

Characterization of Coenzyme Q Biosynthesis Proteins through Integrative Modeling at the Protein-Membrane Interface

Thèse N° 9641

Présentée le 28 juin 2019

à la Faculté des sciences de la vie

Unité du Prof. Dal Peraro

Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Deniz AYDIN

Acceptée sur proposition du jury

Prof. P. De Los Rios, président du jury

Prof. M. Dal Peraro, directeur de thèse

Prof. S.J. Marrink, rapporteur

Prof. S. Vanni, rapporteur

Prof. G. D'Angelo, rapporteur

2019

Acknowledgements

I would like to offer my deepest gratitude to my thesis advisor, **Prof. Matteo Dal Peraro**, who has given me the chance to work in his lab. Matteo has encouraged and supported me throughout my PhD studies and towards my career goals. I have been extremely lucky to have a positive, cheerful, and calm supervisor who was always willing to listen to me while allowing me the room to work in my own way. My PhD has been an amazing experience thanks to him.

I would also like to express my deep appreciation to **Prof. Ruben Abagyan**. Ruben provided me the opportunity to join his lab in UCSD for 7 months, and has actively participated in my learning process during this time, which has been an invaluable experience for me. His hardworking and passionate attitude has been an inspiration to me. He has also continued supporting me beyond this time in his lab, and has been an amazing mentor.

I would like to thank **Prof. David Pagliarini**, **Dr. Danielle Lohman**, **Dr. Andrew Reidenbach**, and **Dr. Jonathan Stefely** for our productive collaboration that changed the course of my PhD. I feel very lucky that I had the opportunity to be involved in this exciting project, from which I learned a lot. I would like to extend my gratitude to Prof. David Pagliarini specifically for his support and encouragement during my postdoc and fellowship applications.

For critical reading of this thesis, and for their highly valued feedback, I would like to thank the experts of my thesis committee: **Prof. Stefano Vanni**, **Prof. Siewert Jan Marrink**, and **Prof. Giovanni D'Angelo**.

I am also very grateful for the valuable support of **Julia Prebandier**, **Sonja Bodmer**, and **Sandra Roux**, who always provided timely help with the administrative tasks throughout my PhD.

A big thank you goes to my fantastic labmates, the **LBMers**: Alessio, Giulia, Sylvain, Luciano, Giorgio, Martina, Chan, Alexandra, Lucien, and Maria. It was great sharing these

Acknowledgements

years with all of you, and it is thanks to you that I have very fond memories of my time in EPFL. I would also like to take this opportunity to specifically thank a few LBMers. **Alessio**, thank you for being the best officemate ever! Your positive and easygoing character and party spirit has turned our office into a comfortable and fun working zone, and I will miss our endless jokes and laughs. **Luciano**, thank you for your invaluable scientific insight, feedback, and guidance, and for always creating the time to help me. **Giorgio**, and **Sylvain**, it was a great pleasure collaborating with you, and learning from you. Finally, **Giulia**, **Chan**, **Alexandra**, and **Martina**, thank you for your everyday support and valuable friendship.

Outside of the lab, I would like to thank **Zeineb**, my closest friend in Lausanne, for the precious moments we shared together, and for being my sister away from home. I have also been extremely lucky to have two lifelong friends, who provided me endless support and care in all stages of my life. Thank you **Deniz Guven**, for the last 21 years, for our fun trips, for always being there for me, for believing in me and for caring about me so much, and for motivating me on a daily basis during my PhD! And thank you to my first friend ever, **Deniz Ilbi**, for being the kindest, the most warm-hearted and considerate person I know; knowing that you have always been and always will be there for me means a lot.

Last, but foremost in my heart, I would like to thank **Stefan** for his presence, dedication, and love, and for patiently supporting me through the difficult time of working on this thesis. A very special word of thanks goes to my loving family, my parents **Nesrin** and **Refhan**, and my sister **Yildiz**, who have made innumerable sacrifices for my education, and unconditionally supported me in every step I took. They have been a constant source of love and strength throughout my life, and this thesis is dedicated to them.

Lausanne, 12 June 2019

Deniz Aydin

Abstract

Integral and peripheral membrane proteins account for one-third of the human proteome, and they are estimated to represent the target for over 50% of modern medicinal drugs. Despite their central role in medicine, the complex, heterogeneous and dynamic nature of biological membranes complicates the investigation of their mechanism of action by both experimental and computational techniques. Among the different membrane bound compartments in eukaryotic cells, mitochondria are highly complex in form and function, and they harbor a unique proteome that remains largely unexplored. A growing number of inherited metabolic diseases are associated with mitochondrial dysfunction, which necessitates the structural and functional elucidation of mitochondrial proteins. In this thesis, we combine experimental and computational methods to explore the activity of COQ8 and COQ9, two functionally elusive proteins of the biosynthetic complex that produces coenzyme Q, a redox-active lipid component of the mitochondrial electron transport chain.

(i) Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8.

We demonstrate that COQ8 has an ATPase function that is activated when it specifically associates with cardiolipin-containing membranes. We identify its interaction surface with the inner mitochondrial membrane, which gives hints about the possible interaction surfaces with other members of the coenzyme Q synthesis machinery and has implications on how it mediates functional interactions with lipids. Collectively, this work reveals how the positioning of COQ8 on the inner mitochondrial membrane is key to its activation, and therefore advances our understanding of the COQ8 function.

(ii) Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9.

We explore the lipid binding activity of COQ9, and we reveal that COQ9 repurposes an ancient bacterial fold to selectively bind aromatic isoprenes, including CoQ intermediates that reside within the bilayer. We elucidate the mechanistic details of its membrane binding process, by which COQ9 warps the membrane surface and creates a tightly sealed hydrophobic region to access its lipid cargo. Finally, we establish a potential molecular interface between COQ9 and COQ7, the enzyme that catalyzes the penultimate step in CoQ biosynthesis, suggesting a model whereby COQ9 presents intermediates to CoQ enzymes to overcome the hydrophobic barrier of the membrane. Collectively, our results provide a mechanism for how a lipid binding protein might access, select, and extract specific cargo from a membrane and present it to a peripheral membrane enzyme.

In conclusion, our work is a good illustration of the interplay between experiment and modeling in protein research and specifically in understanding how proteins perform their action in direct synergy with membrane environments. We anticipate our integrative methodologies and mechanistic findings will prove relevant to other membrane proteins, whose fine functional modulation at the membrane-water interface has been historically challenging to characterize.

Keywords: membrane proteins, biological membranes, molecular modeling, molecular dynamics simulations, coenzyme Q biosynthesis, cardiolipins.

Résumé

Les protéines membranaires intégrales et périphériques représentent un tiers du protéome humain et on estime qu'ils représentent la cible de plus de 50% des médicaments modernes. Malgré leur rôle central en médecine, la nature complexe, hétérogène et dynamique des membranes biologiques complique l'étude de leur mécanisme d'action par des techniques expérimentales et informatiques. Parmi les différents compartiments liés aux membranes dans les cellules eucaryotes, les mitochondries ont une forme et une fonction très complexes et abritent un protéome unique qui reste largement inexploré. Un nombre croissant de maladies métaboliques héréditaires sont associées à un dysfonctionnement mitochondrial, ce qui nécessite une élucidation structurelle et fonctionnelle des protéines mitochondriales. Dans cette thèse, nous combinons des méthodes expérimentales et informatiques pour explorer l'activité de COQ8 et de COQ9, deux protéines fonctionnellement énigmatiques du complexe biosynthétique qui produit la coenzyme Q, un composant lipidique rédox-actif de la chaîne de transport d'électrons mitochondriale.

(i) Modulation lipidique conservée de COQ8, le membre d'une ancienne famille UbiB kinase-like.

Nous démontrons que COQ8 possède une fonction ATPase activée par association spécifique à des membranes contenant de la cardiolipine. Nous identifions sa surface d'interaction avec la membrane mitochondriale interne, indiquant sur les surfaces d'interaction possibles avec d'autres membres du mécanisme de synthèse de la coenzyme Q et a des implications sur la façon dont elle médie les interactions fonctionnelles avec les lipides. Ensemble, ce travail révèle que le positionnement de la COQ8 sur la membrane mitochondriale interne est la clé de son activation et permet ainsi de mieux comprendre la fonction de la COQ8.

(ii) Interactions membranaires, lipidiques et protéiques de COQ9.

Nous explorons l'activité de liaison des lipides de COQ9 et nous révélons que COQ9 réutilise un ancien pli bactérien pour lier sélectivement les isoprènes aromatiques, y compris les intermédiaires de CoQ qui résident dans la bicouche. Nous élucidons les détails mécanistiques de son processus de liaison membranaire, par lequel COQ9 déforme la surface de la membrane et crée une région hydrophobe étroitement scellée pour accéder à sa cargaison lipidique. Enfin, nous établissons une interface moléculaire potentielle entre COQ9 et COQ7, l'enzyme qui catalyse l'avant-dernière étape de la biosynthèse de CoQ, suggérant un modèle dans lequel COQ9 présente des intermédiaires aux enzymes CoQ pour surmonter la barrière hydrophobe de la membrane. Ensemble, nos résultats mettent en évidence un mécanisme expliquant comment une protéine ancrée à un lipide pourrait accéder, sélectionner et extraire une cargaison spécifique d'une membrane et de la présenter à une enzyme membranaire périphérique.

En conclusion, cette thèse illustre bien l'interaction entre l'expérience et la modélisation dans la recherche sur les protéines et permet en particulier de comprendre comment les protéines exercent leur action en synergie directe avec les environnements membranaires. Nous sommes certains que nos méthodologies intégratives et nos découvertes mécanistiques pourront s'appliquer à d'autres protéines membranaires, dont la modulation fonctionnelle fine à l'interface entre le milieu aqueux et la membrane a toujours été difficile à caractériser.

Mots clés : protéines membranaires, membrane biologique, modélisation, simulation de dynamique moléculaire, biosynthèse du coenzyme Q, cardiolipines.

Contents

Acknowledgements	iii
Abstract (English/Français)	v
Table of Contents	ix
List of Tables	xiii
List of Figures	xiv
1 Introduction	1
1.1 Biological Membranes and Membrane Proteins	1
1.2 The Mitochondrial Membrane	3
1.3 Coenzyme Q Biosynthesis	5
1.4 The Importance of Molecular Modeling and Simulations	6
1.5 Objectives of the Thesis	8
2 Molecular Modeling and Simulation Methods	11
2.1 Modeling Biomolecular Interactions with Molecular Mechanics	11
2.1.1 Bonded Interactions	12
2.1.2 Non-Bonded Interactions	14
2.2 Molecular Dynamics Simulations	16
2.2.1 The Ergodic Hypothesis	18
2.2.2 Periodic Boundary Conditions	19
2.2.3 Coarse-Grained Representation of Biomolecules	19
2.3 Ligand Docking Methods	23
	ix

3	Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8	27
3.1	Introduction	28
3.2	Methods	30
3.2.1	Identification and Characterization of Binding Pockets on Nucleotide-Bound and Apo COQ8A	30
3.2.2	Coarse-Grained Molecular Dynamics Simulations of Protein-Membrane Systems	31
3.2.3	Modeling of Transmembrane Domain-Containing COQ8A	32
3.2.4	Identification of Residues Important for Membrane Interaction and Selection of Residues for Mutagenesis Experiments	32
3.2.5	Modeling of COQ8A Mutants	34
3.3	Results	34
3.3.1	Nucleotide Binding to COQ8A Opens Small-Molecule Pockets	34
3.3.2	COQ8 Interacts with the Membrane through its Signature KxGQ Domain	36
3.3.3	Cardiolipin Specifically Increases COQ8 Membrane Interaction and ATPase Activity	38
3.3.4	Cardiolipin Specifically Increases COQ8 Membrane Interaction and ATPase Activity	41
3.4	Discussion	44
3.5	Supplementary Information	45
3.5.1	DNA Constructs and Cloning	45
3.5.2	Recombinant Protein Expression and Purification	46
3.5.3	ADP-Glo Assay	48
3.5.4	Cytosphos ATPase Assay	48
3.5.5	In Vitro [γ - ³² P] Autophosphorylation and ATPase Comparison Assay . .	49
3.5.6	In Vitro [γ - ³² P]ATP ATPase Assay	49
3.5.7	Yeast Drop Assay and Growth Curves	49
3.5.8	Liposomes	50
3.5.9	Liposome Flotation Assay	50

4 Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9 53

4.1	Introduction	54
4.2	Methods	56
4.2.1	Coarse-Grained Molecular Dynamics Simulations	56
4.2.2	Bioinformatics, Docking, and Modeling	59
4.3	Results	62
4.3.1	COQ9 Specifically Binds Aromatic Isoprenes	62
4.3.2	The Amphipathic 10th α -Helix of COQ9 Drives Membrane Association	68
4.3.3	COQ9 Association with Cardiolipin-Rich Membranes Warps the Lipid Bilayer	71
4.3.4	COQ9 Interaction with Aromatic Isoprenes Illuminates Mechanisms of Specificity	73
4.3.5	COQ9 Interaction and Interface with COQ7 Suggest a Lipid Presentation Role	75
4.3.6	COQ9 Lipid- and Membrane-Binding Residues are Critical for CoQ Pro- duction <i>In Vivo</i>	78
4.4	Discussion	80
4.5	Supplementary Information	86
4.5.1	DNA Constructs and Cloning	86
4.5.2	Protein Expression and Purification	86
4.5.3	Biochemical Assays	87
4.5.4	<i>S. cerevisiae</i> Growth Assays	89
4.5.5	Crystallography	89
4.5.6	Liquid Chromatography Mass Spectrometry Lipidomics	91
4.5.7	Sample Preparation for <i>S. cerevisiae</i> Lipids Enriched in Coq9p-FLAG Pull- Downs	93
4.5.8	Sample Preparation for <i>E. coli</i> Lipids Co-Purifying with His ₈ -MBP-COQ9 ^{NA79}	95
4.5.9	Sample Preparation for CoQ Intermediates in <i>S. cerevisiae</i>	95
4.5.10	LC-MS/MS Proteomics Analysis of <i>S. cerevisiae</i>	97
4.5.11	Quantification and Statistical Analysis	98
4.5.12	Experimental Data Availability and Accession Codes	98

5	Conclusions and Perspectives	99
A	Using Energy Grids to Constrain the Search Space of Small Molecules During Docking	103
A.1	Introduction	104
A.2	Methods	106
A.2.1	The Constrained Optimization Strategy	106
A.2.2	General Formulation of the Docking Problem	108
A.2.3	The Objective Function	108
A.2.4	Constraint Terms	110
A.2.5	Training the Energy Grid Map Parameters To Obtain Optimal Constraints	111
A.2.6	Step-by-Step Docking Protocol	113
A.2.7	Generation of the Ligand Conformational Ensemble	114
A.2.8	Metrics to Evaluate Accuracy	115
A.3	Results and Discussion	115
A.3.1	Constrained Rigid and Flexible Docking	116
A.3.2	Evaluation of Successes and Failures	119
A.4	Conclusions	122
	Bibliography	125
	Curriculum Vitae	

List of Tables

3.1	Ranking of Charged/Polar Residues According to their Interaction Frequencies with the Membrane	33
3.2	Designed COQ8A Mutants	34
3.3	Characterization of Binding Pockets on Nucleotide Bound- and Apo-COQ8A . .	35
4.1	Protein Constructs	65
4.2	Data Collection and Refinement Statistics	66
A.1	Flexible Docking of the Astex Diverse Set and Reported Accuracies for <i>pow^{er}</i> - <i>mViE</i> and AutoDock Vina	117

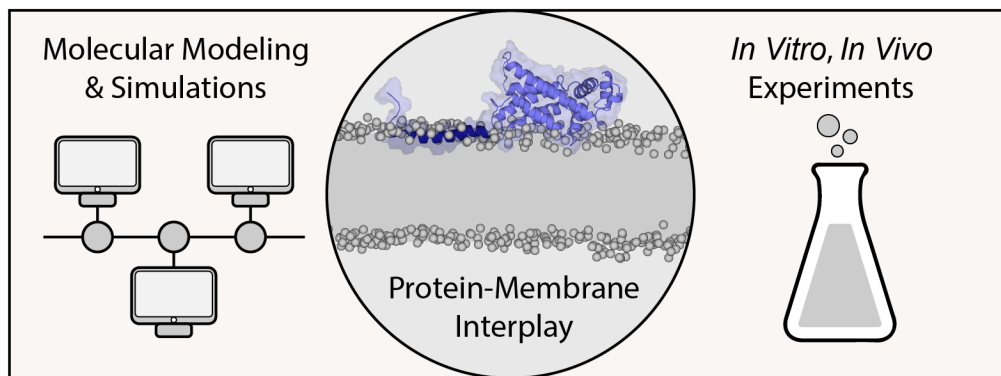
List of Figures

1.1	Protein-Membrane Interaction Modes	3
1.2	Chemical Structure of Glycerolipids	4
1.3	Coenzyme Q Chemical Structure, and the CoQ-Synthome	5
1.4	Eukaryotic CoQ Biosynthesis	7
2.1	Bonded and Non-Bonded Interactions in Molecular Mechanics	12
2.2	Periodic Boundary Conditions	20
2.3	All-Atom and Coarse-Grained (CG) Representations of Proteins	20
2.4	Martini Mapping Examples of Selected Molecules	23
2.5	The Viability Evolution Algorithm	25
3.1	Crystal Structures of Apo and Nucleotide-Bound COQ8A	29
3.2	Nucleotide Binding to COQ8A Opens Small-Molecule Pockets	36
3.3	Cardiolipin Enhances COQ8A Membrane Binding	37
3.4	Cardiolipin Enhances the ATPase Activity and Liposome Binding of COQ8A	40
3.5	COQ8 Interacts with the Membrane through Its Signature KxGQ Domain	43
3.6	Model for How COQ8 Membrane and CL Binding, and ATPase Activity Could Facilitate CoQ Biosynthesis	45
4.1	Lipid-Binding Proteins	55
4.2	LC-MS/MS Analysis of COQ9 Lipid Ligands	63
4.3	COQ9 Binds Aromatic Isoprenes	64
4.4	Crystal Structure of COQ9 with Geranylgeraniol and Extended Helix 10	67
4.5	The 10th α -Helix of COQ9 is Unique, Amphipathic, and Flexible	68

List of Figures

4.6	COQ9 Associates with Membranes in 5 Steps	69
4.7	COQ9 Membrane Association is Strongly Influenced by Cardiolipin	72
4.8	COQ9 Warps Cardiolipin-Containing Membranes	73
4.9	COQ9 Creates a Hub for CoQ Molecules	75
4.10	Biochemical Data Illuminates Possible COQ9-COQ7 Intermolecular Contacts .	76
4.11	COQ9 and COQ7 Predicted Molecular Interface Suggests Substrate Presentation	78
4.12	Coq9p Protein-, Lipid- and Membrane-Disrupting Mutants Perturb Endogenous CoQ Biosynthesis	79
4.13	Coq9p Mutant Growth Inhibition Profile	81
4.14	Phenotypes of $\alpha 10$ Swap Chimeric Proteins	82
A.1	Constrained Ligand Docking Workflow	107
A.2	Constraint Generation from Energy Grid Maps	112
A.3	Docking Success and Failures of <i>pow^{er}</i> - <i>mViE</i> and AutoDock Vina	121

1 Introduction



1.1 Biological Membranes and Membrane Proteins

Biological membranes act as a barrier delimiting the cell from its environment, and allow subcellular compartmentalization to produce discrete organelles. They are composed of a complex repertoire and a dynamic distribution of proteins, lipids and sugars.

Eukaryotic cells invest a significant amount of resources in synthesizing thousands of different lipids.¹ Lipids are amphipathic molecules that can be grouped into three categories: (i) glycerolipids, (ii) sphingolipids, and (iii) sterols.

Glycerolipids are the most abundant membrane lipids,¹ and they are composed of a hy-

drophilic head group and hydrophobic tails. Due to this amphipathic character, lipids arrange themselves into a sheet-like structure known as the lipid bilayer, which is composed of two leaflets where polar head groups interact with the solvent and lipid tails form the hydrophobic core. The diversity of glycerolipids arises from the head group substituent, the length of lipid tails, and the number and position of unsaturations in the lipid tails. Phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI) and phosphatidic acid (PA) are lipids that belong to the glycerolipid class.

The second class of lipids, sphingolipids, have a general structure similar to that of glycerolipids, except that they contain the fatty acid derivative sphingosine instead of glycerol. They consist of different surface-exposed head groups, such as phosphorylcholine (sphingomyelin) or carbohydrates (glycosphingolipids). Sphingolipids are particularly abundant in nerve cells and brain tissues.²

The plasma membrane has the highest abundance of sterols, and the highest sterol to phospholipid ratio. Cholesterol is the major sterol found in plasma membranes of higher eukaryotes, while ergosterol is the major sterol component present in lower eukaryotes.³

While the width of biological membranes is about 4 nm considering the lipid bilayer only, extra constituents protruding the membrane (membrane-associated proteins, lipopolysaccharides) can increase the overall membrane width to 10 nm. Membrane-associated proteins represent about 30% of the whole human genome,⁴ and more than 50% of the targets of all drugs.⁵ Based on their mode of interaction with the membrane, membrane proteins can be classified into two categories: integral and peripheral membrane proteins (Figure 1.1). While lipids ensure permeability for the lipid bilayer, integral membrane proteins typically enable signal transduction and molecular transport. Integral membrane proteins contain hydrophilic regions exposed to the solvent on both sides of the bilayer, and one or more segments embedded into the hydrophobic core of the membrane, typically spanning the entire bilayer (Figure 1.1). These transmembrane domains can be composed of one or more α -helices, or multiple β -strands. On the other hand, peripheral membrane proteins are soluble in water and they do not fully penetrate the hydrophobic core of the membrane. They associate reversibly with the membrane through electrostatic/hydrophobic interactions with lipids,

covalently bound fatty acids (lipidation), or through interacting with other membrane-bound proteins (Figure 1.1). The interaction of peripheral membrane proteins with the membrane is highly influenced by the membrane lipid composition.⁶ They are known to have key roles in non-vesicular lipid transportation, cell signaling, and detecting lipid packing defects.^{6,7}

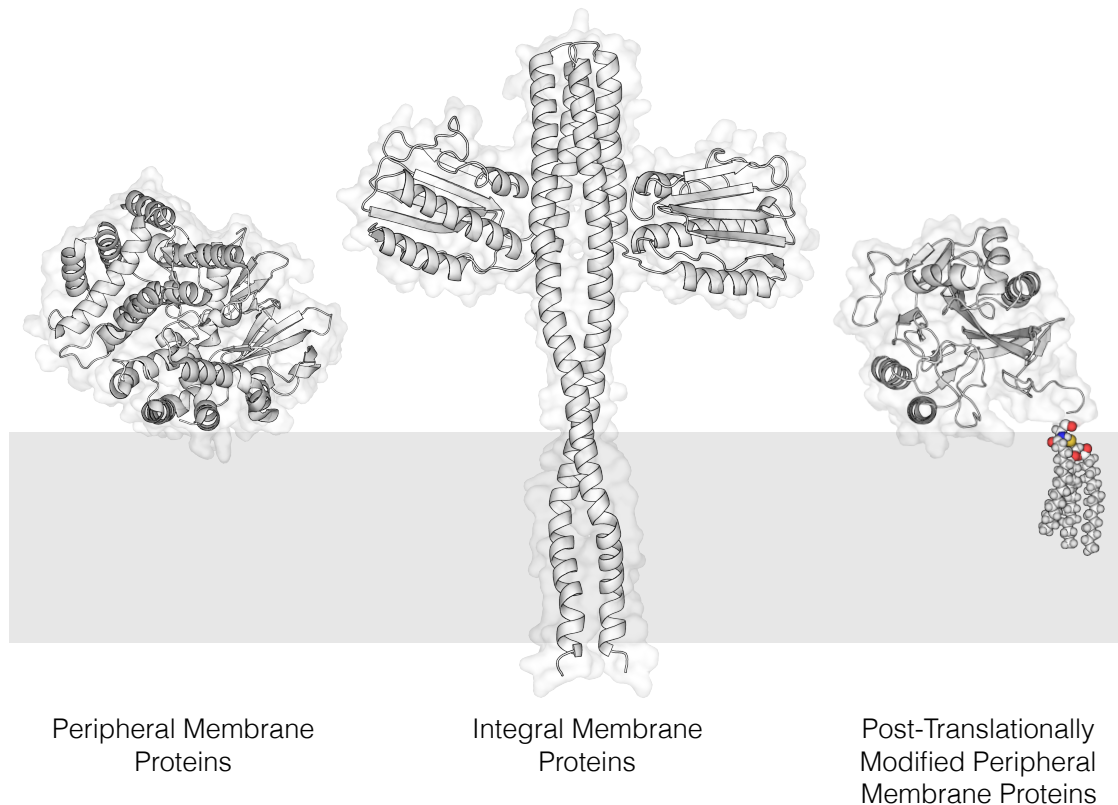


Figure 1.1 – Protein-Membrane Interaction Modes.

