resolution: Analysis of the Enzyme-Substrate and Enzyme-Product Complexes. *J. Mol. Biol.* **1989**, *208*, 679−696.

(54) Gatti, D. L.; Palfey, B. A.; Lah, M. S.; Entsch, B.; Massey, V.; Ballou, D. P.; Ludwig, M. L. The Mobile Flavin of 4-OH Benzoate Hydroxylase. *Science* **1994**, *266*, 110−114.

(55) Kuatsjah, E.; Johnson, C. W.; Salvachúa, D.; Werner, A. Z.; Zahn, M.; Szostkiewicz, C. J.; Singer, C. A.; Dominick, G.; Okekeogbu, I.; Haugen, S. J.; Woodworth, S. P.; Ramirez, K. J.; Giannone, R. J.; Hettich, R. L.; McGeehan, J. E.; Beckham, G. T. Debottlenecking 4-Hydroxybenzoate Hydroxylation in Pseudomonas Putida KT2440 Improves Muconate Productivity from p-Coumarate. *Metab. Eng.* **2022**, *70*, 31−42.

(56) Enroth, C.; Neujahr, H.; Schneider, G.; Lindqvist, Y. The Crystal Structure of Phenol Hydroxylase in Complex with FAD and Phenol Provides Evidence for a Concerted Conformational Change in the Enzyme and Its Cofactor during Catalysis. *Structure* **1998**, *6*, 605−617.

(57) Hiromoto, T.; Fujiwara, S.; Hosokawa, K.; Yamaguchi, H. Crystal Structure of 3-Hydroxybenzoate Hydroxylase from Comamonas Testosteroni Has a Large Tunnel for Substrate and Oxygen Access to the Active Site. *J. Mol. Biol.* **2006**, *364*, 878−896.

(58) Koskiniemi, H.; Metsä-Ketelä, M.; Dobritzsch, D.; Kallio, P.; Korhonen, H.; Mäntsälä, P.; Schneider, G.; Niemi, J. Crystal Structures of Two Aromatic Hydroxylases Involved in the Early Tailoring Steps of Angucycline Biosynthesis. *J. Mol. Biol.* **2007**, *372*, 633−648.

(59) Ryan, K. S.; Howard-Jones, A. R.; Hamill, M. J.; Elliott, S. J.; Walsh, C. T.; Drennan, C. L. Crystallographic Trapping in the Rebeccamycin Biosynthetic Enzyme RebC. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15311−15316.

(60) Greenhagen, B. T.; Shi, K.; Robinson, H.; Gamage, S.; Bera, A. K.; Ladner, J. E.; Parsons, J. F. Crystal Structure of the Pyocyanin Biosynthetic Protein PhzS. *Biochemistry* **2008**, *47*, 5281−5289.

(61) Treiber, N.; Schulz, G. E. Structure of 2,6-Dihydroxypyridine 3-Hydroxylase from a Nicotine-Degrading Pathway. *J. Mol. Biol.* **2008**, *379*, 94−104.

(62) Hekkelman, M. L.; de Vries, I.; Joosten, R. P.; Perrakis, A. AlphaFill: Enriching AlphaFold Models with Ligands and Cofactors. *Nat. Methods* **2023**, *20*, 205−213.

(63) Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276−W279.

(64) Zhu, W.; Shenoy, A.; Kundrotas, P.; Elofsson, A. Evaluation of AlphaFold-Multimer Prediction on Multi-Chain Protein Complexes. *Bioinformatics* **2023**, *39*, btad424.

(65) Soltanikazemi, E.; Quadir, F.; Roy, R. S.; Guo, Z.; Cheng, J. Distance-Based Reconstruction of Protein Quaternary Structures from Inter-Chain Contacts. *Proteins: Struct., Funct., Bioinf.* **2022**, *90*, 720−731.

(66) Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; DiMaio, F.; De Munck, S.; Savvides, S. N.; Baker, D. Improving de Novo Protein Binder Design with Deep Learning. *Nat. Commun.* **2023**, *14*, 2625.

(67) Eppink, M. H. M.; Schreuder, H. A.; van Berkel, W. J. H. Interdomain Binding of NADPH in P-Hydroxybenzoate Hydroxylase as Suggested by Kinetic, Crystallographic and Modeling Studies of Histidine 162 and Arginine 269 Variants*. *J. Biol. Chem.* **1998**, *273*, 21031−21039.

(68) Eppink, M. H. M.; Bunthof, C.; Schreuder, H. A.; van Berkel, W. J. Phe[161] and Arg[166] variants of p-hydroxybenzoate hydroxylase: Implications for NADPH recognition and structural stability. *FEBS Lett.* **1999**, *443*, 251−255.

(69) Pravda, L.; Sehnal, D.; Toušek, D.; Navrátilová, V.; Bazgier, V.; Berka, K.; Vařeková, R. S.; Koča, J.; Otyepka, M. MOLEonline: A Web-Based Tool for Analyzing Channels, Tunnels and Pores (2018 Update). *Nucleic Acids Res.* **2018**, *46*, W368−W373.

(70) Manicki, M.; Aydin, H.; Abriata, L. A.; Overmyer, K. A.; Guerra, R. M.; Coon, J. J.; Dal Peraro, M.; Frost, A.; Pagliarini, D. J. Structure and Functionality of a Multimeric Human COQ7:COQ9 Complex. *Mol. Cell* **2022**, *82*, 4307−4323.e10.