1.2 The Mitochondrial Membrane

As a main source of ATP in eukaryotes, the mitochondria are separated from the cytoplasm by two specialized membranes called the outer and inner mitochondrial membrane. The outer membrane has a smooth and unfolded structure, while the inner membrane, encircled by the outer membrane, has a folded structure, with a large number of inward projections called cristae. The cristae protrude into the matrix, and increase the surface area of the inner membrane, which in turn increases the room for respiratory chain complexes and ATP

Chapter 1. Introduction

synthase.⁸ The two membranes are separated by a 10 nm wide intermembrane space,⁹ which resembles the cytosol and contains enzymes that catalyze the interconversion of adenine nucleotides.

The outer and inner mitochondrial membranes differ in lipid composition. The outer membrane is characterized by a high lipid to protein ratio and has a higher permeability than the inner membrane.¹⁰ While the outer membrane is rich in sterols, the inner membrane contains a high percentage of the unusual lipid, cardiolipin. Cardiolipin is a unique anionic glycerophospholipid that has a dimeric structure. It contains four lipid tails and can carry up to two negative charges, one per phosphate group. It is the hallmark lipid of the mitochondrial inner membrane, where it contributes up to 20% of total lipids.¹¹ Due to its chemical structure with one head group and four acyl chains, cardiolipins adopt a cone-shaped structure and lead to a negative membrane curvature. Cardiolipins are shown to form microdomains,¹¹ and they also play an essential role in the stability of protein supercomplexes.^{12, 13}

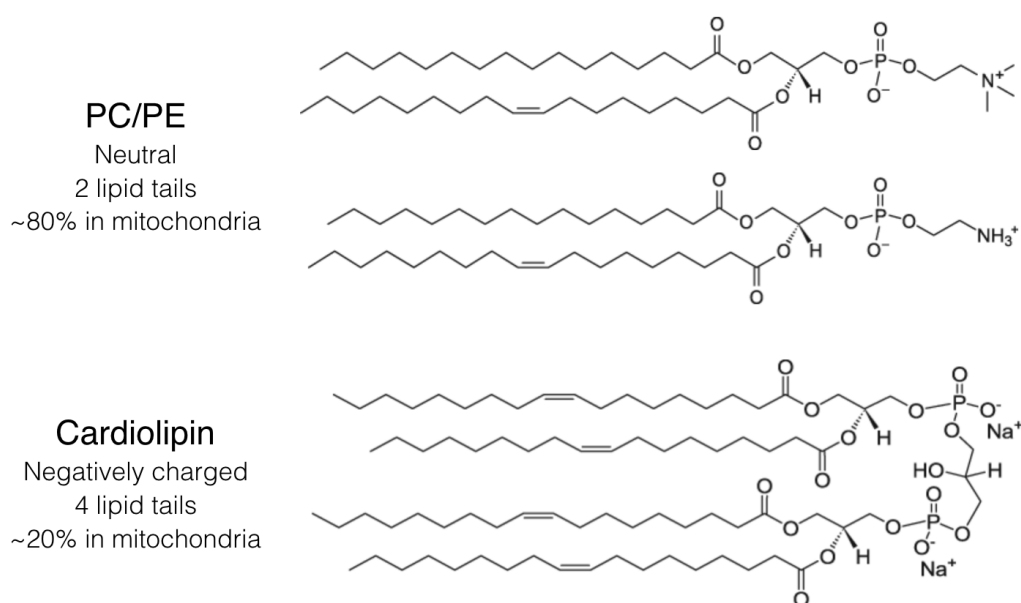


Figure 1.2 – Chemical structure of glycerolipids. Phosphatidylcholine (PC) and phosphatidylethanolamine (PE) are neutral lipids that have two lipid tails, whereas cardiolipin is a negatively charged lipid with four lipid tails. Cardiolipin is the signature lipid of mitochondria.

1.3 Coenzyme Q Biosynthesis

Coenzyme Q (ubiquinone, Q, or CoQ), is a lipid that participates in aerobic cellular respiration to produce ATP within the mitochondria in eukaryotes and the plasma membrane in prokaryotes. CoQ exists in both reduced and oxidized states and the conversion between these states allows it to transfer electrons to substrates and act as a cofactor of enzymatic reactions, including those of the mitochondrial electron transport chain. The quinone head group of CoQ allows it to function as an electron-carrier, while the highly hydrophobic isoprene tail helps to confine it to lipid-rich areas of cells (Figure 1.3). The extraordinarily long hydrophobic tail makes CoQ one of the most hydrophobic molecules in nature.

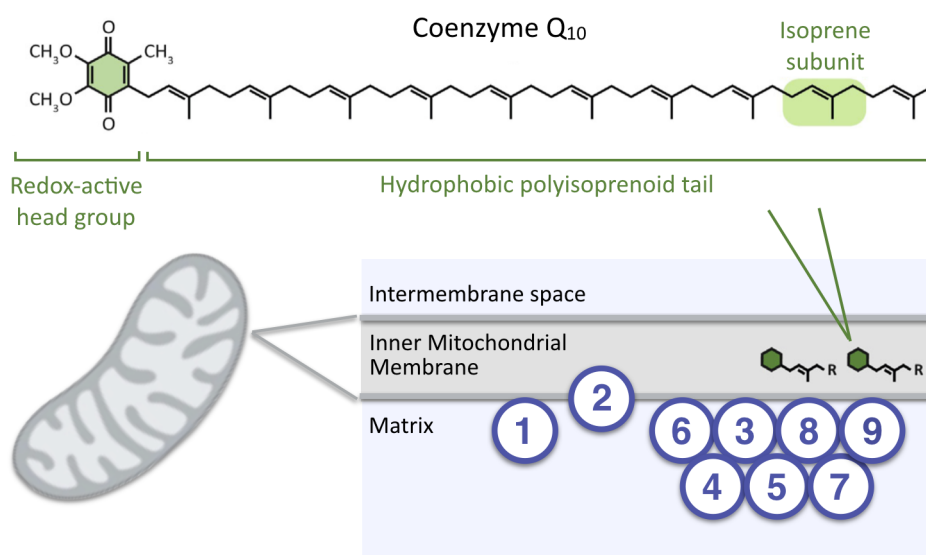


Figure 1.3 – Coenzyme Q Chemical Structure, and the CoQ-Synthome. The chemical structure of coenzyme Q₁₀ (CoQ₁₀) includes a redox-active head group (which allows it to function as a cofactor in the mitochondrial electron transport chain), and a hydrophobic polyisoprenoid lipid tail (which is made of up a varying number of isoprene subunits in different organisms). Proteins crucial for the biosynthesis of CoQ are known to form a biosynthetic complex on the matrix side of the inner mitochondrial membrane. This "metabolon" is termed **complex Q**, or the **CoQ-synthome**.

Genetic and physical evidence indicates that COQ gene products are interdependent, and that CoQ biosynthesis requires a multi-subunit complex of the COQ-encoded enzymes,¹⁴ termed the "complex Q", or the "CoQ-synthome" (Figure 1.3). In eukaryotes, the biosynthesis of CoQ requires the collective action of COQ proteins COQ1-9, and mutations of these proteins can lead to CoQ deficiency, which is further linked to cerebellar ataxia, kidney disease,

isolated myopathy, and childhood-onset multisystemic disorders.¹⁵ These diseases affect multiple organ systems, but primarily the heart, the brain, kidneys and skeletal muscles, i.e. locations where mitochondrial oxidative phosphorylation is needed the most.¹⁶ The complex mechanisms leading to these diseases remain largely unknown because of various gaps in our understanding of CoQ biochemistry.

CoQ is synthesized in all cells and in organisms across all domains of life. Most eukaryotic CoQ is known to be generated at the inner mitochondrial membrane, even though it can also be produced outside of mitochondria.¹⁶ The biosynthesis of CoQ involves (i) head group production, (ii) polyisoprenoid tail production, (iii) attachment of the tail to the head group, and (iv) successive steps of head group modifications (Figure 1.4). CoQ intermediates become increasingly hydrophobic throughout the course of the pathway. CoQ and its intermediates are predicted to diffuse laterally in the hydrophobic core of the phospholipid bilayer, where they can move toward the bilayer surface, but not past the glycerol backbones.¹⁷ It is unclear how peripheral membrane enzymes gain access to hydrophobic CoQ intermediates to perform head group modifications.¹⁶

Currently, X-ray structures of three COQ proteins have been reported:

- Yeast Coq5p (apo (PBD: 4OBX) and S-adenosyl methionine (SAM)-bound (PBD: 4OBW)),¹⁸
- Human COQ8 (apo (PBD: 4PED)¹⁹ and ATP analog-bound (PBD: 5I35)²⁰),
- Human COQ9 (apo (PDBs: 4RHP¹⁵ and 6AWL²¹) and isoprene-bound (PDB: 6DEW)²¹).

COQ4, COQ8, and COQ9 are proteins that have no clear catalytic role in the CoQ biosynthesis pathway, even though they are known to be crucial for the biosynthesis. This thesis focuses on two of these functionally uncharacterized proteins, COQ8 and COQ9, and aims to shed light on the role of these proteins in CoQ biosynthesis.

1.4 The Importance of Molecular Modeling and Simulations

Despite their remarkable role in biology and pharmacology, only 3% of the Protein Data Bank entries refer to membrane proteins. This is mainly due to their hydrophobicity, high flexibility, and lack of stability, which make it extremely challenging to treat these proteins by means of techniques such as X-ray crystallography, NMR or cryo-EM. In this sense, the rise of

1.4. The Importance of Molecular Modeling and Simulations

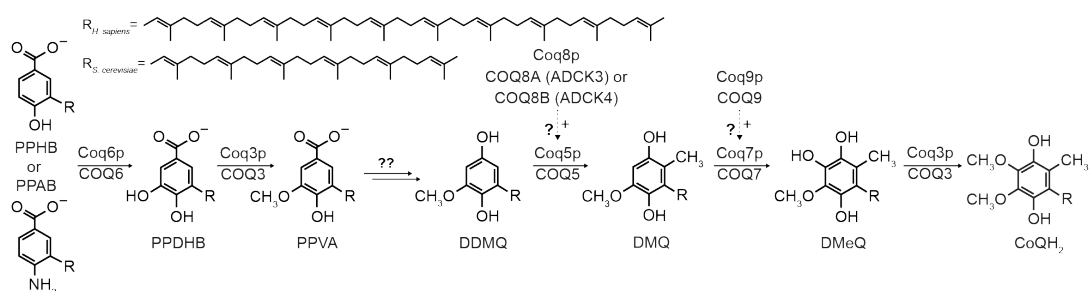


Figure 1.4 – Eukaryotic CoQ Biosynthesis. Head group modifications of eukaryotic CoQ biosynthesis post-attachment of the polyprenyl tail (R) where *S. cerevisiae* (Coq#p) and *H. sapiens* (COQ#) protein names are shown. PPHB, polyphenyl-hydroxybenzoate; PPAB, polyphenyl-aminobenzoate; PPDHB, polyphenyl-dihydroxybenzoate; PPVA, polyphenyl-vanillic acid; DDMQ, demethoxy-demethyl CoQ; DMQ, demethoxy CoQ; DMeQ, demethyl CoQ; CoQH₂, reduced CoQ.

computational tools allows to complete and further analyze structural aspects of membrane proteins that are generally not accessible experimentally.

As membrane biology is surging as a fundamental field of research to understand the processes involved in cellular function, the role played by lipids in the modulation of the function of peripheral and integral membrane proteins is nowadays recognized as crucial. A reflection of this fact is found in the recent advances in techniques (e.g., mass spectrometry, lipid-overlay and lipid pull-down assays, liposome microarray-based assays, affinity-purification lipidomics, cryo-EM of membrane proteins) that aim at providing a more faithful characterization of membranes and membrane-protein interactions. As a result, we now have a better classification of the lipidome repertoire that counts today more than 40,000 different lipid species; however, little is still known about their functional role, and even less about their subtle interplay with membrane proteins. As these experimental advances are rapidly changing our classical understanding of biological membranes, it still remains a major challenge to acquire high-resolution molecular data for membranes and specific lipid-protein interactions. Additionally, a growing number of lipid-related proteins are associated with newly identified disease-related genes, which calls for a functional elucidation in order to improve diagnostics and develop new therapeutic interventions. In this context, optimally positioned at the crossroads of chemistry, biology, physics and life sciences, computational modeling and molecular simulations are powerful (and possibly the only) tools able to integrate the outcomes of these developing techniques, providing a molecular and mechanistic picture of the membrane-protein interplay. In summary, at the dawn of the lipidomics era, it

is absolutely required to combine experimental studies with molecular models that mimic as close as possible the physiological state of biological membranes.

Recent developments on the computational front include a number of tools that allow the precise multiscale modeling of the membrane. In particular, tools like CHARMM-GUI²² and *LipidBuilder*²³ enable the accurate generation of lipid bilayer geometries and the fine-tuning of the lipid composition by altering both the head type and the hydrophobic tail. Moreover, a growing number of validated parameters for lipids in different force fields, both at the coarse-grained (MARTINI²⁴) and the atomistic level (AMBER,²⁵ CHARMM²⁶) allow the investigation of membranes and associated membrane proteins at an extremely precise level of detail, spanning time scales that go from nanoseconds to milliseconds, and length scales up to micrometers. The rise of these tools requires a deep rationalization about their potential and limitations in complementing both the traditional and the most innovative experimental techniques that are being used nowadays in the field of membrane proteins.

1.5 Objectives of the Thesis

The integration of molecular modeling and simulations with structural and biochemical data represents an effective approach for studying the mechanism and function of proteins in their biological environments. This thesis combines computational and experimental techniques to characterize two poorly understood members of the CoQ biosynthesis pathway.

(1) Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8.

COQ8 is a member of the protein kinase-like UbiB superfamily, which includes protein and lipid kinases, and ATPases. Even though COQ8 was initially thought to act as a protein kinase, its structural characterization has revealed features that inhibit protein kinase activity. As COQ8 is known to be localized to the inner mitochondrial membrane, in this thesis, we used coarse-grained molecular dynamics simulations to explore the protein-membrane interaction of COQ8 and the inner mitochondrial membrane. We then used biochemical experiments to validate our computational findings and to further understand how this interplay enhances CoQ biosynthesis. We demonstrate that COQ8 has an ATPase function that is enhanced when it specifically associates with cardiolipin-containing membranes, and we reveal that a conserved

domain on COQ8 drives the membrane association electrostatically.

(2) Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9.

The structural characterization of COQ9 has revealed structural similarity to the bacterial TetR family of transcriptional regulators, and the presence of a lipid-binding pocket. In addition, COQ9 physically and functionally interacts with COQ7, the enzyme that catalyzes the penultimate step in CoQ biosynthesis. Nevertheless, which lipid(s) COQ9 binds, how it accesses these lipid ligands, and how this lipid binding functionality might assist other CoQ pathway enzymes have remained unanswered. In this thesis, we explore the membrane, lipid, and protein binding activity of COQ9 through combined computational and experimental studies. We report the stepwise mechanism for its membrane binding process, and we describe a potential molecular interface between COQ9 and COQ7, suggesting a model whereby COQ9 acts as a lipid presenter to overcome the hydrophobic barrier of the membrane.

This thesis is organized as follows. First, computational techniques and concepts used in this work will be detailed (Chapter 2). Next, results of molecular modeling and simulation techniques revealing the membrane interaction of COQ8 will be described, and the experimental work performed by our experimental collaborators to confirm our results will be presented (Chapter 3). As a continuation of this project, computational and experimental results for the functional elucidation of COQ9 will be presented (Chapter 4). For both of these projects, computational methods performed in our lab will be detailed in the Methods sections of these Chapters, while experimental methods used by our experimental collaborators will be presented in the respective Supplementary Information sections. Finally, conclusions and perspectives of these studies will be presented (Chapter 5). A study related to the computational methods described in this thesis will be presented in the Appendix, which focuses on the development of a new optimization method for ligand docking that uses pharmacophoric probes as constraints.

2 Molecular Modeling and Simulation

Methods

2.1 Modeling Biomolecular Interactions with Molecular Mechanics

Molecular mechanics calculations use a force field that describes the potential energy of the system through the sum of molecular interactions as a function of atomic positions. This approximation decreases the computational cost of simulating large and complex biological systems, while providing a good level of accuracy. Force field parameters are obtained from experimental or theoretical data. An important feature of force field parameters is that they should be transferable; in other words, a set of parameters developed and tested for a particular atom or group of atoms should be valid for the same atom/group of atoms in a different molecule.

Considering a system composed of N particles where atoms are described as spheres connected by springs, the interactions are defined as a potential energy, U , which can be defined as the sum of bonded and non-bonded interactions:

$$U_{total} = U_{bonded} + U_{non-bonded} \quad (2.1)$$

The minima of U_{total} represent stable conformations, while conformational transitions can be described through moving along the potential energy landscape.

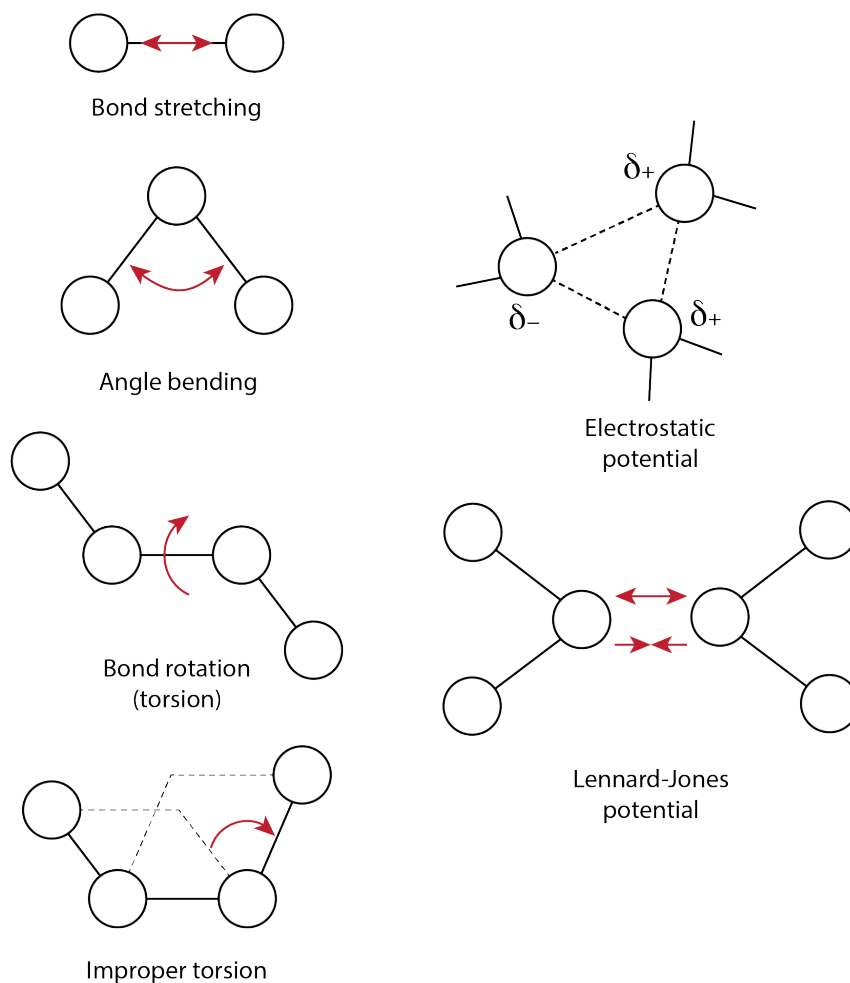


Figure 2.1 – Bonded and Non-Bonded Interactions in Molecular Mechanics. Bond stretching, angle bending, and proper and improper bond torsions add up to form the bonded interactions, while electrostatic and van der Waals interactions add up to form the non-bonded interactions.

2.1.1 Bonded Interactions

The bonded potential, U_{bonded} , describes the intramolecular interactions, and includes four contributions: bond stretching (2-body), bond angle (3-body), and torsional angle (4-body) interactions. A special type of torsional interaction (called improper torsion) is mainly used to maintain planarity in molecular structures (Figure 2.1). U_{bonded} can therefore be expressed as:

$$U_{bonded} = U_{bond} + U_{angle} + U_{torsion} + U_{improper} \quad (2.2)$$

Bond stretching

The first term of the bonded potential, U_{bond} , represents the energy associated with the bond stretching between two covalently bonded atoms (Figure 2.1). The most common potential used in molecular dynamics programs is the harmonic bond potential, expressed as:

$$U_{bond} = \sum_{bonds} k_r (r - r_{eq})^2 \quad (2.3)$$

where k_r is the force constant, r is the distance between two covalently bonded atoms, and r_{eq} is the equilibrium (reference) bond length. Since there are several force field components, the bond will deviate slightly from this reference value to be able to compensate for other contributors to the total potential energy.

Angle bending

The second term of the bonded potential, U_{angle} , represents the energy associated with the distortion of an angle between three consecutive atoms (Figure 2.1). Since forces between bonded atoms are very strong, the energy required for stretching or compressing a bond from equilibrium is high. Distorting an angle away from equilibrium requires less energy; therefore force constants for angle bending are proportionately smaller. As with U_{bond} , U_{angle} is also described using a harmonic potential:

$$U_{angle} = \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \quad (2.4)$$

where k_θ is the force constant, θ is the angle between three consecutive atoms, and θ_{eq} is the equilibrium angle.

Torsional angles

The third term of the bonded potential, $U_{torsion}$, concerns four consecutive atoms, and is divided into two types: proper and improper torsion potentials (Figure 2.1). The proper torsion potential is used to constrain rotations around a bond, and are modelled using a form of a

cosine series:

$$U_{torsion} = \sum_{torsions} k_{\phi} [1 + \cos(n\phi - \phi_{eq})] \quad (2.5)$$

where k_{ϕ} is the force constant, ϕ is the torsional angle between four consecutive atoms, and ϕ_{eq} is the phase factor, or the equilibrium torsional angle (the angle where the potential passes through its minimum value). n is the periodicity of the torsional potential that determines the number of minima in the torsional potential.

Improper torsional angles

The last term of the bonded potential, $U_{improper}$, is commonly considered as an additional term in the force field to maintain stereochemistry at chiral centers and more generally, to maintain the planarity of molecules, such as the aromatic rings. The improper torsional angle can be modelled using a harmonic potential since it would have one minimum to maintain planarity:

$$U_{improper} = \sum_{improvers} k_{\omega} (\omega - \omega_{eq})^2 \quad (2.6)$$

where k_{ω} is the force constant, ω is the torsional angle between four consecutive atoms, and ω_{eq} is the reference torsional angle.

2.1.2 Non-Bonded Interactions

In addition to the bonded interactions between atoms, force fields also consider non-bonded interactions, which act both between independent molecules, and between atoms in the same molecule. Non-bonded interactions are generally the sum of two contributions: van der Waals interactions and electrostatic interactions.

$$U_{non-bonded} = U_{vdW} + U_{Coulomb} \quad (2.7)$$

Van der Waals interactions

The first term of the non-bonded potential, U_{vdW} , describes the attractive and repulsive

interactions between non-bonded atoms. It is commonly expressed as the Lennard-Jones potential:

$$U_{vdW} = \sum_{pairs} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.8)$$

where ϵ_{ij} is the potential well depth, σ_{ij} is the collision diameter (the distance where the Lennard-Jones potential is zero), and r_{ij} is the distance between two atoms i and j .

At short distances, the atomic repulsive forces lead to large positive values of U_{vdW} , converging to infinity as the distance approaches zero. At longer distances, the attractive London dispersion forces lead to small negative values that converge to zero as the distance approaches infinity. Since van der Waals interactions are short-ranged, a cutoff and a switching distance are defined in order to decrease the computational cost in molecular dynamics simulations. Beyond the cutoff distance, van der Waals interactions are ignored. To guarantee the continuity of the potential, a switching function is applied to U_{vdW} at the switching distance so that it equals to zero at the cutoff distance.

Electrostatic interactions

The second term of the non-bonded potential, $U_{Coulomb}$, arises from the uneven distribution of charge in molecules. The electrostatic interaction between two atoms, in the same molecule or in different molecules, is generally modelled by a Coulomb potential:

$$U_{Coulomb} = \sum_{pairs} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.9)$$

where q_i and q_j are partial atomic charges of atoms i and j , r_{ij} is the distance between the two atoms, and ϵ_0 is the electrical permittivity of vacuum.

Since electrostatic interactions are long-range interactions, the calculation of electrostatic contribution is the most computationally expensive part of the force field in molecular dynamics simulations. Considering larger system sizes and the inclusion of explicit water molecules, the calculation of all possible pairwise interactions would make simulations infeasible. One possible solution to this problem would be to truncate the forces at a defined cutoff distance. However, this method is known to introduce artifacts at and beyond the cutoff

distance.²⁷ The most common method used in overcoming the problem of computing these long-range interactions is the Particle Mesh Ewald (PME) based on an Ewald summation.²⁸ In this method, the electrostatic potential within the cutoff distance is calculated in the real space, whereas beyond the cutoff distance it is computed in the reciprocal space by extrapolating charge positions on a grid through the fast Fourier transformation.

2.2 Molecular Dynamics Simulations

Molecular dynamics (MD) is a computational method where the dynamics of a number of particles is simulated by the numerical integration of Newton's equations of motion for discrete timesteps δt on the femtosecond timescale, in order to obtain a trajectory of positions $\mathbf{r}_i(t)$ and velocities $\dot{\mathbf{r}}_i(t)$ (Equation 2.10). Each point in the phase space represents a unique set of positions and velocities. Initial velocities are usually randomly assigned from a Maxwell distribution for the desired temperature. Initial coordinates of proteins are taken from X-ray/NMR/cryo-EM structures or computational models while setting up biological systems.

Given the initial set of conditions, the net force $\mathbf{F}_i(t)$ acting on a particle is then calculated by summing up the forces that particles exert on each other. These forces arise from the interaction of particles, and hence the changes in the potential energy as a function of the separation distances of particles. Once the net force is calculated for each particle, its position and velocities are updated for a small increment in time and the procedure is repeated. For each particle i , this can be defined as:

$$\mathbf{a}_i(t) = \ddot{\mathbf{r}}_i(t) = \frac{d^2\mathbf{r}_i(t)}{dt^2} = \frac{\mathbf{F}_i(t)}{m_i}, \text{ where } \mathbf{F}_i(t) = -\frac{\partial U(\mathbf{r})}{\partial \mathbf{r}_i(t)} \quad (2.10)$$

where $\mathbf{F}_i(t)$ is the force exerted on particle i , $U(\mathbf{r})$ is the potential energy of the system, \mathbf{r}_i and m_i are the position and mass of the particle.

In MD, this numerical integration is performed using timesteps that are smaller than the fastest characteristic vibration of the system under investigation. In atomistic biological systems, the fastest motion corresponds to the covalent bond between a heavy atom and a hydrogen, which is why the typical timesteps for classical atomistic MD are in the order of 1 fs.

Coarse-grained representation of biological systems explained in Section 2.2.3 allows an increase in the timestep, where timesteps of 15-30 fs can be used depending on the biomolecules used in the system. This reduces the simulation wall-clock time and enables reaching higher time and length scales. Coarse-graining also leads to the loss of degrees of freedom and therefore enables smoothening the free energy landscape and eliminating fine-scale free energy barriers.

Several numerical integration algorithms have been developed for molecular dynamics calculations, where the Verlet algorithm is the most widely used.²⁹ Where the acceleration of a particle i is defined as in Equation 2.10, the new position of the particle is obtained using a Taylor expansion at time $t+\delta t$ and $t-\delta t$:

$$\begin{aligned}\mathbf{r}_i(t + \delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t) + \frac{1}{2}\mathbf{a}_i(t)\delta t^2 \\ \mathbf{r}_i(t - \delta t) &= \mathbf{r}_i(t) - \mathbf{v}_i(t) + \frac{1}{2}\mathbf{a}_i(t)\delta t^2\end{aligned}\tag{2.11}$$

Summing these two equations, we can obtain:

$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \mathbf{a}_i(t)\delta t^2\tag{2.12}$$

According to this equation, the Verlet algorithm uses the position and acceleration at time t , along with the position at time $t - \delta t$ to compute the new position at time $t + \delta t$. The Verlet algorithm does not use velocities; however, they can be computed if needed using:

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2\delta t}\tag{2.13}$$

MD simulations provide detailed information about the fluctuations and conformational changes of biomolecules, therefore allowing the investigation of structure, function and dynamics of biological systems. MD simulations provide an accurate dynamic view of molecular behavior, serving as a link between microscopic intermolecular interactions and experimental macroscopic quantities. Nowadays, molecular dynamics simulation timescales can reach micro/milliseconds, and system sizes can go up to tens of millions of atoms. It has therefore become a prominent tool in biology, chemistry and physics, to generate new

hypotheses that pave the way for new experiments, or to confirm existing hypotheses about the structural and dynamic mechanisms of model systems.

2.2.1 The Ergodic Hypothesis

Classical statistical mechanics considers Hamiltonian systems (systems where the equations of motion can be derived from a Hamiltonian principle), where the Hamiltonian, H , is equal to the sum of kinetic and potential energies of a system, and therefore it is equal to the total energy:

$$H = H(\mathbf{r}, \mathbf{p}) = K(\mathbf{p}) + U(\mathbf{r}) \quad (2.14)$$

where \mathbf{r} and \mathbf{p} are respectively the position and the momentum of a particle, $K(\mathbf{p})$ is the kinetic energy, and $U(\mathbf{r})$ is the potential energy.

In statistical mechanics, experimental observables are defined as ensemble averages:

$$\langle A \rangle_{ensemble} = \iint d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) \quad (2.15)$$

where $A(\mathbf{p}^N, \mathbf{r}^N)$ is the observable of interest expressed as a function of \mathbf{p} and \mathbf{r} (the momenta and positions of the system), and $\rho(\mathbf{p}^N, \mathbf{r}^N)$ is the probability density of the ensemble.

On the other side, in molecular dynamics, a physical quantity A is calculated as the time average of A along the trajectory:

$$\langle A \rangle_{time} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\mathbf{p}^N(t), \mathbf{r}^N(t)) dt \approx \frac{1}{M} \sum_{t=1}^M \int_{t=0}^{\tau} A(\mathbf{p}^N, \mathbf{r}^N) \quad (2.16)$$

where τ is the total simulation time, M is the number of timesteps in the simulation.

The ergodic hypothesis states that the ensemble average equals to the time average:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \quad (2.17)$$

The ergodic hypothesis allows the calculation of thermodynamic properties from molec-

ular dynamics simulations and it is therefore the basis of molecular dynamics. If one allows the system to evolve in time indefinitely, the system will eventually pass through all states. Even though molecular dynamics simulations cannot be infinitely long, it is important that one ensures to sample a sufficient amount of the phase space, taking into account the characteristic timescale of the observable under investigation.

2.2.2 Periodic Boundary Conditions

MD simulations are typically performed on a system where the atoms are confined to a simulation box of a specific size and shape. Since molecules found on the surface of the box would experience different forces than the other molecules, MD simulations necessitate the proper treatment of this surface effect. It is most common to use periodic boundary conditions (PBC), where the simulation box is virtually replicated in all directions, which mimics the presence of an infinite bulk surrounding the model system. In the course of the simulation, when an atom leaves the simulation box at one side, an identical atom enters the box at the same time on the opposite side with the same velocity. Particle Mesh Ewald (PME, previously discussed in Section 2.1.2) requires the use of PBC and charge neutrality of the system in order to accurately calculate the total Coulombic interaction.

PBCs do not completely eliminate the effects of using a finite simulation box. These effects are especially pronounced when the system size is small. It is therefore necessary to set up systems that are large enough in order to avoid the influence of PBC on diffusion and dynamics. The simulation box has to be set large enough to avoid the molecule interacting with its periodic image when using PBC, but sufficiently small not to be very computationally demanding. For biological systems, a 12-15 Å water padding is commonly used when solvating a protein.

2.2.3 Coarse-Grained Representation of Biomolecules

Coarse-graining of biomolecules is a popular method in molecular modeling that replaces atomistic level detail with lower resolution coarse-grained (CG) representation. CG representation enables the exploration of biological processes on time and length scales inaccessible

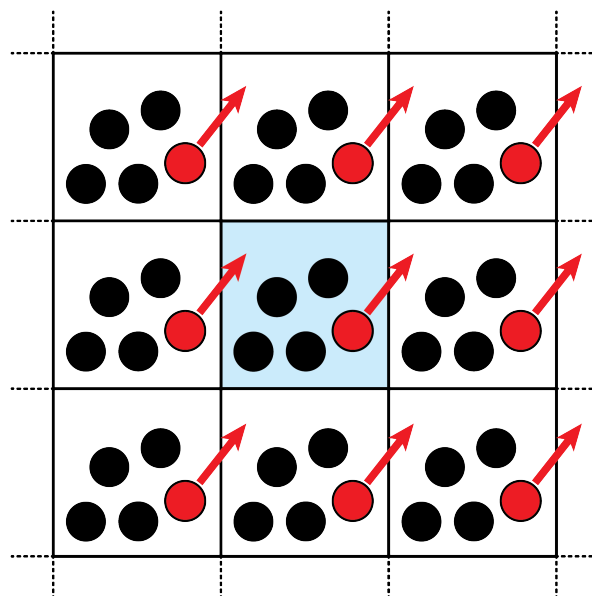


Figure 2.2 – Periodic Boundary Conditions. The particles and trajectories in the central simulation box (blue) are replicated in all directions. When a particle leaves the box at one side (red), it is replaced by an image particle that enters the box on the opposite side with the same velocity.

with an atomic level representation, such as protein-membrane interactions, and membrane protein self-assembly and gating.³⁰ Due to loss of atomistic degrees of freedom, larger integration timesteps, and faster sampling through a smoothened energy landscape, CG systems are faster than atomistic ones by two to three orders of magnitude.³⁰

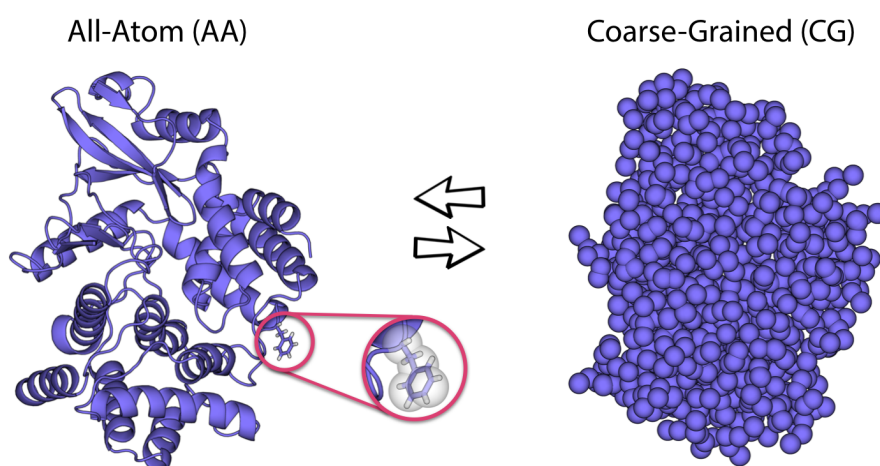


Figure 2.3 – All-Atom and Coarse-Grained (CG) Representations of Proteins. Several atoms are grouped into a single CG particle to switch from an all-atom to a CG representation. The coarse-graining process reduces the number of particles used to describe the system, allowing reaching higher spatial/temporal scales in MD simulations. Through *backmapping* methods, one can also reconstruct the all-atom structure from the CG representation, which enables a multi-scale approach.

The typical potential energy function for a CG system resembles the previously described atomistic molecular mechanics based energy function, expressed as the summation of bonded and non-bonded interactions:

$$U_{CG-total} = U_{pseudo-bond} + U_{pseudo-angle} + U_{pseudo-torsion} + U_{vdW} + U_{Coulomb} \quad (2.18)$$

The pseudo-bond and pseudo-angle potentials for the CG beads are generally expressed as a harmonic potential, while a cosine-series is used to describe the pseudo-torsion potential. For the Martini force field used in this thesis, the pseudo-bond potential can be written as:

$$U_{pseudo-bond} = \sum_{pseudo-bonds} k_r (r - r_{eq})^2 \quad (2.19)$$

where k_r is the force constant, r is the distance between two pseudo-bonded CG beads, and r_{eq} is the equilibrium (reference) pseudo-bond length.

Similarly, the pseudo-angle potential can be written as:

$$U_{pseudo-angle} = \sum_{pseudo-angles} k_\theta [\cos(\theta) - \cos(\theta_{eq})]^2 \quad (2.20)$$

where k_θ is the force constant, θ is the angle between three pseudo-bonded CG beads, and θ_{eq} is the equilibrium angle.

The pseudo-torsion potential can be expressed as:

$$U_{pseudo-torsion} = \sum_{pseudo-torsions} k_\phi [1 + \cos(n\phi - \phi_{eq})] \quad (2.21)$$

where k_ϕ is the force constant, ϕ is the torsional angle between four pseudo-bonded CG beads, and ϕ_{eq} is the equilibrium torsional angle. n is the periodicity of the torsional potential that determines the number of minima in the torsional potential.

The non-bonded interactions are generally composed of van der Waals and electrostatics contributions. The van der Waals potential describes the attractive and the repulsive interactions between non-bonded atoms and is commonly expressed as the Lennard-Jones

potential:

$$U_{vdW} = \sum_{pairs} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.22)$$

where ϵ_{ij} is the potential well depth (interaction strength), σ_{ij} is the distance at which the Lennard-Jones potential is equal to zero, r_{ij} is the distance between two beads i and j .

In addition to van der Waals interactions, charged CG beads also interact through a shifted Coulombic potential expressed as:

$$U_{Coulomb} = \sum_{pairs} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (2.23)$$

where q_i and q_j are partial atomic charges of atoms i and j , r_{ij} is the distance between the two atoms, and ϵ_0 is the electrical permittivity of vacuum. A relative dielectric constant, ϵ_r , is used for screening Coulombic interactions.

The CG-MD simulations performed in this thesis use the Martini 2.2P force field, originally developed for lipids²⁴ and extended to other biomolecules.^{31,32} In the Martini CG model, a four-to-one mapping is used, meaning that four heavy atoms and associated hydrogens are combined to represent one CG particle (Figure 2.4). CG particles exist in four types: nonpolar (N), apolar (I), polar (P) and charged (Q). Each particle type is additionally divided into sub-types based on hydrogen-bonding capabilities (d=donor, a=acceptor, da=donor-acceptor, or 0=none), or degree of polarity (from 1=low polarity, to 5=high polarity). A special class of CG beads (S) is also used to model rings, which have a two/three-to-one mapping.

Currently, Martini parameters exist for a wide range of lipids, peptides, proteins, sugars, polymers, solvents, nucleic acids and nanoparticles (Figure 2.4).³² In addition, the simplistic mapping procedure and available scripts allow the building of topologies for custom peptides and proteins. These features allow a realistic modeling of complex systems with Martini. It is important to note that Martini, like any other model, has its limitations. The current version of Martini necessitates a preset secondary structure and consequently uses structure constraints to maintain helical and extended secondary structure elements throughout the simulation, even though these elements are able to move relative to one another. This means that the

Martini force field does not make it possible to observe conformational changes and protein folding.³²

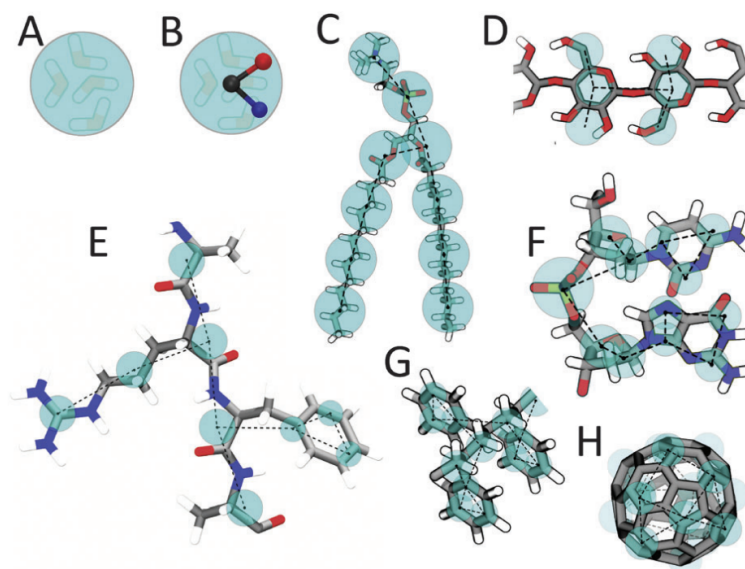


Figure 2.4 – Martini Mapping Examples of Selected Molecules. Martini CG beads are shown in cyan beads overlaying the atomistic structure. (A) CG water particle representing four water molecules. (B) Polarizable CG water particle with embedded charges. (C) DMPC lipid. (D) Polysaccharide fragment. (E) Peptide. (F) DNA fragment. (G) Polystyrene fragment. (H) Fullerene molecule. (Figure by Marrink and Tieleman)³²

2.3 Ligand Docking Methods

Molecular docking is a foundational tool used in structure-based drug design to predict the preferred binding mode of a ligand in the binding site of a protein. The prediction of the binding of a ligand to a protein requires efficient sampling of the conformational space of the flexible ligand/receptor, and the minimization of an accurate fitness function. Ligand docking programs therefore consist of two basic components: an optimization algorithm, and an energy scoring function. The energy scoring function can be distinguished into three types:

(1) Force-field based scoring functions, where molecular mechanics force fields are used to compute the interaction energy and estimate the binding affinity. Examples include the scoring functions of ICM,³³ DOCK,³⁴ GOLD,³⁵ SwissDock,³⁶

(2) Empirical scoring functions, which use protein-ligand complexes with known bind-

ing affinities as a training set. Multiple linear regression analysis is then used to determine the weight constants of a function composed of molecular interaction terms. Examples include the scoring functions of AutoDock Vina,³⁷ GlideScore,³⁸ X-Score.³⁹

(3) Knowledge-based scoring functions, which use interatomic contact frequencies and/or distances in a large database of known protein-ligand complexes. The observed frequency distributions of favorable and unfavorable molecular interactions are used to extract pairwise energy potentials.

A quick and intelligent search over the vast search space necessitates the use of optimization algorithms which find the optimal solution of the computationally demanding fitness function, corresponding to the best ligand pose. Many different optimizers have been implemented in ligand docking programs, the most common ones being evolutionary algorithms (EAs), implemented in AutoDock,⁴⁰ GOLD³⁵ and SwissDock³⁶ and particle-swarm optimizations (PSOs), implemented in PSO@AutoDock.⁴¹

Our in-house integrative modeling and optimization protocol termed parallel optimization workbench for enhancing resolution (*power*),⁴² is a framework that makes use of the MD-based conformational ensemble docking approach for macromolecular structure prediction. It allows the resolution of a large set of optimization problems through the addition of specific modules. It is an object oriented code developed in Python, and supports parallel computing by exploiting MPI libraries. The protocol was shown to be efficient in sampling the protein conformational search space by selecting an ensemble of structures closely resembling that found in nature, as it enabled the near-atomistic arrangement determination of the aerolysin heptamer pore.^{42,43} A new constrained optimizer named *mViE* (memetic viability evolution),^{44,45} has also been recently implemented within this framework.

In mViE, candidate solutions can survive and reproduce if they satisfy a set of viability criteria, defined on the problem objectives and constraints (Figure 2.5). The viability criteria are modified during evolution, where solutions are driven towards desired regions of the search space. Individuals undergo evolution and show adaptive learning capabilities while exploring independently their constrained space. The information learned by multiple of these local search units are combined through a global search operator, where an adaptive

scheduler decides between advancing the local search units or combining them through global search operators.⁴⁵

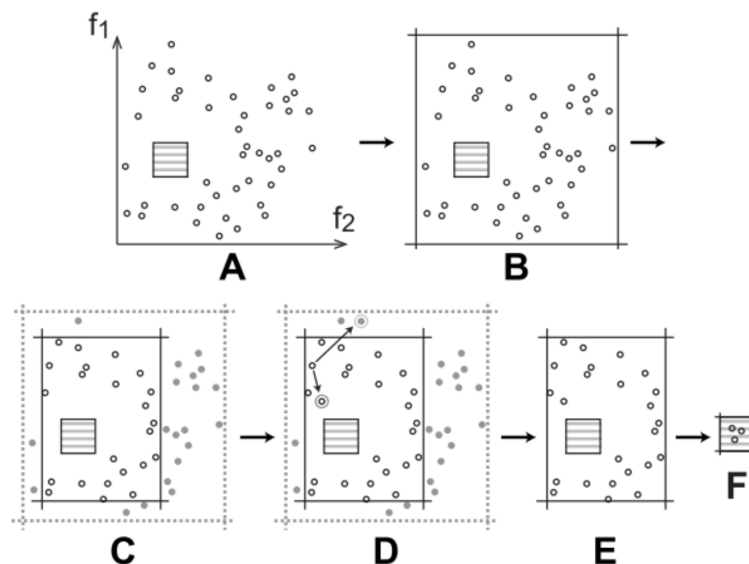


Figure 2.5 – The Viability Evolution Algorithm. The algorithm consists of defining viability boundaries, creating an initial population, repetitive reproduction, elimination and boundary updates until the conditions are satisfied. (A) The population under evolution is in a two dimensional space defined by the objective functions, f_1 and f_2 . Individuals are randomly generated (black circles), where it is very unlikely that an individual can satisfy the target viability conditions (the viability boundaries shown in gray stripes). (B) The viability boundaries are initially set to encompass all individuals. (C) The viability boundaries are tightened gradually towards the target boundaries and individuals that fall outside viability boundaries are eliminated (gray shaded circles). (D) In order to evade the tightened region and to drive further approach to the target boundaries, each individual in the population is allowed to reproduce by making one mutated copy at each iteration. Mutated copies that fall within the viability boundaries are allowed to stay with the parent while the ones that fall outside are marked for elimination. (E) Individuals marked for elimination are removed. (F) The process described in (C-E) is repeated until viability boundaries match the target ones. Finally, the algorithm returns the population of solutions that satisfy the user-defined criteria, and the user may choose one of them. (Figure adapted from Maesani et al.⁴⁵).

mViE was previously successfully used for macromolecular assembly structure prediction in our lab⁴⁶ and is also highly suitable for use in protein-ligand docking, as it enables the identification of suitable conformations from ensembles without the need for an exhaustive exploration of the vast universe of ligand and receptor conformations. In addition, contrary to majority of other evolutionary algorithms where competitively reproducing individuals cause the premature convergence to a sub-optimal solution, viability-based evolution maintains population diversity, resulting in the discovery of alternative solutions. In the Appendix section

of this thesis, a protein-ligand docking protocol recently developed as an extension to the (*pow^{er}*) framework will be presented, where this constrained optimization will be used to carry out an evolutionary ligand docking process.

Well-curated public benchmarks are used by method developers for proper assessment of algorithms and for the establishment of the baseline feasibility of the new technique. Two of the most common benchmarking sets were used in our study: the Astex-85 set,⁴⁷ having high-quality crystal structures of therapeutically relevant targets and drug-like ligands, and the PDBbind dataset,⁴⁸ a collection of experimentally measured binding affinity data for all biomolecular complexes deposited in the Protein Data Bank.

3 Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

This Chapter includes published work from the following manuscripts. While these two interdisciplinary works spanned a large number of techniques and involved work from different labs, the molecular modeling and simulation data were produced by me.

“Conserved Lipid and Small-Molecule Modulation of COQ8 Reveals Regulation of the Ancient Kinase-Like UbiB Family.”

Reidenbach AG, Kemmerer ZA, **Aydin D**, Jochem A, McDevitt MT, Hutchins PD, Wilkerson EM, Stark JL, Stefely JA, Johnson IE, Bingman CA, Markley JL, Coon JJ, Dal Peraro M, Pagliarini DJ. *Cell Chemical Biology* (2018) 25(2): 154-165. doi: 10.1016/j.chembiol.2017.11.001

“Cerebellar Ataxia and Coenzyme Q Deficiency Through Loss of Unorthodox Kinase Activity.”

Stefely JA, Licitra F, Laredj L, Reidenbach AG, Kemmerer ZA, Grangeray A, Jaeg T, Minogue CE, Ulbrich A, Hutchins PD, Wilkerson EM, Ruan Z, **Aydin D**, Hebert AS, Guo X, Freiburger EC, Reutenauer L, Jochem A, Chergova M, Johnson IE, Lohman DC, Rush MJP, Kwiecien NW, Singh PK, Floyd BJ, Forsman U, Sindelar PJ, Westphall MS, Pierrel F, Dal Peraro M, Kannan N, Bingman CA, Coon JJ, Isope P, Puccio H, Pagliarini DJ.

Molecular Cell (2016) 63(4): 608-620. doi: 10.1016/j.molcel.2016.06.030

3.1 Introduction

The ancient UbiB atypical protein kinase-like family is conserved across all three domains of life.⁴⁹ The main function of protein kinase-like superfamily members is to catalyze ATP-dependent phosphorylation.⁵⁰ The founding member of the UbiB family, *E. coli* UbiB, and its orthologs, *S. cerevisiae* Coq8p and human COQ8A (ADCK3) and COQ8B (ADCK4) support the aerobic biosynthesis of the redox-active lipid coenzyme Q (CoQ).⁵¹ *S. cerevisiae* Coq8p and human COQ8A are collectively called COQ8. Mutations in COQ8A and COQ8B result in neurologic and kidney disorders associated with CoQ deficiency, likely via the destabilization of complex Q — a protein complex on the inner mitochondrial membrane that synthesizes CoQ (see Chapter 1, Section 1.3). Despite these clinical findings, the specific role and the biochemical activity of UbiB proteins remain unclear.

Since COQ8 belongs to the protein kinase-like family, it has been proposed that COQ8 might fill a regulatory gap in CoQ biosynthesis by phosphorylating other proteins in the CoQ pathway.⁵² However, to date, the only biochemical activities observed for COQ8 is *cis* autophosphorylation with no *trans* protein kinase activity; and a low-level ATPase activity.^{19,20} This necessitates the exploration of other models for its activity.

Recently, Stefely et al. have solved the crystal structure of the apo human COQ8A^{NΔ254} (Figure 3.1B).¹⁹ The kinase domain of COQ8A adopts an atypical protein kinase-like fold. When compared to a well-characterized protein kinase structure (e.g. protein kinase A (PKA), Figure 3.1A), it has a unique domain, termed the “KxGQ domain”. This domain is composed of an N-terminal extension (GQα1-4, shown in blue in Figure 3.1B) and an N lobe insert (GQα5-6, shown in green in Figure 3.1B) that folds and positions on top of the protein kinase cleft, while the signature KxGQ motif occludes the substrate binding pocket, inhibiting protein kinase activity (Figures 3.1A, 3.1B)).¹⁹

The protein kinase-like domain of COQ8A is preceded by a transmembrane helix on the N-terminal, which is a region rich in small amino acids spaced at *i*, *i*+4 forming an evolutionarily conserved, extended Gly-zipper motif (215-AxxxGxxxGxxxG-227).⁵⁴ It was further shown experimentally that the transmembrane domain of COQ8A self-associates in *E.coli*, via the Gly-zipper motif.⁵⁴ The GxxxG motif is statistically overrepresented in membrane

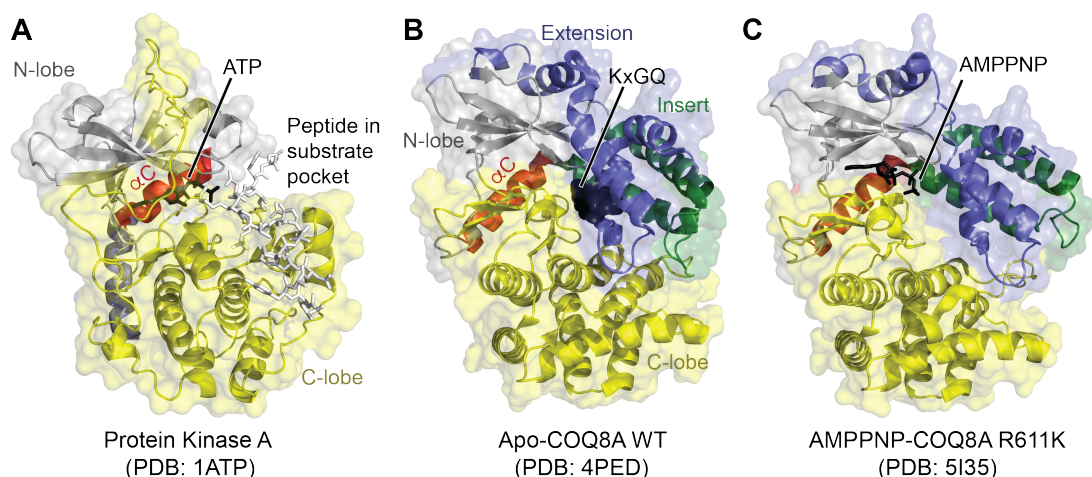


Figure 3.1 – Crystal Structures of Apo and Nucleotide-Bound COQ8A. (A) Crystal structure of protein kinase A (PKA) (PDB: 1ATP)⁵³ colored by domains, with ATP shown in black sticks and the peptide substrate analog in the binding cleft shown in white sticks. (B) Crystal structure of apo COQ8A^{NΔ254} WT (PDB: 4PED)¹⁹ colored by domains, and the KxGQ motif residues represented with black spheres. (C) Crystal structure of COQ8A^{NΔ254} R611K bound to AMPPNP (PDB: 5I35)²⁰ colored by domains, with AMPPNP shown in black sticks.

proteins, and specifically in transmembrane homodimers. The addition of an appropriately spaced small residue (G, A, S) leads to a strong helix packing, and results in the formation of homooligomers.⁵⁵

Characterization of the protein kinase-like and the transmembrane domains of COQ8A have provided a strong foundation for understanding the COQ8 structure and function. Protein kinase-like family members are known to be regulated by lipids and small molecules.^{56–59} Considering that the COQ8A active site is sterically inaccessible to proteinaceous substrates, and that CoQ biosynthesis proteins are localized to the inner mitochondrial membrane, COQ8 might also mediate functional interactions with lipids, similar to other protein kinase-like family members. The role of specific lipids in governing COQ8 function, and how COQ8 localizes to and orients itself on the inner mitochondrial membrane remain unexplored. The examination of how COQ8A interacts with the inner mitochondrial membrane would be important for further investigation of how lipid and membrane interactions might alter or enhance COQ8 activity.

In addition, despite the characterization of the protein kinase-like and the transmembrane domains of COQ8A, much less is known about the behavior of the full-length endoge-

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

nous form of COQ8A. These two domains are connected with a linker domain, short enough to enable the interaction of the protein kinase-like domain with the inner mitochondrial membrane from the matrix side, subsequent to the insertion of the transmembrane domain into the membrane. Since other proteins involved in the biosynthesis of coenzyme Q are localized on the same side of the inner mitochondrial membrane, the identification of the orientation of COQ8A relative to the bilayer might give hints about possible interaction surfaces of this protein with other members of the complex Q.

In this Chapter, we combined computational and biochemical analyses to explore the nature of COQ8A membrane binding and activation. We revealed that COQ8 possesses evolutionarily conserved ATPase activity that is activated by binding to membrane containing cardiolipin, the signature mitochondrial lipid enriched at the inner mitochondrial membrane where COQ8 is known to reside. Our findings advance our understanding of the COQ8 function, and reveal how the positioning of COQ8 on the inner mitochondrial membrane is key to its activation. The biochemical experiments included in this Chapter were led by Dr. Andrew G. Reidenbach, Dr. Jonathan A. Stefely, and Prof. David J. Pagliarini in University of Wisconsin-Madison.

3.2 Methods

This section presents the detailed description of the computational methods used in this study. For experimental methods, please refer to Section 3.5.

3.2.1 Identification and Characterization of Binding Pockets on Nucleotide-Bound and Apo COQ8A

The X-ray structures of apo COQ8A (PDB: 4PED) and nucleotide-bound COQ8A (PDB: 5I35) were subjected to the standard protein preparation procedure implemented in Maestro (Maestro package version 9.8 from Schrödinger, LLC). SiteMap⁶⁰ module in Maestro was used with its default settings to identify the top 5 ranked binding sites on the two structures. SiteMap first locates binding sites by grouping site points that are in spatial proximity to the protein and have favourable interaction energies with the protein. Consequently, for each returned site,

it constructs hydrophilic (further divided into donor and acceptor) and hydrophobic maps, and evaluates pockets through their physicochemical properties. The top two ranking pockets found on nucleotide-bound COQ8A are associated with the signature KxGQ domain of COQ8A, hence were named KxGQ Pocket 1 and KxGQ Pocket 2. The corresponding pockets on apo COQ8A ranked the first and the fourth, respectively. Table 3.3 summarizes the characteristics of these pockets.

3.2.2 Coarse-Grained Molecular Dynamics Simulations of Protein-Membrane Systems

All simulations were performed using the GROMACS⁶¹ simulation package version 5.0.4. The systems were described with the MARTINI 2.2P^{24,62} coarse-grained (CG) force field, together with the ElNeDyn⁶³ approach to maintain the secondary and tertiary structure of the protein. Coarse-grained molecular dynamics (CG-MD) simulations based on the MARTINI force field have been previously used with success for different systems to investigate protein-lipid interactions.⁶⁴⁻⁶⁷

The systems were built with center of mass of COQ8A (PDB: 4PED) 70 Å away from the surface of the membrane and checked for whether the protein approaches and interacts with the membrane. Inner mitochondrial membrane (IMM) models were built according to the experimental molecular ratio of CL observed for bovine heart mitochondria, namely 15 to 20% of the phosphorus content.⁶⁸ Therefore, IMM mimics were featuring a lipid concentration of 16% CL/41% POPC/37% POPE/6% DSPC, while generic membrane models featuring 100% POPC lipids were built as control. The IMM model was approximated to have a symmetrical concentration of lipids between the two leaflets since precise lipidomic information about concentration asymmetry is not known. Lipid bilayers of 130 x 130 Å² were generated using the insane (INSert membrANE)⁶⁹ method of MARTINI. Systems were solvated with polarizable CG water and counter ions were added to neutralize the system. Following solvation, systems were energy minimized with a timestep of 5 fs. Successive equilibrations with decreasing restraints were performed in order to obtain a fully equilibrated system (force constants of 1000, 500, 250 and 0 kJ/mol applied on protein particles, 10 ns of run with a timestep of 10 fs for each). In the production phase, which is 4 μs for all simulations, the protein, membrane

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

and the aqueous phase (water and ions) were coupled independently to an external bath at 310 K by using the v-rescale thermostat. The pressure was coupled to a pressure bath at 1 bar by semi-isotropic Parrinello-Rahman pressure coupling. Three MD repeats with randomized initial velocities were performed for each membrane system and they yielded consistent results. Average occupancy of cardiolipins was calculated using VMD's VolMap plugin.⁷⁰

3.2.3 Modeling of Transmembrane Domain-Containing COQ8A

Khadria et al. have created 5 models for the COQ8A transmembrane domain dimer through a computational prediction algorithm, where Model 2 is reported as the best structural candidate according to the ranking based on the energy scores, and the agreement of computational and experimental mutagenesis data.⁵⁴ The chain A of the best model, Model 2, was used as an initial model for the transmembrane domain of our structure. The 24 residue long linker sequence was added to the C terminus of the helix with PyMOL.⁷¹ The linker was modeled as a random coil, based on the secondary structure prediction obtained from PSIPRED.⁷² The atomistic model of the transmembrane+linker domain was then converted to a CG representation using the MARTINI mapping.²⁴ An early timestep from the previously discussed CG simulation of the protein kinase-like domain of COQ8A with the 16% CL membrane was used as the initial conformation in order to start with an equilibrated membrane, where the protein does not interact with the membrane yet, allowing space for the protein to be linked to the transmembrane+linker domain. The transmembrane domain was inserted in the membrane manually in VMD⁷⁰ by shifting some lipid molecules to create space in the membrane, in a location close to the protein kinase-like domain.

3.2.4 Identification of Residues Important for Membrane Interaction and Selection of Residues for Mutagenesis Experiments

In Section 3.3.2, the interaction surface of COQ8A with the mitochondrial membrane is identified as the positively charged surface of COQ8A, mainly comprising the GQ α 1, GQ α 4 and GQ α 5 domains. In order to identify specific residues important for the membrane interaction of COQ8A, kinase domain residues in the CG-MD simulation were ranked based on their interaction frequencies with the membrane (Table 3.1). The interaction frequency of a residue

j with a lipid molecule L , noted ξ_L^j , is the time average of instantaneous contacts (Eq. 3.1), starting from the frame where the first contact is recorded (t_0) until the end of the simulation (T). A contact is counted when a charged/polar residue is within the first interaction shell cutoff, $\delta_L = 6.5 \text{ \AA}$ of a phosphate head (Eq. 3.2).

$$\xi_L^j = \frac{1}{T - t_0} \sum_{t=t_0}^T I_L^j(t) \quad (3.1)$$

$$I_L^j(t) = \begin{cases} 1, & \text{if } d_{jL}(t) \leq \delta_L \\ 0, & \text{if } d_{jL}(t) > \delta_L \end{cases} \quad (3.2)$$

Table 3.1 – Ranking of Charged/Polar Residues According to their Interaction Frequencies with the Membrane. Positively charged residues are highlighted in gray.

Residue	Helix	ξ_L^j
R262	$\alpha 1$	0.924
T266	$\alpha 1$	0.810
N259	$\alpha 1$	0.798
S370	$\alpha 5$	0.756
N374	$\alpha 5$	0.720
N373	$\alpha 5$	0.628
K269	$\alpha 1$	0.620
Q366	$\alpha 5$	0.601
S367	$\alpha 5$	0.594
N369	$\alpha 5$	0.590
K314	$\alpha 4$	0.589
Q311	$\alpha 4$	0.538
R265	$\alpha 1$	0.523
N318	$\alpha 4$	0.517
S382	$\alpha 5$	0.513
N317	$\alpha 4$	0.505
K310	$\alpha 4$	0.478
N380	$\alpha 5$	0.187

The ranking based on interaction frequencies revealed that GQ $\alpha 1$ and GQ $\alpha 4$ helices interact with the membrane through the positively charged Arg and Lys residues (highlighted in gray in Table 3.1), whereas the membrane interaction surface of GQ $\alpha 5$ comprises polar residues Gln, Ser and Asn. In order to narrow down the list of residues important for membrane

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

interaction, we have defined 4 mutants as summarized in Table 3.2, where each mutant is named after the mutated or electrostatically knocked out domain for ease of understanding.

Table 3.2 – Designed COQ8A Mutants. The residues listed for each mutant are either (i) mutated to alanines (in electrostatic maps (Figure 3.5A) and in mutagenesis experiments (Figure 3.5D-G)), or (ii) altered by the removal of charges of the polarizable CG beads, to knockout Coulomb interactions without changing the residue types (in CG-MD simulations (Figure 3.5B)).

Mutant	Altered Residues
GQ α 1 mutant (Triple A)	R262, R265, K269
GQ α 4 mutant	K310, K314
GQ α 1 α 4 mutant	R262, R265, K269, K310, K314
GQ α 5 mutant	Q366, S367, N369, S370, N373, N374

3.2.5 Modeling of COQ8A Mutants

For Electrostatic Surface Potential Analysis: Atomistic models of the different GQ α mutated COQ8A proteins were built in PyMOL⁷¹ by mutating the following residues listed in Table 3.2 to alanines. The effect of these mutations on the electrostatic potential surface of the protein was visualized through the ICM-REBEL module of the ICM program.³³

For CG-MD Simulations: Residues listed in Table 3.2 for each mutant were altered by the removal of charges of the polarizable CG beads, to knockout Coulomb interactions without changing the residue types. CG-MD simulations were then performed for these four mutants following the procedure explained in Section 3.2.2.

3.3 Results

3.3.1 Nucleotide Binding to COQ8A Opens Small-Molecule Pockets

The apo COQ8A structure previously showed that its UbiB family specific KxGQ domain occludes the typical protein kinase substrate binding pocket (Figures 3.1A, 3.1B).¹⁹ The apo structure does not have any obvious hydrophobic binding pockets that could bind a lipid. Nevertheless, nucleotide binding to COQ8A could open a substrate binding pocket. To test this idea, our collaborators crystallized COQ8A^{N Δ 254}R611K with an ATP analogue, adenosine 5'-(β,γ -imido)triphosphate (AMPPNP), and solved its X-ray structure at a resolution of 2.3 Å (Figure 3.1C).

Overall, the apo and nucleotide-bound COQ8A structures are largely similar (RMSD=3.3 Å; Figures 3.1, 3.2A). Even with a nucleotide bound, the KxGQ domain occludes the typical peptide substrate binding site and hinders *trans* protein phosphorylation. The comparison of apo and nucleotide-bound COQ8A shows that the most significant difference between the two structures arises from deviations in two loops in the KxGQ domain: the GQ α 2-GQ α 3 loop and the GQ α 5-GQ α 6 loop (Figure 3.2A). The movement of these two regions due to nucleotide binding opens two hydrophobic pockets in the KxGQ domain (KxGQ pockets 1 and 2, Table 3.3), which are not present in the apo COQ8A structure (Figure 3.2B). KxGQ pocket 1 is near the putative phosphoryl acceptor substrate binding region, and it could potentially bind the kinase's substrate. In contrast, KxGQ pocket 2, which is partially formed by the "x" of the KxGQ motif, has the potential to bind an allosteric effector molecule that could communicate with the active site through the KxGQ motif. KxGQ pocket 2 is additionally very close to the membrane interacting surface of COQ8A (the positively charged surface comprising GQ α 1, GQ α 4 and GQ α 5, see Section 3.3.2). This pocket might be accessible to a CoQ pathway intermediate residing in the membrane, which would enable the head group to interact with the protein while the hydrophobic lipid tail would stay embedded in the membrane.

Table 3.3 – Characterization of Binding Pockets on Nucleotide Bound- and Apo-COQ8A.

Property	Remarks	Nucleotide Bound-COQ8A		Apo-COQ8A	
		KxGQ Pocket 1	KxGQ Pocket 2	KxGQ Pocket 1	KxGQ Pocket 2
SiteScore	>1 is promising	1.092	1.057	1.051	0.689
Dscore	Distinguishes "druggable" pockets	0.949	1.114	0.989	0.598
Exposure	Lower is better. Average for good binding is 0.49	0.436	0.535	0.460	0.721
Enclosure	Higher is better. Average for good binding is 0.78	0.835	0.703	0.774	0.570
Volume (Å ³)		1079.42	230.50	360.49	120.39
Hydrophobic/ philic balance		0.402	1.815	0.379	0.015
Ranking		1	2	1	4

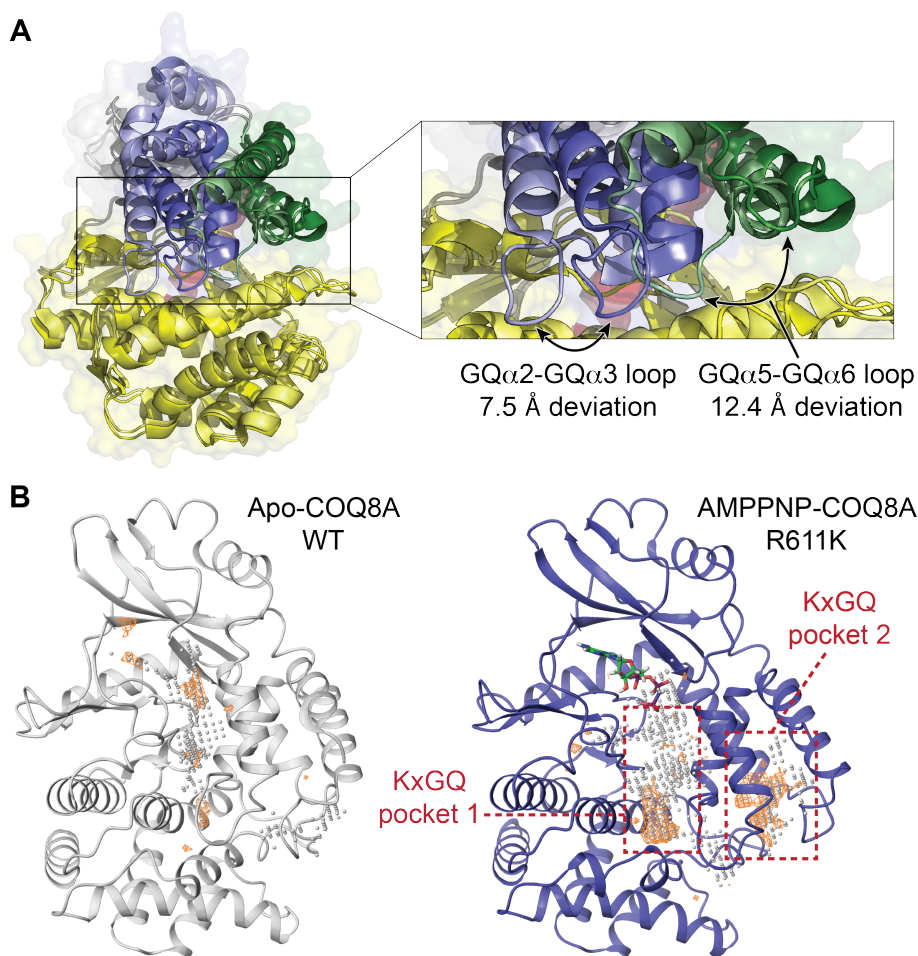


Figure 3.2 – Nucleotide Binding to COQ8A Opens Small-Molecule Pockets. (A) Superposition of apo COQ8A WT (PDB: 4PED, light colors) and AMPPNP-COQ8A R611K (PDB: 5I35, dark colors), where the region that differs the most is highlighted in the right panel. (B) Two top-ranking binding pockets identified by SiteMap on the nucleotide-bound COQ8A structure (PDB: 5I35, in blue cartoon representation) and the corresponding pockets on the apo COQ8A structure (PDB: 4PED, in gray cartoon representation). The accessible areas of the pockets are shown in white beads and the hydrophobic regions are shown in orange mesh.

3.3.2 COQ8 Interacts with the Membrane through its Signature KxGQ Domain

Mitochondria are characterized by a high percentage of cardiolipins (CL), a unique, anionic, four-tail lipid that significantly influences the correct functioning of mitochondrial membrane proteins.^{73–75} Mature COQ8 resides on the inner mitochondrial membrane (IMM) facing the matrix where the other complex Q proteins reside and where it can potentially interact with CoQ intermediates.^{76–78} This suggests that membrane association might be an important

modulator of its activity, as has been shown for other membrane-bound kinases.⁷⁹ To determine the preferential interaction surface of COQ8 with the mitochondrial membrane, we analyzed the electrostatic surface of the protein kinase-like domain of COQ8A. Examination of the electrostatic surface potential of the soluble domain of COQ8A revealed an extensive positive patch, generated primarily by the signature “KxGQ domain”. This domain is formed by the GQ α 1 (residues 260-270), GQ α 4 (residues 309-320) and GQ α 5 (residues 366-381) helices; however, the main contributors to the positive patch are GQ α 1 and GQ α 4 (Figure 3.3A). We hypothesized that this positive surface on COQ8A likely interacts with negatively charged phospholipids found in the IMM, such as cardiolipins.

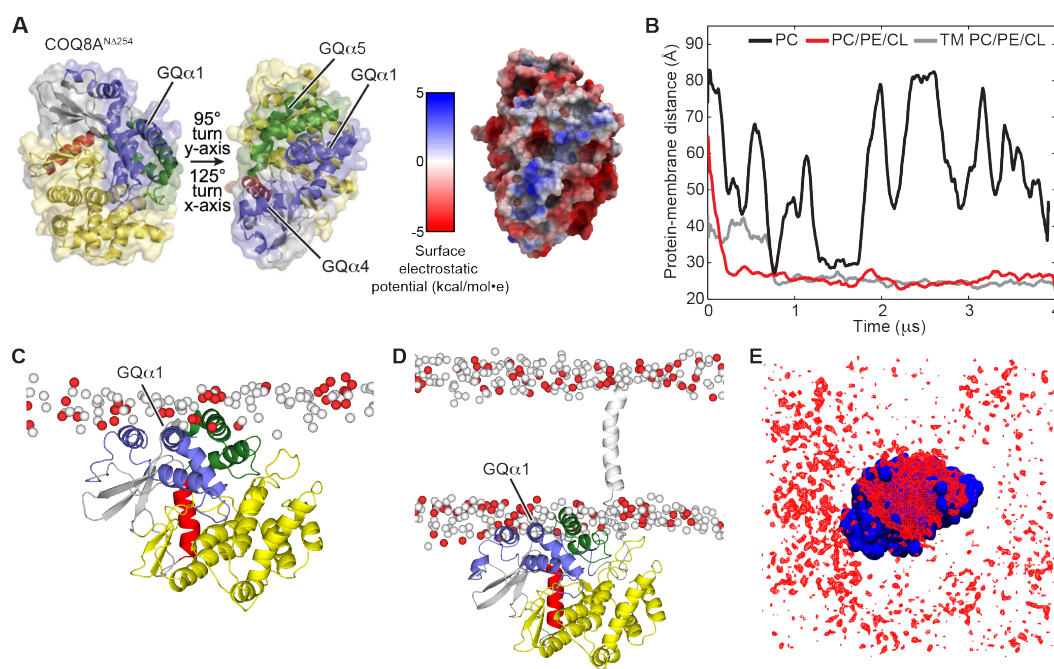


Figure 3.3 – Cardiolipin Enhances COQ8A Membrane Binding. (A) The two structures of COQ8A^{NΔ254} on the left are colored according to Stefely et al. (2015) and turned to show the orientation of the electrostatic maps. The KxGQ domain is colored in blue and green in the structures on the left. On the right, the electrostatic map of COQ8A^{NΔ254} shows the positively charged region spanning GQ α 1, GQ α 4, and GQ α 5 (negative [-5 kcal/(mol·e) charge] red, via white, to positive [+5 kcal/(mol·e) charge] blue). (B) Time evolution of the distance between the center of mass of the protein and the center of mass of the phosphate heads of the leaflet it interacts with, for CG-MD simulations of (i) COQ8A^{NΔ254} with a PC membrane (black) or (ii) with a PC/PE/CL membrane (red), and (iii) the COQ8A^{NΔ162} model (transmembrane (TM) domain-containing COQ8) with a PC/PE/CL membrane (gray). (C) Snapshot from a CG-MD simulation ($t \sim 4 \mu\text{s}$) showing the interaction of COQ8A^{NΔ254} with a PC/PE/CL membrane. CL phosphate heads are shown in red, and PC/PE phosphate heads are shown in white. (D) Snapshot from a CG-MD simulation ($t \sim 4 \mu\text{s}$) showing the interaction of the COQ8A^{NΔ162} model (TM domain-containing COQ8) with a PC/PE/CL membrane. CL phosphate heads are shown in red, and PC/PE phosphate heads are shown in white. (E) Average occupancy of CL phosphate heads (red) when COQ8A^{NΔ254} (blue) is centered throughout the trajectory.

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

To explore the nature of COQ8 membrane binding and activation, we used molecular dynamics to simulate the binding of COQ8A to membranes. We used coarse-grained molecular dynamics (CG-MD) simulations to investigate the binding of COQ8A to PC/PE/CL (47/37/16%), mimicking inner mitochondrial membrane (IMM)-like membranes, or to PC alone, mimicking a non-IMM membrane model. In the simulations performed with PC/PE/CL membrane composition, the protein initially diffuses in the aqueous environment and subsequently binds to the lipid bilayer through the previously identified positive electrostatic patch (Figure 3.3B, 3.3C). Invariably, the presence of CL at physiological concentration (Horvath and Daum, 2013) is required to induce rapid (within the first microsecond) membrane association (3.3B). While a stable interaction with the membrane is seen with the 16% CL membrane, in simulations featuring a pure POPC bilayer, even though the protein approaches the bilayer several times, no stable binding event is observed (Figure 3.3B). This suggests that membrane interaction is strongly influenced by CL. Interestingly, anchoring of COQ8A at the IMM is associated with local reorganization of lipids, featuring a strong segregation of CL species to the COQ8A interface (Figure 3.3E).

Endogenous COQ8 contains a transmembrane domain that anchors it to the IMM.¹⁹ In order to test whether an intact transmembrane domain would still allow the soluble domain to interact with the membrane with the same surface that we have found through CG-MD simulations of the soluble domain, we modelled the transmembrane domain-containing COQ8A, including both the transmembrane and kinase-like domains. In the CG-MD simulations of the full model, COQ8A interacts with the PC/PE/CL membrane with the same positively charged region (Figure 3.3D). Even though the linker region between the transmembrane and soluble domains was modelled as an extended loop in the initial state, during the simulation it rearranges to allow the membrane interaction of the soluble domain.

3.3.3 Cardiolipin Specifically Increases COQ8 Membrane Interaction and ATPase Activity

To validate the computationally observed membrane association of COQ8A, and to test the effect of membrane association on its activity, our collaborators first used a liposome flotation assay to determine how effectively COQ8 binds to liposomes. Without the transmembrane

domain, COQ8A^{NΔ250} exhibited moderate affinity for PC/PE/CL liposomes; however, the addition of the transmembrane domain enabled near-complete COQ8A^{NΔ162} and Coq8p^{NΔ41} liposome binding (Figure 3.4A). Strikingly, this association markedly enhanced the ATPase activity of COQ8A^{NΔ162} and Coq8p^{NΔ41} (Figure 3.4C), which is several orders of magnitude higher than autophosphorylation (Figure 3.4D). These data imply that UbiB proteins, like other protein kinase-like family members, may be activated endogenously by binding to particular membrane environments.

To determine whether the COQ8 membrane binding and activity is driven by a particular membrane component in these liposome flotation assays, we tested the ability of multiple individual lipids to activate COQ8 when reconstituted in PC carrier liposomes. Remarkably, from among the eight individual species tested, only CL was able to activate COQ8 (Figure 3.4E). This increase in activity was matched by the superior ability of CL to facilitate binding of COQ8 to liposomes, indicating that CL activation of COQ8 is directly linked to its ability to mediate COQ8-membrane interactions (Figure 3.4B). This experimental result confirmed our computational finding that CL mediates membrane interaction of COQ8. Interestingly, COQ8 lacking its transmembrane domain still preferentially bound to CL-containing liposomes (Figure 3.4B) but was not activated by them (Figure 3.4C), suggesting that the transmembrane domain itself is a regulatory feature of COQ8.

To test whether liposome/CL binding specifically enhances the ATPase activity of COQ8, we performed [$\gamma^{32}\text{P}$]ATP kinase reactions with liposomes to follow the location of the gamma phosphate. We also assessed autophosphorylation using SDS-PAGE and potential phosphorylation of copurifying lipid substrates or CL using thin-layer chromatography. In these experiments, no autophosphorylation or lipid phosphorylation was observed, further supporting the hypothesis that COQ8 acts as an ATPase (Figure 3.4F). Collectively, these data demonstrate that CL enables COQ8 membrane binding and enhances COQ8 ATPase activity.

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

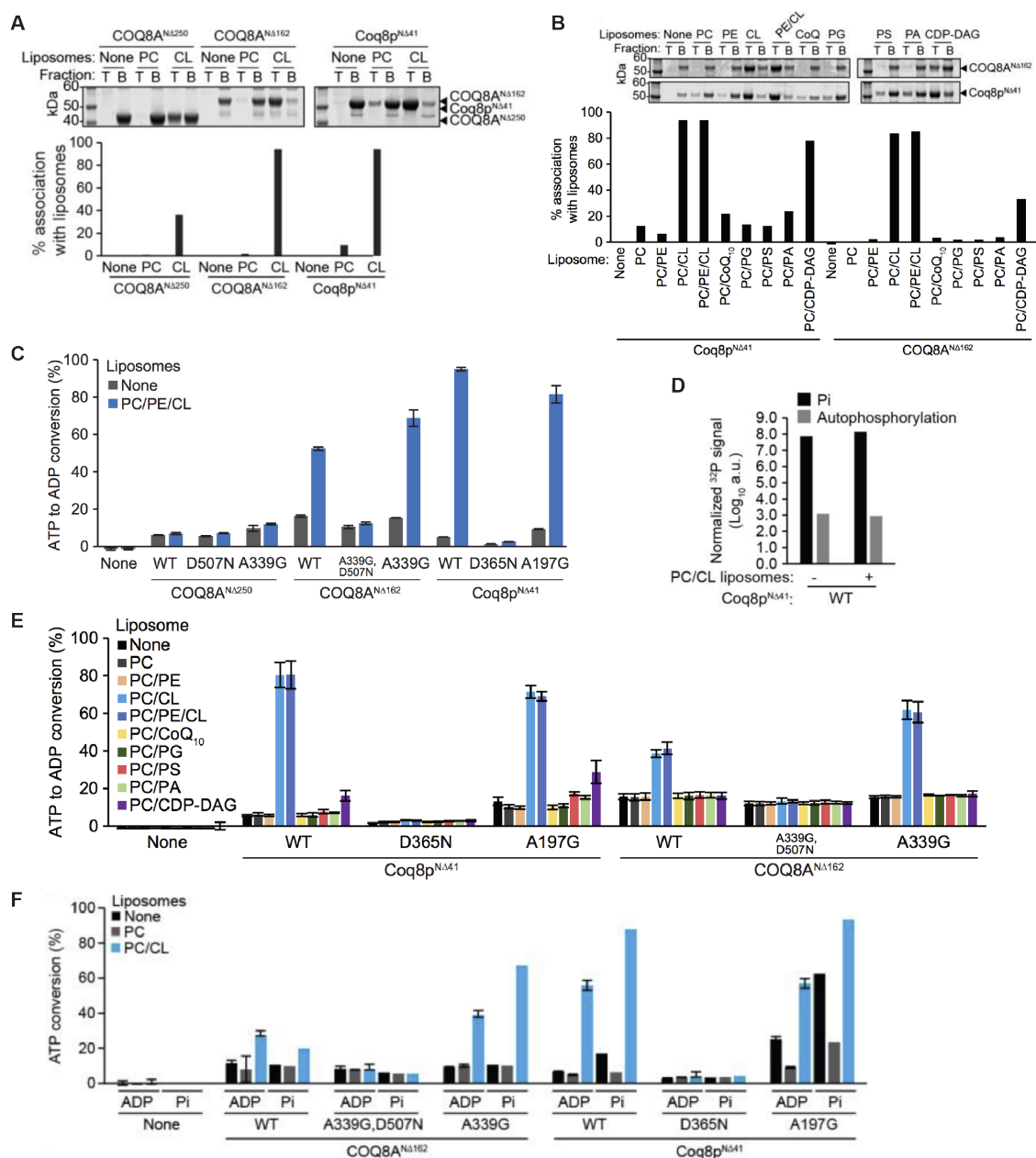


Figure 3.4 – Cardiolipin Enhances the ATPase Activity and Liposome Binding of COQ8A. (A) Coomassie-stained SDS-PAGE of the liposome flotation assay with COQ8 variants and PC/PE/CL liposomes. T, top fraction; B, bottom fraction. (B) Coomassie-stained SDS-PAGE and the liposome flotation assay with transmembrane (TM) domain-containing COQ8 and a panel of PC-based liposomes. T, top fraction; B, bottom fraction. (C) ADP-Glo assay with COQ8 variants and PC/PE/CL liposomes. Error bars represent SD of two independent experiments performed in technical triplicate. (D) Direct comparison of Coq8^{NΔ41} ATPase (Pi) and autophosphorylation activities with or without PC/CL liposomes using [γ -³²P]ATP. (E) ADP-Glo assay with a panel of liposomes and TM domain-containing COQ8. Error bars represent SD of two independent experiments performed in technical duplicate. (F) ADP-Glo (ADP produced) and [γ -³²P]ATP ATPase assay (Pi produced) of COQ8 variants with PC or PC/CL liposomes performed in parallel.

3.3.4 Cardiolipin Specifically Increases COQ8 Membrane Interaction and ATPase Activity

Ranking of residues based on their interaction frequencies with the membrane in our CG-MD simulations revealed that GQ α 1 and GQ α 4 helices interact with the membrane through positively charged Arg and Lys residues, whereas the membrane interaction surface of GQ α 5 comprises polar residues Gln, Ser and Asn (Table 3.1). In order to confirm the importance of these specific residues in membrane interaction, we designed 4 mutants (GQ α 1, GQ α 4, GQ α 1 α 4, GQ α 5) as summarized in Table 3.2, where each mutant is named after the mutated or electrostatically knocked out domain for ease of understanding.

Previous examination of the electrostatic potential surface of COQ8A had revealed an extensive positive electrostatic patch which was found to interact with negatively charged cardiolipins (CL) found in the inner mitochondrial membrane. Visualizing the effect of these mutations on the electrostatic surface of the protein revealed that the positive electrostatic surface of COQ8A, which is crucial for membrane interaction, disappears in GQ α 1 α 4, where all positively charged residues interacting with the membrane are altered. In GQ α 1 and GQ α 4, the positive surfaces associated with α 4 and α 1 remain, respectively, which still might enable membrane interaction for both. It is important to note that the surfaces observed in GQ α 1 and GQ α 4 add up to form the positive electrostatic patch of the wild-type (WT) COQ8A. While the alteration of positively charged α 1 and α 4 residues weaken the electrostatic potential of the interaction surface, the alteration of polar α 5 residues does not alter the surface. Overall, from this electrostatics analysis, it is possible to predict that GQ α 1 α 4 will not bind to the membrane, while GQ α 5 will. The comparison of the behavior of GQ α 1 and GQ α 4 mutants in the CG-MD simulations will reveal whether α 1 and α 4 have a complementary effect on membrane interaction or whether knocking out one of them will suffice to eliminate the interaction of ADCK3 with the membrane.

CG-MD simulations were performed with the soluble kinase-like domain of COQ8A for the 4 designed mutants (Figure 3.5B), and as predicted in the electrostatic surface potential analysis, the alteration of α 1+ α 4 was found to prevent binding to CL-containing membranes. The alteration of α 5 did not affect membrane association. Furthermore, individual assessment

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

of $\alpha 1$ and $\alpha 4$ domains revealed that the alteration of $\alpha 1$ alone is enough to remove binding, whereas the alteration of $\alpha 4$ is not. Overall, this analysis suggests that $\alpha 1$ is the indispensable domain for membrane binding and that it compensates for the electrostatic knockout of $\alpha 4$.

$\alpha 1$ residues might be important for (i) driving the initial encounter with the membrane through electrostatic interactions to ensure the correct binding surface or (ii) the subsequent interaction with the membrane to establish a stable orientation and to retain the protein on the surface. To substantiate the role of the $\alpha 1$ domain and to assess the importance of specific $\alpha 1$ residues in membrane binding we checked the time evolution of the distance between the polarizable CG bead of positively charged $\alpha 1$ and $\alpha 4$ residues and the closest phosphate head (Figure 3.5C). R262 is the first residue to interact and also to establish a strong interaction with the membrane, specifically with cardiolipins. The interaction of $\alpha 4$ residues with the membrane takes place at a later stage in WT (920 ns), supporting the conclusion that $\alpha 1$ residues (in particular R262) are the crucial residues for the membrane interaction of ADCK3.

Collectively, CG-MD simulations enabled the prediction that three positively charged residues of the GQ $\alpha 1$ helix (R262, R265, and K269) are the key anchor points driving the initial encounter of the soluble domain with the membrane via electrostatic interactions with cardiolipin phosphate heads (Figure 3.5B and 3.5C). To validate the importance of these residues, our collaborators purified a triple mutant of COQ8A^{N Δ 250} (R262A,R265A,K269A; Triple A) and measured its ability to bind liposomes. Indeed, this mutant demonstrated a large decrease in liposome association (Figures 3.5D and 3.5E).

When the corresponding mutations were introduced into COQ8, its ability to rescue $\Delta coq8$ respiratory deficiency was significantly diminished, suggesting that these residues may have important ramifications for membrane binding and/or orientation in vivo (Figures 3.5F and 3.5G). Altogether, our simulations and supporting experiments reinforce the importance of CL for COQ8 membrane association and suggest that the COQ8-CL interaction is driven by conserved residues in the signature KxGQ domain.

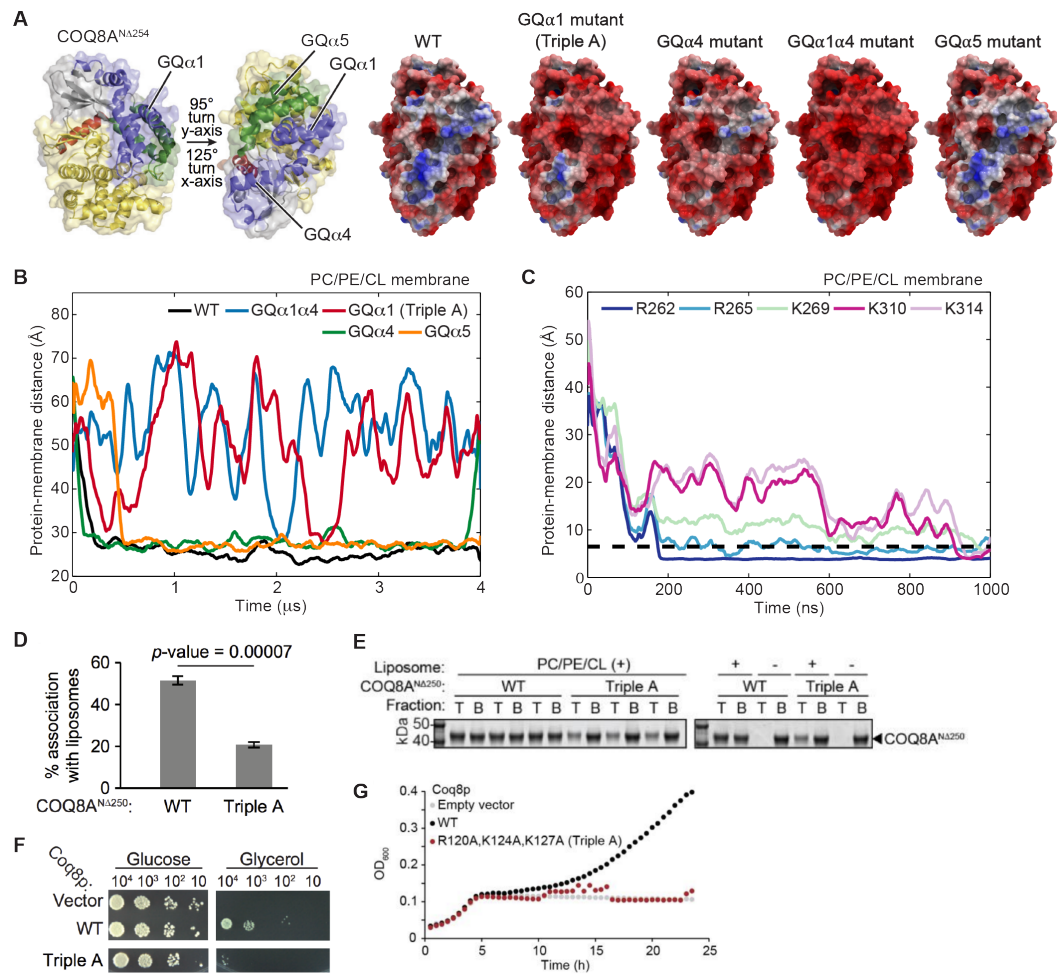


Figure 3.5 – COQ8 Interacts with the Membrane through Its Signature KxGQ Domain. (A) The two structures of COQ8A^{NΔ254} on the left are colored according to Stefely et al. (2015) and turned to show the orientation of the electrostatic maps. The KxGQ domain is colored in blue and green in the structures on the left. On the right, there are electrostatic maps of COQ8A^{NΔ254} showing the positively charged region spanning GQα1, GQα4, and GQα5 (negative [-5 kcal/(mol·e) charge] red, via white, to positive [+5 kcal/(mol·e) charge] blue), illustrating the effect of alanine mutations on different regions of the positive electrostatic patch proposed as the membrane interaction surface of COQ8A. (B) CG-MD simulations of different GQα mutated COQ8A proteins and PC/PE/CL membranes. Time evolution of the distance between the center of mass of the protein and the center of mass of the phosphate heads of the leaflet with which it interacts. GQα1 (Triple A) mutant, R262A, R265A, K269A; GQα4 mutant, K310A, K314A; GQα1α4 mutant, R262A, R265A, K269A, K310A, K314A; GQα5 mutant, Q366A, S367A, N369A, S370A, N373A, N374A. (C) Time evolution of the distance between positively charged residues on COQ8A GQα1 and GQα4 and the closest phosphate head, during the CG-MD simulation of WT COQ8A protein and PC/PE/CL membrane. (D) Percent protein associating with liposomes from a liposome flotation assay with COQ8A^{NΔ250} WT and the Triple A mutant (R262A, R265A, K269A). Error bars represent SD of three independent experiments. (E) Coomassie-stained SDS-PAGE of top and bottom fraction from the liposome flotation assay in panel D. +, with PC/PE/CL liposomes; -, no liposomes. (F) Serial dilutions of $\Delta coq8$ transformed with Coq8p WT or the Triple A mutant (R120A, K124A, K127A) on synthetic complete glucose (2% w/v) or glycerol (3% w/v) containing media. (G) Representative growth curve of $\Delta coq8$ yeast transformed with WT COQ8 or the Triple A mutant in Ura-, glucose (0.1% w/v) and glycerol (3% w/v) respiratory media.

3.4 Discussion

Coenzyme Q was identified as a key component of the mitochondrial electron transport chain, yet there are several unanswered questions regarding biosynthesis. One unknown feature of its biosynthesis pathway is the unexplained need for multiple auxiliary proteins that have no clear catalytic role in the pathway, including COQ8. Based on primary sequence alignments, COQ8 was initially identified as an atypical protein kinase, proposed to fill a regulatory gap in the CoQ biosynthesis by phosphorylating other proteins in the CoQ pathway.⁵² Contrary to this initial proposition, and despite extensive experimentation, no COQ8 *trans* protein kinase activity has been observed until now, which necessitates the investigation of other models for determining its role in pathway.

Multiple members of the protein kinase-like superfamily become activated by binding to endogenous molecules found at their sites of cellular activity, such as the activation of PKC or Akt by diacylglycerol and phosphatidylinositol 3,4,5-trisphosphate, respectively, at the plasma membrane.^{59,80,81} Furthermore, protein-lipid interactions in general are becoming increasingly recognized as regulatory and structural features of proteins.^{82–84} These indicate that it may be important to investigate the biochemical function of COQ8 in its native environment.

Similar to PKC and Akt, our data also indicate that COQ8 is activated via interaction with a specific lipid, CL. CL is a unique anionic, four-tailed lipid found at particularly high concentrations in the IMM where it influences the activity and stability of various mitochondrial membrane proteins.^{73–75,85} and promotes the formation of respiratory chain supercomplexes.⁸⁶ In addition to mere binding, our CG-MD simulations indicate that CL also directs the orientation of the soluble COQ8 domain along the membrane surface, perhaps thereby enabling COQ8's known interactions with other COQ proteins^{76,87} and/or with membrane-embedded CoQ precursors. Furthermore, CL is thought to be distributed nonuniformly throughout the IMM,⁸⁸ suggesting that the interaction of COQ8 with CL might seed complex Q formation at strategic sub-mitochondrial locations. These observations suggest models whereby the proper positioning of COQ8 at CL-rich domains on the mitochondrial inner membrane activates its ATPase activity as a requisite first step in enabling complex Q assembly

and/or CoQ biosynthesis, and that exposure to CoQ intermediates at those sites can further enhance COQ8 function (Figure 3.6). Within this context, CG-MD is shown once more to be a powerful resource to understand the molecular determinants underlying protein-lipid interplay.

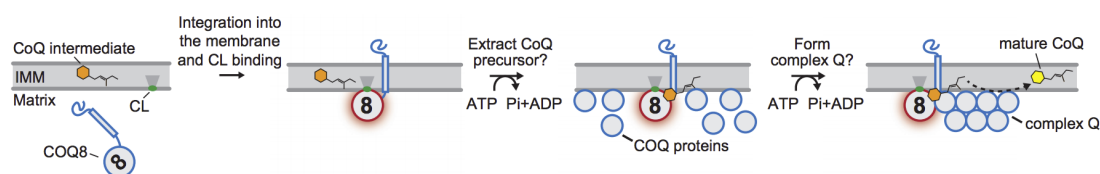


Figure 3.6 – Model for How COQ8 Membrane and CL Binding, and ATPase Activity Could Facilitate CoQ Biosynthesis. First, COQ8 is imported into the matrix. Binding to CL then facilitates COQ8's association with, and insertion into, the IMM with its soluble domain properly oriented along the membrane surface. Next, the CL-induced activation of COQ8 (red glow) allows it to advance CoQ biosynthesis by coupling the hydrolysis of ATP to the extraction of CoQ precursors out of the IMM and/or to the formation of complex Q.

More broadly, these results suggest an important and underappreciated connection between CoQ and CL, two quintessential mitochondrial lipids that enable oxidative phosphorylation. Indeed, CL has long been known to enable electron transfer between CI and CIII,⁸⁹ and in plants, CL was found to have a positive effect on the electron transfer rate from CoQ to the photosynthetic reaction center.⁹⁰ These observations in conjunction with our data suggest a model whereby enhanced CL biosynthesis may augment oxidative phosphorylation function in part by enabling the biosynthesis of CoQ via COQ8.

3.5 Supplementary Information

This Section details the work done by our experimental collaborators to support our computational findings. The experimental work included in this Chapter was led by Dr. Andrew G. Reidenbach and Prof. David J. Pagliarini in University of Wisconsin-Madison.

3.5.1 DNA Constructs and Cloning

Cloning of COQ8A^{NΔ250}, Coq8p^{NΔ41}, PKA, and p426 GPD COQ8 plasmids was previously described.^{19,91} For MBP-UbiB^{CΔ47} and COQ8A^{NΔ162} standard PIPE cloning methods⁹² were

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

used. PIPE reactions were DpnI digested and transformed into DH5 α competent *E. coli* cells. Plasmids were isolated from transformants and DNA sequencing was used to identify those containing the correct constructs. pVP68K, a plasmid for expression of recombinant proteins in bacteria [8His-cytoplasmically-targeted maltose-binding protein (MBP) with a linker including a tobacco etch virus (TEV) protease recognition site fused to the protein construct (8His-MBP-[TEV]-Protein)], has been described previously.⁹³ Oligonucleotides were purchased from IDT (Coralville, IA, USA). For yeast rescue experiments, COQ8 was cloned into the p426 GPD vector.⁹⁴ Point mutations were introduced by PCR-based mutagenesis and confirmed by DNA sequencing.

3.5.2 Recombinant Protein Expression and Purification

Human COQ8A^{N Δ 162} and COQ8A^{N Δ 250}

Method adapted from.^{19,91} COQ8A constructs were overexpressed in *E. coli* by autoinduction.⁹⁵ Cells were isolated and frozen at -80°C until further use. For protein purification, cells were thawed, resuspended in Lysis Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM 2-mercaptoethanol (BME) or 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 0.25 mM phenylmethylsulfonyl fluoride (PMSF)] (4°C). The cells were lysed by sonication (4°C , 75% amplitude, 10 s x 3). The lysate was clarified by centrifugation (15,000 g, 30 min, 4°C). The cleared lysate was mixed with cobalt IMAC resin (Talon resin) and incubated (4°C , 1 h). The resin was pelleted by centrifugation (700 g, 2 min, 4°C) and washed four times with Wash Buffer [50 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM BME, 0.25 mM PMSF, 10 mM imidazole]. His-tagged protein was eluted with Elution Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM BME or 0.5 mM TCEP, 0.25 mM PMSF, 100 mM imidazole]. The eluted protein was concentrated with a MW-cutoff spin filter (50 kDa MWCO) and exchanged into Storage Buffer [20 mM HEPES (pH 7.2), 300 mM NaCl, 10% glycerol, 5 mM BME or 0.5 mM TCEP]. The concentration of 8His-MBP-[TEV]-COQ8A^{N Δ 250} was determined by its absorbance at 280 nm ($\epsilon=96,720\text{ M}^{-1}\text{ cm}^{-1}$) (MW = 89.9 kDa) and 8His-MBP-[TEV]-COQ8A^{N Δ 162} ($\epsilon=96,720\text{ M}^{-1}\text{ cm}^{-1}$) (MW = 99.26 kDa). The fusion protein was incubated with Δ 238TEV protease (1:50, TEV/fusion protein, mass/mass) (1 h, 20°C). The TEV protease reaction mixture was mixed with cobalt IMAC resin (Talon resin), incubated (4°C , 1 h). The resin was pelleted by centrifugation (700

g, 3 min, 4 °C). The unbound COQ8A was collected and concentrated with a MW-cutoff spin filter (30 kDa MWCO) and exchanged into storage buffer. The concentration of COQ8A^{NΔ250} was determined by its absorbance at 280 nm ($\epsilon=28,880 \text{ M}^{-1} \text{ cm}^{-1}$) (MW = 45.6 kDa). The concentration of COQ8A^{NΔ162} was determined by its absorbance at 280 nm ($\epsilon=28,880 \text{ M}^{-1} \text{ cm}^{-1}$) (MW = 54.99 kDa). The protein was aliquoted, frozen in N₂(l), and stored at -80 °C. Fractions from the protein preparation were analyzed by SDS-PAGE.

Yeast Coq8p^{NΔ41}

Coq8p^{NΔ41} plasmid constructs were transformed into RIPL competent *E. coli* cells for protein expression. 8His-MBP-[TEV]-Coq8p^{NΔ41} was overexpressed in *E. coli* by autoinduction. Cells were isolated by centrifugation, frozen in N₂(l), and stored at -80 °C until further use. For protein purification, cells were thawed on ice, resuspended in Lysis Buffer [50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM BME, 0.25 mM PMSF, 1 mg/mL lysozyme, pH 7.5] and incubated (1 h, 4 °C). The cells were lysed by sonication (4 °C, 6 V, 60 s x 4). The lysate was clarified by centrifugation (15,000 g, 30 min, 4 °C). The cleared lysate was mixed with cobalt IMAC resin (Talon resin) and incubated (4 °C, 1 h). The resin was pelleted by centrifugation (700 g, 5 min, 4 °C) and washed four times with Wash Buffer [50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM BME, 0.25 mM PMSF, 10 mM imidazole, pH 7.5] (10 resin bed volumes). His-tagged protein was eluted with Elution Buffer [50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM BME, 100 mM imidazole, pH 7.5]. The eluted protein was concentrated with a MW-cutoff spin filter (50 kDa MWCO) and exchanged into storage buffer [50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, 1 mM BME, pH 7.5]. The concentration of 8His-MBP-[TEV]-Coq8p^{NΔ41} was determined by its absorbance at 280 nm ($\epsilon=109,210 \text{ M}^{-1} \text{ cm}^{-1}$) (MW = 96.2 kDa). The fusion protein was incubated with Δ238TEV protease (1:50, TEV/fusion protein, mass/mass) (1 h, 20 °C). The TEV protease reaction mixture was mixed with cobalt IMAC resin (Talon resin) and incubated (4 °C, 1 h). The unbound Coq8p^{NΔ41} was isolated and concentrated with a MW-cutoff spin filter (30 kDa MWCO) and exchanged into Storage Buffer. The concentration of Coq8p^{NΔ41} was determined by its absorbance at 280 nm ($\epsilon=41,370 \text{ M}^{-1} \text{ cm}^{-1}$) (MW = 52 kDa). The protein was aliquoted, frozen in N₂(l), and stored at -80 °C. Fractions from the protein preparation were analyzed by SDS-PAGE.

3.5.3 ADP-Glo Assay

ADP-Glo (Promega) was performed according to the manufacturer's instruction with the following modifications. All solutions were diluted in HBS (150 mM NaCl, 20 mM HEPES pH 7.5). Coq8p^{NΔ41} (1 μM), COQ8A^{NΔ250} (2 or 4 μM), COQ8A^{NΔ162} (2 μM), or MBP-UbiB^{CΔ47} (0.5 μM) were mixed with liposomes (~3.33 mM), ATP (100 μM), MgCl₂ (4 mM), CMK (20 μM) or DMSO (0.2% v/v), Triton X-100 (1 mM) or reduced Triton X-100 (1 mM) (Sigma Aldrich), and CoQ headgroup analogs (Sigma Aldrich) dissolved as 200 mM stock solutions in DMSO (2-PP, 2-propylphenol; 2-MeO-6-MP, 2-methoxy-6-methylphenol; 4-methylcatechol; 4-HB, 4-hydroxybenzoic acid; 2-MeO-HQ, 2-methoxyhydroquinone; 4-MeOP, 4-methoxyphenol; 2-AP, 2-allylphenol; HQ, hydroquinone; 3,4-diHB, 3,4-dihydroxybenzoic acid; MeHQ, methylhydroquinone; 2,6-diMeO-HQ, 2,6-dimethoxyhydroquinone) (1 mM) (final concentrations for reaction components). Reactions were incubated (30 °C, 45 min). Then ADP-Glo Reagent (5 μL) was added and incubated (40 min, r.t. [~21 °C], covered). Kinase Detection Reagent was added (10 μL) and incubated (40 min, r.t., covered). Luminescence was read using default values on a Biotek Cytation 3 plate reader. An ADP/ATP standard curve was made according to the manufacturer's instructions. Error bars represent SD of technical triplicate measurements unless otherwise specified.

3.5.4 Cytophos ATPase Assay

The Cytophos ATPase assay (Malachite green) was performed according to the manufacturer's instruction with the following modifications. All solutions were diluted in HBS (150 mM NaCl, 20 mM HEPES pH 7.5). COQ8A^{NΔ250} (1 μM final), Coq8pND41 (1 μM final), COQ8A^{NΔ162} (1 μM final) or MBP-UbiB^{CΔ47} (0.25 μM final) with 100 μM ATP, 4 mM MgCl₂, 1 mM 2-alkylphenol, 0.5 mM CoQ₁, 1 mM Triton X-100 or reduced Triton X-100. CoQ₁ was reduced by adding 1.2 fold molar excess of NaBH₄ (Sigma) (10 min, r.t.). Reactions were incubated (30 °C, 45 min) and 35 μL of Cytophos reagent was added and incubated (10 min, r.t.). Absorbance was measured at 650 nm. Reactions in Figure 2E were incubated for 10 min rather than 45 min and [Coq8p^{NΔ41}] was 0.5 μM. The standard curve ranged from 0–50 μM phosphate. Error bars represent SD of technical triplicate measurements unless otherwise specified.

3.5.5 In Vitro [γ - 32 P] Autophosphorylation and ATPase Comparison Assay

Method for Figure 3.4D. Coq8p^{N Δ 41} (4 μ M) was mixed with [γ - 32 P]ATP (0.25 μ Ci/ μ L, 100 μ M [ATP]_{total}) and MgCl₂ (4 mM) in an aqueous buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) and incubated (30 °C, 45 min, 20 μ L total reaction volume) (final concentrations for reaction components). 10 μ L of reaction was quenched with 10 μ L 0.75 M potassium phosphate pH 3.3, 1 μ L of quenched reaction was spotted on a PEI-cellulose TLC plate and developed using the method above. The other 10 μ L of the reaction was quenched with 4 μ L 4 x LDS buffer and 10 μ L was loaded onto an SDS-PAGE gel. [γ - 32 P]ATP was separated from Coq8p^{N Δ 41} by SDS-PAGE (10% Bis-Tris gel, MES buffer, 150 V, 80 min). The gel was stained with Coomassie Brilliant Blue, dried under vacuum, and imaged by digital photography. The same storage phosphor screen was exposed to the TLC plate and the SDS-PAGE gel (1 hr) and then imaged with a Typhoon (GE) to generate the phosphorimages used for quantification of Pi and autophosphorylation.

3.5.6 In Vitro [γ - 32 P]ATP ATPase Assay

Method for Figure 3.4F Unless otherwise indicated, COQ8A^{N Δ 162} (2 μ M) or Coq8p^{N Δ 41} (1 μ M) was mixed with liposomes (3.33 mM), [γ - 32 P]ATP (0.01 μ Ci/ μ L, 100 μ M [ATP]_{total}), and MgCl₂ (4 mM) in an aqueous buffer (20 mM HEPES, 150 mM NaCl, pH 7.5) and incubated (20 mL total volume, 30 °C, 45 min, 700 rpm) (final concentrations for reaction components). Reactions were quenched with 0.75 M potassium phosphate pH 3.3 (20 μ L, 4 °C). 1 μ L of quenched reaction was spotted on PEI-cellulose TLC (Millipore) and developed with 0.5 M LiCl in 1 M formic acid(aq). After drying, a storage phosphor screen was exposed to the PEI-cellulose TLC plate (~5 hours) and then imaged with a Typhoon (GE) to generate the phosphorimages.

3.5.7 Yeast Drop Assay and Growth Curves

Yeast cultures for drop assays. *S. cerevisiae* (BY4742) Δ coq8 yeast were transformed as previously described⁹⁶ with p426 GPD plasmids encoding Coq8p variants and grown on uracil drop-out (Ura⁻) synthetic media plates containing glucose (2%, w/v). Individual colonies of yeast were used to inoculate Ura⁻ media (2% glucose, w/v) starter cultures, which were incubated (30 °C, ~18 h, 230 rpm). Serial dilutions of yeast (10⁴, 10³, 10², or 10 yeast cells)

Chapter 3. Conserved Lipid Modulation of Ancient Kinase-Like UbiB Family Member COQ8

were dropped onto Ura⁻ agar media plates containing either glucose (2%, w/v) or glycerol (3%, w/v) and incubated (30 °C, 2 d). To assay yeast growth in liquid media, yeast from a starter culture were swapped into Ura⁻ media with glucose (0.1%, w/v) and glycerol (3%, w/v) at an initial density of 5x 10⁶ cells/mL. The cultures were incubated in a sterile 96 well plate with an optical, breathable cover seal (shaking at 1140 rpm). Optical density readings were obtained every 10 min.

3.5.8 Liposomes

Liposomes were made by drying down lipids in a 5 mL plastic tube under Ar(g) at room temperature until a film was left. Liposomes were made of the following lipid (Avanti Polar Lipids) compositions: PC, PC/NBD-PE 99.9/0.1; PC/PE, PC/PE/NBD-PE 89.9/10/0.1; PC/CL, PC/CL/NBD-PE 89.9/10/0.1; PC/PE/CL, PC/PE/CL/NBD-PE 79.9/10/10/0.1; PC/CoQ₁₀, PC/CoQ₁₀/NBD-PE 97.9/2/0.1; PC/PG, PC/PG/NBD-PE 89.9/10/0.1; PC/PS, PC/PS/NBD-PE 89.9/10/0.1; PC/PA, PC/PA/NBD-PE 89.9/10/0.1; PC/CDP-DAG, PC/CDP-DAG/NBD-PE 89.9/10/0.1; all mol %. The lipids were dried in a vacuum chamber overnight at 25-30 inHg vacuum. The dry lipid film was reconstituted in HBS buffer (20 mM HEPES pH 7.5, 150 mM NaCl) at 30–35 °C for one hour with occasional pipetting to resuspend the lipid film. The total concentration of lipids in solution was 10 mM. 2 μL of the liposomes were taken before and after extrusion and diluted with 22 μL of HBS to determine how much the liposomes were diluted during extrusion by measuring the fluorescence of NBD-PE (excitation: 460 nm, emission: 535 nm). Liposomes were extruded through 100 nm membranes (Avanti Polar Lipids), 15 passes, 30–35 °C. Liposomes were made fresh before each experiment.

3.5.9 Liposome Flotation Assay

Liposome flotation assay is adapted from Connerth et al.⁹⁷ with the following modifications. Liposomes (100 μL) were mixed with protein (50 μL 6 x COQ8) at 4 °C then incubated (r.t., ~ 10 min). Final concentrations of reagents are as follows: 2–4 μM protein and 3.33 mM liposomes in HBS (150 mM NaCl, 20 mM HEPES pH 7.5). 2.72 M sucrose (110 μL) was added to the protein liposome mixture (1.15 M [sucrose] final). The sucrose-liposome-protein mixture (250 μL) was added to the ultracentrifuge tube. The sucrose gradient was made by layering 300 μL HBS

3.5. Supplementary Information

0.86 M sucrose, 250 μ L HBS 0.29 M sucrose, and 150 μ L HBS on top of the sucrose-liposome-protein mix. This gradient was centrifuged (240,000 g, 1 h, 4 °C). 450 μ L was removed for the top layer and 450 μ L for the bottom layer. Liposome location was determined by mixing 8 μ L of top or bottom fraction with 16 μ L HBS and reading NBD-PE fluorescence for (excitation: 460 nm, emission: 535 nm). To concentrate the protein from the top and bottom fractions, a CHCl_3 :MeOH precipitation was performed according to Wessel and Flugge.⁹⁸ Methanol (1800 μ L, 4 volumes) was added to a 450 μ L fraction. After thorough mixing chloroform (450 μ L, one volume) was added and vortexed. Water (1350 μ L, three volumes) was added and the mixture was vortexed again then centrifuged immediately at full speed (\sim 4,200 g, 5 min). A white disc of protein should form between the organic layer (bottom) and the aqueous layer (upper). Discard most of the upper aqueous layer, and be careful not to disturb the protein pellet. Methanol was added (1000 μ L) to the tube, inverted 3 times, and centrifuged at full speed (5 min, 16,000 g). All of the liquid was removed and the pellet was air or vacuum dried. The precipitated protein pellet was resuspend in 30 μ L 1 x LDS with 10 mM DTT, incubated (95 °C, \sim 10 min), and analyzed with 4–12% Novex NuPAGE Bis-Tris SDS-PAGE (Invitrogen) gels (1 hr, 150 V). Band quantification was done with imaging and densitometry on a LI-COR Odyssey CLx (700 nm) using Image Studio v5.2 software. Error bars for Figure 4G represent SD of three independent floats. Student's t-test was used to determine statistical significance.

4 Membrane, Lipid, and Protein

Interactions of Coenzyme Q Biosynthesis Protein COQ9

This Chapter is adapted from the following published manuscript:

"An Isoprene Lipid-Binding Protein Promotes Eukaryotic Coenzyme Q Biosynthesis."

Lohman DC*, **Aydin D***, Vonbank H, Smith R, Linke V, Weisenhorn E, McDevitt MT, Hutchins P, Wilkerson E, Russell J, Stefely MS, Beebe ET, Jochem A, Bingman CA, Coon JJ, Dal Peraro M, Pagliarini DJ.

Molecular Cell (2019) 73(4): 763-774.

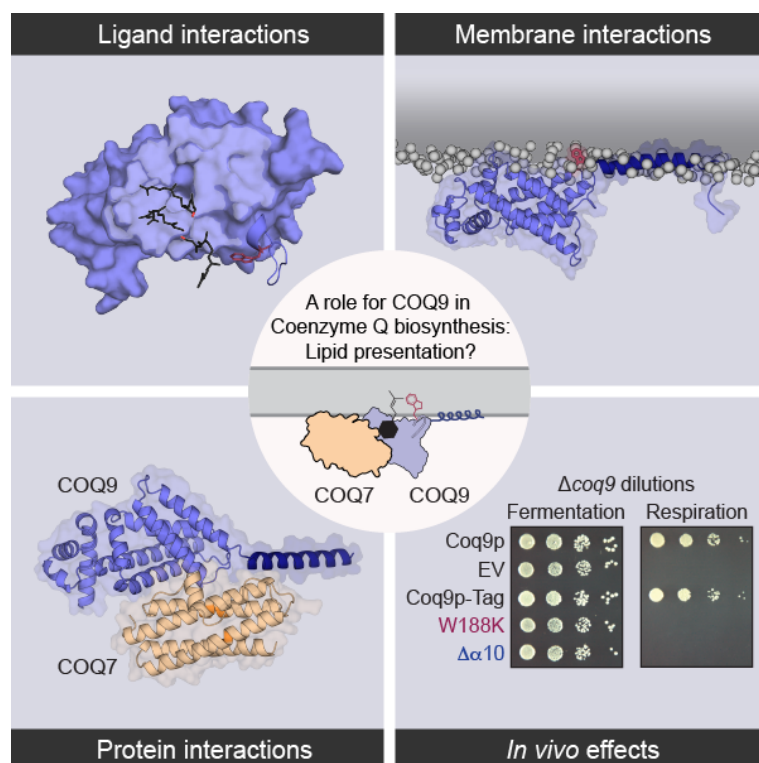
doi: 10.1016/j.molcel.2018.11.033

Author Contributions: *D.C.L. and **D.A.** contributed equally. D.C.L., **D.A.**, M.D.P., and D.J.P. conceived of the project and its design and wrote the manuscript. D.C.L., H.C.V.B., R.W.S., M.S.S., E.T.B., and A.J. prepared samples and performed biochemical experiments. **D.A.** performed computational experiments. C.A.B. solved the crystal structures. V.L., E.W., M.T.M., P.H., and E.M.W. acquired mass spectrometry data. D.C.L., **D.A.**, E.T.B., C.A.B., P.H., J.R., J.J.C., M.D.P., and D.J.P. analyzed data.

The biosynthesis of coenzyme Q presents a paradigm for how cells surmount hydrophobic barriers in lipid biology. In eukaryotes, CoQ precursors—among nature's most hydrophobic molecules—must somehow be presented to a series of enzymes peripherally associated with the mitochondrial inner membrane. In this Chapter, we reveal that this process relies on custom lipid-binding properties of COQ9. We show that COQ9 repurposes the bacterial TetR fold to bind aromatic isoprenes with high specificity, including CoQ intermediates that

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

likely reside entirely within the bilayer. We reveal a process by which COQ9 associates with cardiolipin-rich membranes and warps the membrane surface to access this cargo. Finally, we identify a molecular interface between COQ9 and the hydroxylase COQ7, motivating a model whereby COQ9 presents intermediates directly to CoQ enzymes. Overall, our results provide a mechanism for how a lipid binding protein might access, select, and deliver specific cargo from a membrane to promote biosynthesis.



4.1 Introduction

The synthesis and transport of biological lipids in a largely aqueous cellular terrain poses a longstanding problem in biology.^{1,99} A requisite first step in these processes is to access and separate a lipid from its local membrane environment. One of nature's strategies for dealing with such hydrophobicity challenges is to employ lipid binding proteins, which possess diverse modes of action (Figure 4.1) Transmembrane enzymes, like stearyl-CoA desaturase, penetrate biological membranes and bring enzymatic function to the hydrophobic milieu.¹⁰⁰ Lipid chaperones, like acyl carrier protein^{101,102} or Ups1/Mdm35,^{97,103,104} solubilize lipids in a hydrophobic pocket and transport them to soluble enzymes or insert them into membranes.

Lipid lifases, like saposin C¹⁰⁵ or PlsC¹⁰⁶ act at the membrane-water barrier to partially lift the lipid from the bilayer. Biosynthesis and spatiotemporal distribution of cellular lipids, from fatty acids to sterols, may depend on one or more of these strategies.

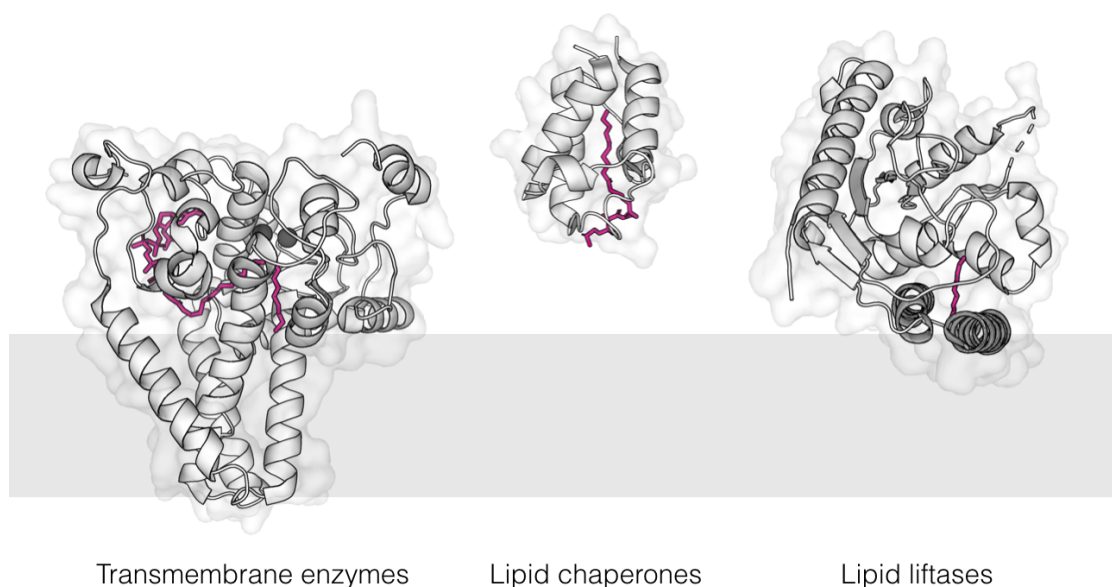


Figure 4.1 – Lipid-Binding Proteins. Examples illustrated here include transmembrane enzymes (stearoyl-CoA desaturase (PDB:4ZYO)), lipid chaperones (acyl carrier protein (PDB:2FAE)), and lipid lifases (bacterial acyltransferase PlsC (PDB: 5KYM)). Proteins are placed with respect to the membrane (represented by the gray area), where lipids bound to proteins are shown in pink sticks.

Coenzyme Q is a redox-active isoprene lipid that serves as a cofactor for numerous enzymes across all domains of life.¹⁶ Deficiency in CoQ is linked to many rare inborn errors of metabolism, and to common diseases including Parkinson's and type II diabetes mellitus.^{16,107,108} The molecular structure of CoQ is remarkable: its extraordinary isoprene tail—composed of 50 carbons in humans—places it among the most hydrophobic molecules in nature. This hydrophobicity keeps CoQ sequestered within the inner leaflet of the inner mitochondrial membrane where it passes electrons between complexes I & II and complex III in the electron transport chain. The CoQ tail is attached to an immature head group precursor at an early step in CoQ biosynthesis. In eukaryotes, enzymes within complex Q (a.k.a., the CoQ synthome) then modify this precursor to generate mature CoQ (see Figure 1.4).^{16,109} For these modifications to occur, complex Q enzymes must gain access to these hydrophobic precursors,

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

which, like mature CoQ,^{110,111} are thought to reside within the hydrophobic portion of the membrane¹⁶; however, how this access is granted remains unknown.

Recently, it was shown that COQ9—a poorly characterized protein associated with CoQ biosynthesis that structurally resembles the TetR family (TFR) of proteins¹⁵—can bind lipids, suggesting that it may help complex Q overcome this hydrophobic challenge. Loss of COQ9 function in many organisms results in severe CoQ deficiency and underlies human disease, demonstrating its central role in this pathway.^{112–117} However, which lipid(s) COQ9 binds, how it accesses this cargo, and how this binding may assist other pathway enzymes have remained open questions.

Here, we investigate the molecular mechanism of COQ9 lipid-binding biochemically and computationally. We find that COQ9 binds aromatic isoprene lipids, accesses membranes via a terminal amphipathic helix, and that both of these functions are crucial to its biological role in CoQ production. We further establish a likely molecular interface between COQ9 and COQ7—the hydroxylase that catalyzes the penultimate step in CoQ biosynthesis. Collectively, these data suggest models in which COQ9 enables CoQ biosynthesis by accessing CoQ intermediates from the leaflet of the mitochondrial inner membrane and presenting them to a biosynthetic enzyme in the pathway.

4.2 Methods

This section presents the detailed description of the computational methods used in this study. For experimental methods, please refer to Section 4.5.

4.2.1 Coarse-Grained Molecular Dynamics Simulations

Coarse-grained molecular dynamics (CG-MD) simulations were used to investigate: (i) the interaction of COQ9 and COQ7 with bilayers containing different lipid compositions, (ii) the flexibility of the $\alpha 9$ - $\alpha 10$ hinge allowing different structural placements of $\alpha 10$. All simulations

were performed using the GROMACS¹¹⁸ simulation package version 5.1.2.⁶¹ The systems were described with the MARTINI 2.2P^{24,62} force field, together with the ElNeDyn (Elastic Network in Dynamics) approach to maintain the secondary structure of the protein.^{24,63} All simulations of COQ9 were performed using the crystal structure of COQ9 where the $\alpha 10$ is in an extended conformation (PDB: 6AWL, chain A). Simulations of COQ7 were performed using a homology model of COQ7 (see below). For visualization purposes, selected frames from the CG-MD simulations were backmapped into atomistic coordinates using backward.py.¹¹⁹ Images were produced with PyMOL (Schrödinger LLC, Version 2.0) and videos were rendered with VMD.^{70,71}

Protein-Membrane Systems

CG-MD simulations based on the MARTINI force field have been previously used with success for different systems to investigate protein-lipid interactions.^{64–67} In CG-MD simulation systems, the center of mass of COQ9 was 70 Å away from the surface of the membrane. Since the last 7 C-terminal residues of COQ9 were not solved in the crystal structure, these residues were built to be disordered using Modeller version 9.11.¹²⁰ Inner mitochondrial membrane (IMM) models were built according to the CL composition of mammalian mitochondria, namely 10 to 20% of total mitochondrial phospholipids.⁶⁸ Therefore, IMM mimics were featuring a lipid concentration of 16% CL/41% POPC/37% POPE/6% DSPC, while generic membrane models featuring 100% POPC lipids were built as control. The IMM model was approximated to have a symmetrical concentration of lipids between the two leaflets since precise lipidomic information about concentration asymmetry is not known. Lipid bilayers of $160 \times 160 \text{ Å}^2$ were generated using the insane (INSert membrANE) method of MARTINI.⁶⁹ Systems were solvated with polarizable CG water, and ions were added to the system to obtain a 150 mM NaCl aqueous solution, with extra ions added to neutralize the system. Following solvation, systems were energy minimized with a timestep of 5 fs. Successive equilibrations with decreasing restraints were performed in order to obtain a fully equilibrated system (force constants of 1000, 500, 250 and 0 kJ/mol applied on protein particles, 10 ns of runs with a timestep of 5 fs for each). In the production phase (10 μs of run with a timestep of 15 fs), the protein, membrane and the aqueous phase (water and ions) were coupled independently to an external bath at 310

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

K by using the v-rescale thermostat. The pressure was coupled to a pressure bath at 1 bar by semi-isotropic Parrinello-Rahman pressure coupling. Three MD repeats with randomized initial velocities were performed for each membrane system and they yielded consistent results (Fig. 4.7A, 4.7B, 4.8B, 4.8D). The minimum pairwise distances between the protein and the membrane were calculated using the gmx pairdist tool of Gromacs,¹¹⁸ which takes into account the distance to the periodic image of the membrane. The MSD and lateral diffusion coefficients were calculated using the gmx msd tool of Gromacs.¹¹⁸ Average occupancies of lipids were calculated using VMD's VolMap plugin and visualized using VMD.⁷⁰ CG-MD simulations of COQ7 were performed to investigate the binding of COQ7 to $130 \times 130 \text{ \AA}^2$ IMM models (16% CL/41% POPC/37% POPE/6% DSPC). The procedure applied to COQ9 was used.

Protein-Membrane Systems Involving CoQ Molecules: We used the readily available CG topology for CoQ (ubiquinol), which has been shown to be in good agreement with experimental data.¹²¹ Ideally, we would use DMQ in these simulations, but parametrizing DMQ at CG resolution was beyond the scope of this study. CoQ is very similar to DMQ at the CG level of resolution. To investigate how CoQ interacts with COQ9, we performed CG-MD simulations of COQ9 with a PC/PE/CL/CoQ₁₀ membrane. To build this system, four CoQ molecules were first placed in solution near the membrane surface without COQ9, and the system was energy minimized, equilibrated (25 ns with a time step of 5 fs) and simulated for 3 microseconds where CoQ molecules diffused inside the membrane, positioning their aromatic head groups on an inner plane compared to the lipid phosphate head groups, confirming the results previously obtained.¹²¹ The last frame of this simulation was used to add COQ9 to the system, where it was placed near the membrane surface, minimized and equilibrated through the procedure described above for protein-membrane systems, and this system was finally simulated for 10 microseconds.

Protein in Solution Systems

The MARTINI force field allows protein tertiary arrangements, whereas local secondary structure is predefined and imposed. COQ9 was solvated with polarizable CG water and ions were added to the system to obtain a 150 mM NaCl aqueous solution, with extra ions added to

neutralize the system. Following solvation, systems were energy minimized with a timestep of 5 fs. Successive equilibrations with decreasing restraints were performed in order to obtain a fully equilibrated system (force constants of 1000, 500, 250 and 0 kJ/mol applied on protein particles, 10 ns of runs with a timestep of 5 fs for each). In the production phase (10 μ s of run with a timestep of 15 fs), the protein and the aqueous phase (water and ions) were coupled independently to an external bath at 310 K. No pressure coupling was applied.

4.2.2 Bioinformatics, Docking, and Modeling

Bioinformatics Analysis of Secondary Structures of TFR proteins

Secondary structures were assigned to the representative set of TFR sequences as defined by Yu and colleagues¹²² and to COQ9 using ICM-Pro molecular modeling software.³³ Each helix was aligned and centered independently for all the reference TFRs, and the mean helix length and the standard deviation was computed for each helix. This analysis therefore does not take into account loop regions. For each residue position of each helix the percentage of TFRs having a residue that belongs to this helix was calculated.

Ligand Docking

Ligand docking and scoring was performed using the ICM-Pro molecular modeling software.^{33,123} ICM ligand docking is based on biased probability Monte Carlo (BPMC) optimization of the ligand internal coordinates in the set of grid potential maps of the receptor, where the grid potential maps account for van der Waals, hydrogen-bonding, hydrophobic, and electrostatic interactions between ligand and receptor. The crystal structure of COQ9 where the $\alpha 10$ is in an extended conformation (PDB: 6AWL, chain A) was used as the receptor. DMQ₅ was used as the ligand instead of DMQ₁₀ because part of the hydrophobic tail of DMQ₁₀ is expected to reside in the membrane and because using such a long ligand increases the degrees of freedom and decreases the docking accuracy. The ligand was represented as an all-atom model and considered fully flexible. The standard docking procedure of ICM was used without the use of any experimentally derived information. Docking poses were evaluated with the

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

ICM ligand binding score:

$$S_{bind} = E_{int} + \Delta S_{Tor} + E_{vdw} + \alpha_1 E_{el} + \alpha_2 E_{hb} + \alpha_3 E_{hp} + \alpha_4 E_{sf} \quad (4.1)$$

where E_{int} is the internal strain of the ligand, $T=300$ K, ΔS_{Tor} is the conformational entropy loss of the ligand upon binding, α_i are constants independent from the ligand and receptor. E_{vdw} , E_{el} , E_{hb} , E_{hp} , E_{sf} are van der Waals, electrostatic, hydrogen bonding, and nonpolar and polar atom solvation energy differences between bound and unbound states, respectively. To compare the docking pose to the co-crystallized isoprenes (PDB: 6DEW), chains A-F and co-crystallizing isoprenes were superimposed to the docked complex. The RMSD between the isoprene chain of DMQ₅ and that of farnesol (bound to chain A) were calculated on 15 matching carbon atoms and was found to be 1.5 Å. To visualize the DMQ₅-bound COQ9 on the membrane, the globular domain of COQ9 was superimposed in the docked complex to the globular domain of COQ9 in a membrane-bound state obtained from the CG-MD simulation. Since the ligand found contacts with W240 in the docking procedure, the crystal structure conformation of W240 (PDB: 6AWL, chain A) was used for the membrane-bound state of COQ9 for visualization purposes in Figure 4.11D.

Homology Modeling of COQ7

At present, there are no crystal structures available for human COQ7. Pfam¹²⁴ identifies COQ7 as a di-iron carboxylate protein that belongs to the ferritin-like superfamily. Members of this protein family are known to have a four-helix bundle folding architecture with a conserved iron binding motif.¹²⁵ Several attempts were previously made to model Coq7 homologues (*P. aeruginosa*, *R. norvegicus*, *S. cerevisiae*) using homology modeling techniques.^{125–127} Here, we used a similar approach to model the structure of COQ7 from *H. sapiens* starting from its sequence as obtained from the UniProt database (accession number Q99807).¹²⁸ SWISS-MODEL automated comparative protein modeling server¹²⁹ was used in Automated Mode to construct a homology model of the human COQ7. The N-terminal 37 residues were removed from the sequence since they had low conservation and since this region was predicted to be

disordered; the C-terminal 32 residues were not modeled since this region was missing in all found templates and was not included in the predicted four-helix bundle. SWISS-MODEL uses two scores for model evaluation: GMQE and QMEAN. GMQE (Global Model Quality Estimation) is expressed as a number between 0 and 1, and a higher value means higher model reliability. QMEAN¹³⁰ estimates the quality of the model compared to experimental structures, and a higher value means better agreement between the model structure and experimental structures of similar size, where scores around zero indicate a good degree of nativeness and scores lower than or equal to -4.0 indicate a low quality model. The best template identified by SWISS-MODEL for COQ7 is a ferritin (PDB: 5HJH, 1.88 Å resolution), which has a 15% sequence identity to COQ7. The generated model has a GMQE score of 0.43 and QMEAN score of -3.09. Despite the initial low sequence identity with templates, COQ7 shares the well-defined structural fold of the ferritin-like superfamily (as determined by Pfam¹²⁴). Therefore, modeling its four-helix bundle structure was facilitated by helical secondary structure prediction and the strong conservation of the iron binding motif. The resulting model was consistent with previously published models of Coq7p.^{125–127}

Homology Modeling of Coq9p

COQ9 sequences from *H. sapiens* (accession number O75208) and *S. cerevisiae* (accession number Q05779) were obtained from the UniProt¹²⁸ database and were aligned using Clustal Omega.¹³¹ This alignment was used to model yeast Coq9p with Modeller version 9.11.¹²⁰ The MD refinement parameter was turned off and 5 models were generated, where the model with the lowest molpdf score was chosen as the final model.

Protein-Protein Docking

To model the COQ9-COQ7 complex, the protein-protein docking server ZDock was used.¹³² In predicting protein-protein complexes, ZDock takes into account shape complementarity, electrostatics and desolvation free energy.¹³³ The crystal structure of COQ9, where the $\alpha 10$ is in an extended conformation (PDB: 6AWL, chain A), and the homology model generated for COQ7 were used as input structures. Protein-protein docking was guided by (i) the experimen-

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

tal data (Fig. 4.10A), where residues that disrupted binding were forced to be included in the interface and residues that did not affect binding were forced not to be included and (ii) the preferred membrane orientation of COQ9 and COQ7 as obtained from CG-MD simulations. Ten poses were obtained where the intermolecular contacts of Fig. 4.10A were fulfilled and the ZDock score of the least favored pose was lower by at most 30% of the best pose. Seven of the ten complex conformations were not occluding the membrane binding surface of COQ9 as identified through CG-MD simulations, and the top two conformations of this cluster were not occluding the membrane binding surface of COQ7. These two conformations showed only minor differences and an equivalent complex architecture having an RMSD as low as 5.2 Å. The best ZDock pose between these two conformations was therefore used as representative of COQ9-COQ7 interactions and is reported in Figure 4.11.

4.3 Results

4.3.1 COQ9 Specifically Binds Aromatic Isoprenes

Recently, the fortuitous co-crystallization of COQ9 with a mixture of phospholipids was reported.¹⁵ This discovery, along with structural homology to ligand binding domains of the TFRs and a predominantly hydrophobic cavity, defined COQ9 as a lipid-binding protein. Yet, the identity of bona fide COQ9 ligands remains unknown.

To identify endogenous lipid ligands of COQ9, we purified a tagged version of its yeast ortholog (Coq9p-FLAG) from *Saccharomyces cerevisiae* and identified ligands by mass spectrometry. Among various co-purifying lipids were several CoQ intermediates (Figures 4.2B, 4.2C). Membrane-associated proteins purified from endogenous sources are often accompanied by non-specific lipids. Therefore, to distinguish specific lipid interactions from non-specific binders, we mutated the highly conserved, surface-exposed tryptophan that resides near the co-crystallizing phospholipid in our previous COQ9 structure (Figure 4.2D) and is essential for COQ9 function.¹⁵ Interestingly, this mutation was not well tolerated in *S. cerevisiae* (see Figure 4.12), so we performed these experiments in *E. coli*, which lack a COQ9 homolog.

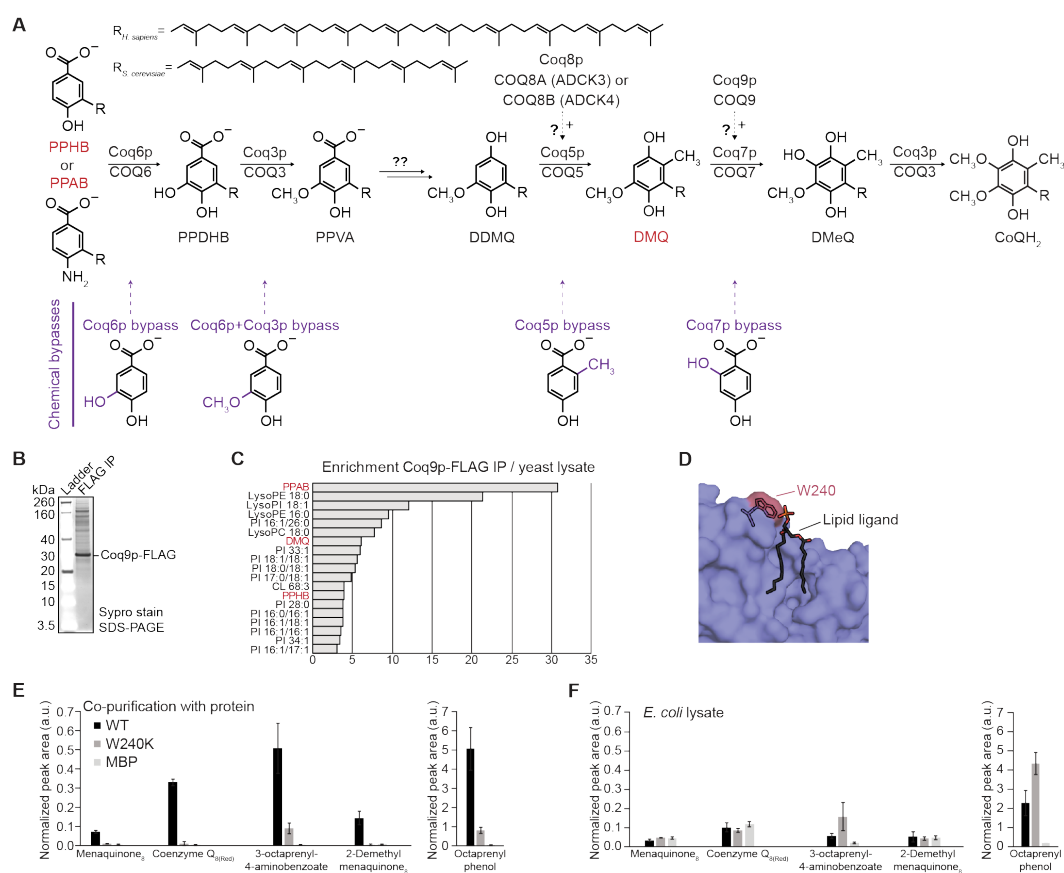


Figure 4.2 – LC-MS/MS Analysis of COQ9 Lipid Ligands, Related to Figures 4.3 and 4.12. (A) Head group modifications of eukaryotic CoQ biosynthesis post-attachment of the polyprenyl tail (R) where *S. cerevisiae* (Coq#p) and *H. sapiens* (COQ#) protein names are shown. PPHB, polyprenyl-hydroxybenzoate; PPAB, polyprenyl-aminobenzoate; PPDHB, polyprenyl-dihydroxybenzoate; PPVA, polyprenyl-vanillic acid; DDMQ, demethoxy-demethyl CoQ; DMQ, demethoxy CoQ; DMeQ, demethyl CoQ; CoQH₂, reduced CoQ. Below, chemical head group bypasses used in Figures 4.12 and 4.13, where pre-modified moieties are in purple. (B) Sypro-stained SDS-PAGE of Coq9p-FLAG immuno-purification from *S. cerevisiae*. (C) Top ranking identified lipids enriched (enrichment > 3, p-value < 0.05) in Coq9p-FLAG pull-down from *S. cerevisiae* compared to whole cell lysate. CoQ intermediates are highlighted in red in panels A and C. (D) Co-crystallized phospholipid with COQ9, where W240 is in red and the phospholipid is in black stick, PDB ID 4RHP. (E) LC-MS/MS quantification of enriched *E. coli* lipids from purified protein in Fig. 4.3, where WT is His₈-MBP-COQ9^{NΔ79}, W240K is His₈-MBP-COQ9^{NΔ79}, W240K and MBP is His₆-MBP. Lipid peak areas are normalized to the total peak area per sample and error bars represent standard deviation (n=3). (F) LC-MS/MS quantification of *E. coli* lipids in Fig. 1 from the *E. coli* lysates expressing recombinant protein, with bar colors described in panel F: WT is His₈-MBP-COQ9^{NΔ79}, W240K is His₈-MBP-COQ9^{NΔ79}, W240K and MBP is His₆-MBP. Lipid peak areas are normalized to the total peak area per sample and error bars represent standard deviation (n=3).

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

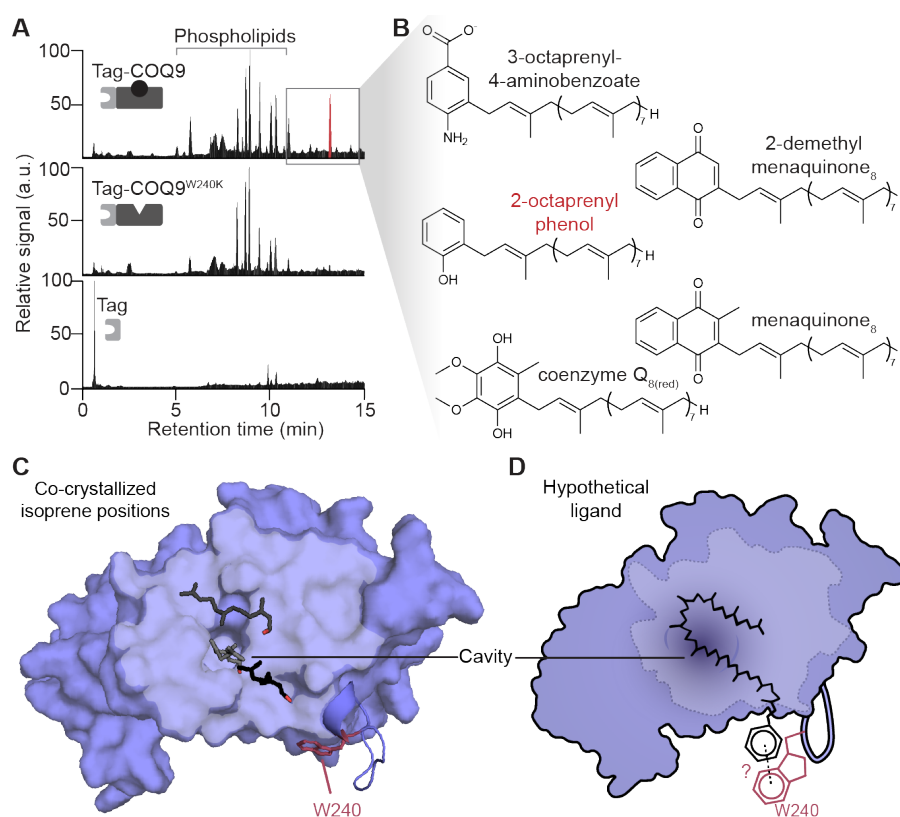


Figure 4.3 – COQ9 Binds Aromatic Isoprenes. (A) LC-MS/MS total ion chromatograms of lipids co-purifying with Tag-COQ9^{NΔ79}, Tag-COQ9^{NΔ79, W240K}, and Tag expressed in *E. coli* where “Tag” is His₈ followed by maltose binding protein. (B) Chemical structures of identified enriched lipids (fold-change > 4, p-value < 0.01) in Tag-COQ9^{NΔ79} compared to Tag-COQ9^{NΔ79, W240K} and Tag. The enriched lipid with the highest intensity signal is highlighted in red. (C) Representative isoprene positions (stick) observed on a hydrophobic surface and in the cavity of COQ9^{NΔ79, Δ31} with W240 in red and the hydrophobic surface in light blue (PDB: 6DEW). Of six isoprenes that co-crystallized in various positions on four of the six COQ9 chains (A,B,D,F) in the unit cell, the three shown are representative of distinct placements and are all mapped to a single COQ9 molecule (chain A). (D) Model for how an aromatic long-chain isoprene could potentially bind on the surface of COQ9, where the COQ9 cavity is shaded, the hydrophobic surface is outlined, and potential π - π stacking interactions with W240 on α 7- α 8 loop are highlighted. See also Figures 4.2, 4.4 and Table 4.2.

Specifically, we purified tagged (His-MBP) versions of human COQ9^{NΔ79} (the version of COQ9 whose structure was solved previously; see Table 4.1), COQ9^{NΔ79, W240K} (this same construct with the tryptophan mutated to lysine), and His-MBP alone from *E. coli*, and again identified co-purifying lipids by mass spectrometry (Figure 4.3A). Consistent with our previous work, phospholipids were the most abundant species co-purifying with COQ9¹⁵; however, the W240K construct also co-purified with many phospholipids (Figure 4.3A). Remarkably, all identified lipids specifically enriched in WT COQ9^{NΔ79} were aromatic isoprenes, including

CoQ(reduced), menaquinone (a bacteria-specific CoQ-like molecule), and their various precursors (Figure 4.3B, 4.2E). This enrichment in lipids resembling CoQ was not due to a global change in lipid levels in the *E. coli* expressing the recombinant protein (Figure 4.2F).

Table 4.1 – Protein Constructs, Related to Figures 4.3, 4.10, 4.11 and 4.12.

Experiment	Expression system	Protein construct description
<i>S. cerevisiae</i> APMS	<i>S. cerevisiae</i>	(“Coq9p-FLAG”) Full length <i>S. cerevisiae</i> Coq9p with C-terminal linker (LDLE) and FLAG tag
<i>E. coli</i> APMS	<i>E. coli</i>	(“His ₈ -MBP-tev-COQ9 ^{NΔ79} ”) Solubility tag (His8-MBP) with linker and TEV protease cleavage site N-terminal to <i>H. sapiens</i> COQ9 amino acids 79-318 (N-terminal 79 residues deleted)
Crystallography	<i>E. coli</i>	(“COQ9 ^{NΔ79} ” and “COQ9 ^{NΔ79,CA31} aka COQ9 ^{Δα10} ”) <i>H. sapiens</i> COQ9 amino acid residues 79-218 or 79-287 (C-terminal 31 residues deleted that correspond to the 10 th alpha helix) with two additional amino acids residual from TEV protease cleavage
Co-purification of COQ7	Cell-free wheat germ extract	(“SII-COQ9 ^{NΔ45} ” or “His ₆ -COQ7 ^{NΔ38} ”) N-terminally tagged <i>H. sapiens</i> protein with 45 or 38 N-terminal amino acids deleted, respectively
<i>S. cerevisiae</i> growth	<i>S. cerevisiae</i>	(“Coq8p-FLAG” and “Coq9p-FLAG”) Full length <i>S. cerevisiae</i> proteins with C-terminal linker (LDLE) and FLAG tag

To understand how COQ9 accommodates isoprenes, we co-crystallized COQ9 with the isoprene lipid geranylgeraniol (Figure 4.3C), as CoQ intermediates themselves have low solubility and are not commercially available. We performed these co-crystallization experiments with a truncated protein (COQ9^{NΔ79,CA31}) to remove a C-terminal amphipathic helix that previously co-crystallized near the lipid binding site, possibly occluding ligand entry into the binding pocket when COQ9 is not associated with a membrane (Figure 4.5B). We determined the geranylgeraniol:COQ9^{NΔ79,CA31} structure to 2 Å (Table 4.2, PDB: 6DEW). Six isoprene positions on multiple COQ9 chains in the unit cell were well-defined in the electron density (Figures 4.4A, 4.4B), with isoprenes commonly positioned along a hydrophobic surface of COQ9 and in the hydrophobic cavity (Figures 4.3C, 4.4D).

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

Table 4.2 – Data Collection and Refinement Statistics. Statistics for the highest-resolution shell are shown in parentheses.

	COQ9^{NA79} P21 (PDB: 6AWL)	COQ9^{NA79, CA31} GG (PDB: 6DEW)
Wavelength	1.033	1.127
Resolution range	48.69 – 2.0 (2.071 – 2.0)	48.0 - 2.0 (2.071 - 2.0)
Space group	P 1 21 1	C 2 2 21
Unit cell	38.11 97.39 63.65 90 95.4 90	116.45 222.78 130.78 90 90 90
Total reflections	159831 (15769)	1545884 (159984)
Unique reflections	31219 (3066)	114146 (11326)
Multiplicity	5.1 (5.1)	13.5 (14.1)
Completeness (%)	99.90 (100.00)	99.68 (99.87)
Mean I/sigma(I)	12.38 (1.25)	15.48 (1.51)
Wilson B-factor	36.31	37.83
R-merge	0.09055 (1.355)	0.1274 (1.759)
R-meas	0.101 (1.51)	0.1325 (1.824)
R-pim	0.04409 (0.658)	0.03598 (0.4802)
CC1/2	0.998 (0.48)	0.998 (0.622)
CC*	1 (0.805)	1 (0.876)
Reflections used in refinement	31212 (3066)	114124 (11326)
Reflections used for R-free	2000 (197)	2045 (197)
R-work	0.1974 (0.3016)	0.1692 (0.2626)
R-free	0.2454 (0.3183)	0.2094 (0.3055)
CC(work)	0.957 (0.710)	0.967 (0.817)
CC(free)	0.962 (0.679)	0.947 (0.730)
Number of non-hydrogen atoms	3564	9744
Macromolecules	3313	9183
Ligands	46	106
Solvent	205	455
Protein residues	409	1138
RMS(bonds)	0.003	0.019
RMS(angles)	0.64	1.59
Ramachandran favored (%)	97.52	97.69
Ramachandran allowed (%)	1.98	2.13
Ramachandran outliers (%)	0.50	0.18
Rotamer outliers (%)	1.12	0.51
Clashscore	4.21	4.09
Average B-factor	43.21	52.11
Macromolecules	42.82	51.61
Ligands	63.56	102.08
Solvent	44.83	50.57
Number of TLS groups	N/A	36

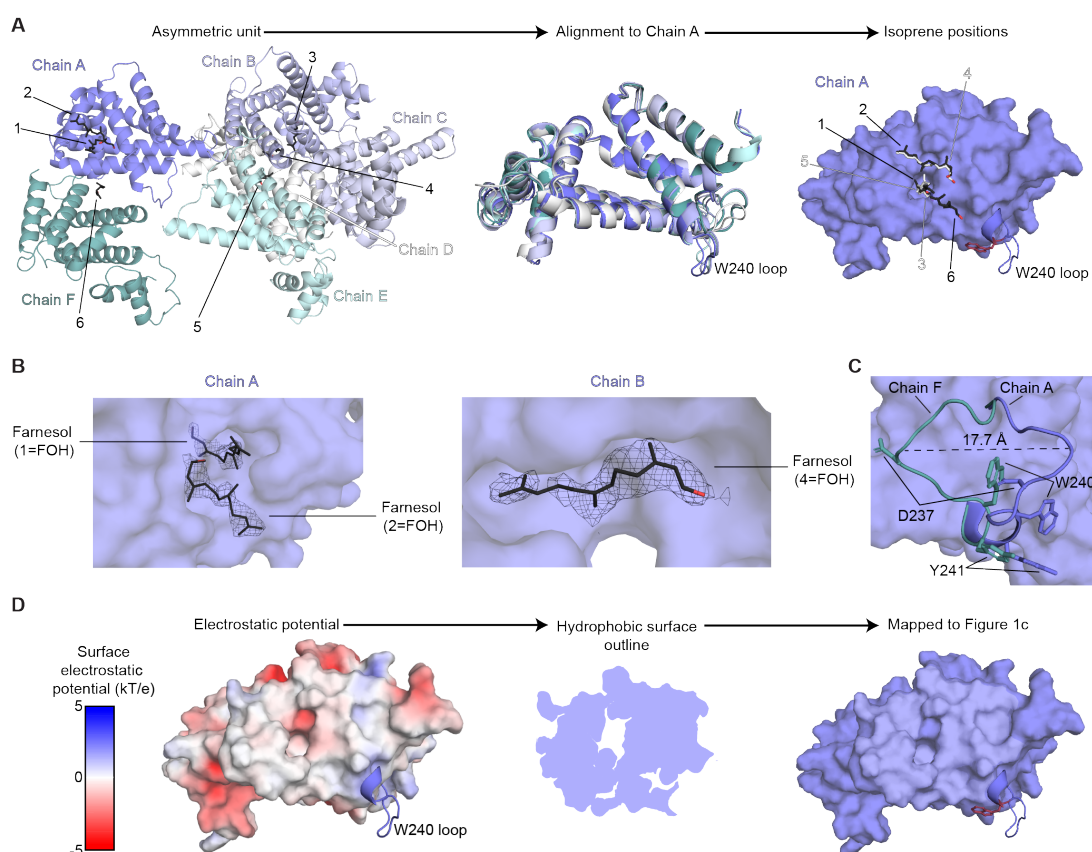


Figure 4.4 – Crystal Structure of COQ9 with Geranylgeraniol and Extended Helix 10, Related to Figures 4.3 and 4.5. (A) Six modeled isoprene positions (black, #1-6) on six COQ9 chains (A-F) in the asymmetric unit (PDB: 6DEW), where Chains A-F are in blue, light blue, blue-white, white, light cyan, and light teal respectively. All chains and co-crystallizing isoprenes were then aligned to Chain A. Isoprene positions shown in Fig. 4.3 are in black and those co-crystallizing in similar positions are in white. (B) Composite simulated annealing omit maps showing electron density (1σ) for the most ordered isoprenes (FOH ligands in Chains A and B). Similar electron density that was somewhat less well defined was observed at the same sites in all of the protomers. (C) Overlay of COQ9^{NΔ79,CA31} chains A (blue) and F (teal), where conformations of the $\alpha 7$ - $\alpha 8$ loop are in cartoon and COQ9 residues important for COQ7 binding (D237, W240, Y241) are highlighted. The difference between loop conformations is 17.7 Å, at its widest point (dotted line). (D) Electrostatic potential of COQ9 Chain A generated by the Adaptive Poisson-Boltzmann Solver (Jurrus et al., 2018). The $\alpha 7$ - $\alpha 8$ loop is excluded from the surface. Negative electrostatic potential is red, positive is blue, and zero is white. The hydrophobic surface was outlined and mapped to Figure 4.3C.

Additionally, the $\alpha 7$ - $\alpha 8$ loop that contains W240 adopted multiple conformations in the crystal lattice, moving up to 17.7 Å, indicating dynamic movement of the residue that conferred ligand specificity above (Figure 4.4C). The multiple isoprene positions suggest a model whereby the long-chain isoprene of a CoQ-like molecule would occupy the hydrophobic regions of COQ9 while the aromatic head group extends towards a flexible W240 (Figure 4.3D).

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

Together, the co-purification of COQ9 with CoQ intermediates from an endogenous context in yeast, its selective co-purification with several aromatic isoprenes from bacteria, and its crystallization with an isoprene lipid suggest that binding CoQ isoprene intermediate(s) is central to COQ9's molecular function. To our knowledge, this is the first lipid binding protein with specificity for this lipid class.

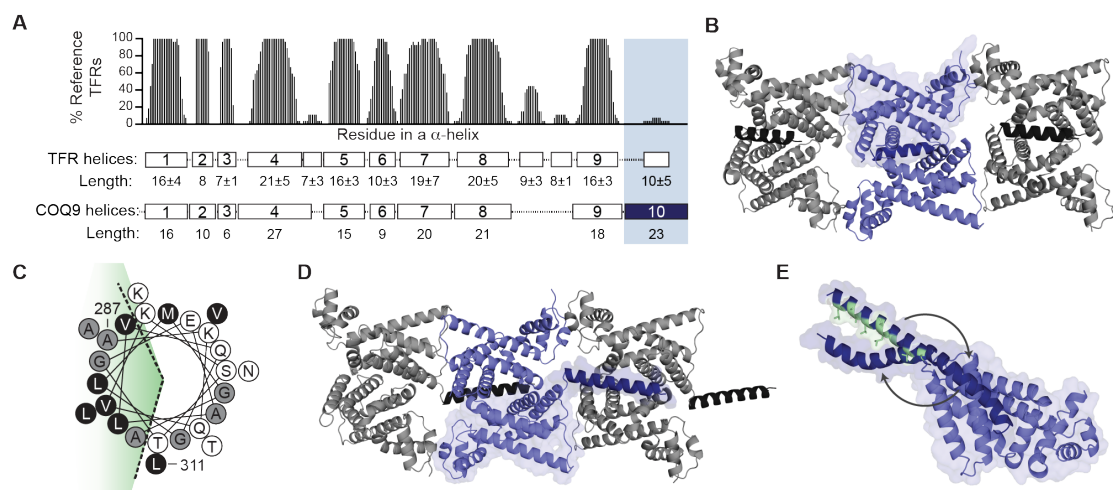


Figure 4.5 – The 10th α -Helix of COQ9 is Unique, Amphipathic, and Flexible. (A) Consensus helices of TetR Family Regulators with average length and standard deviation across 27 representative structures (Yu et al., 2010). COQ9 helices and lengths are mapped below with $\alpha 10$ in dark blue. The presence and length of COQ9's $\alpha 10$ is uncommon among TFRs. (B) Modeled position of $\alpha 10$ in the original COQ9 structure (PDB: 4RHP) where the $\alpha 9$ - $\alpha 10$ loop was not resolved. Symmetry mates are in gray, a single COQ9 protomer is outlined in surface, and $\alpha 10$ is in either dark blue or black. (C) Helical wheel of $\alpha 10$ (A287-L311) with the hydrophobic face highlighted in green, hydrophobic residues in black, A and G in gray, and polar residues in white. (D) Position of $\alpha 10$ in the structure reported in this work (PDB: 6AWL) where the $\alpha 9$ - $\alpha 10$ loop was resolved and shows that $\alpha 10$ extends away from globular COQ9 to form crystal contacts with an adjacent COQ9 molecule. Symmetry mates are in gray, a single COQ9 protomer is outlined in surface, and $\alpha 10$ is in either dark blue or black. (E) Snapshots of COQ9 with three positions of $\alpha 10$ (dark blue) sampled during CG-MD solution simulations, where hydrophobic residues are green. See also Figure 4.4 and Table 4.2.

4.3.2 The Amphipathic 10th α -Helix of COQ9 Drives Membrane Association

Within mitochondria, CoQ isoprene intermediates are thought to be either embedded in the inner mitochondrial membrane (IMM), like mature CoQ, or to be channeled through the membrane-associated complex Q in a 'metabolon'-like fashion.¹⁶ We sought to define how COQ9 might access CoQ intermediates from the IMM and how COQ9 accessed the membrane-embedded lipids reported here (Figure 4.3). In addition to its core TetR fold—a

base architecture that can enable binding of specific metabolites, including hydrophobic polyketides and fatty acids^{122, 134, 135}—COQ9 also possesses a distinct, amphipathic C-terminal 10th α -helix (α 10) not typical among TFR proteins (Figures 4.5A, 4.5C).

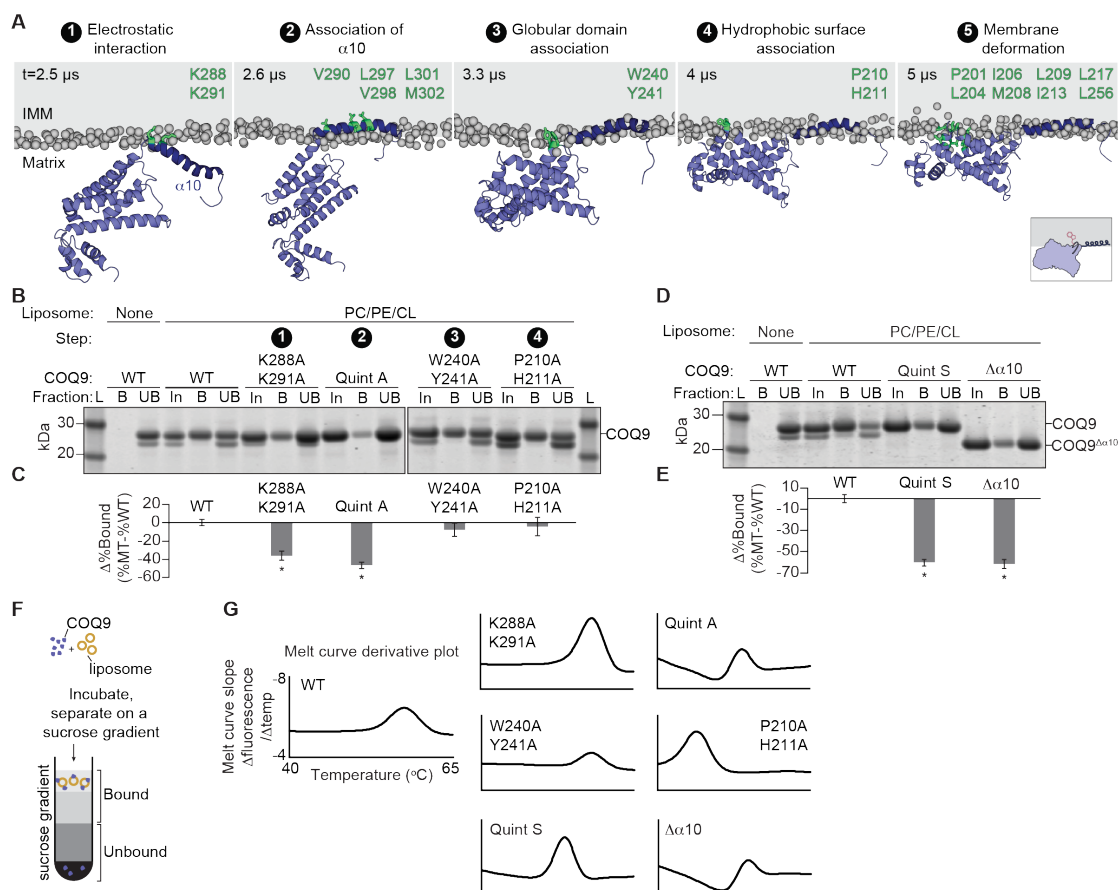


Figure 4.6 – COQ9 Associates with Membranes in 5 Steps. (A) Snapshots representative of the multi-step COQ9 membrane binding mechanism observed in CG-MD simulations (phospholipid head groups in gray and membrane-interacting residues of COQ9 in green; COQ9 model based on PDB: 6AWL). (B) SDS-PAGE of liposome co-floitation of COQ9^{N Δ 79} alanine mutants with PC/PE/CL/NBD-PE (68.9/9/22/0.1%) liposomes where L is ladder; In, protein input; B, liposome bound protein; and UB, unbound protein. (C) Quantification of liposome bound COQ9 alanine mutants (MT) in panel B relative to WT by densitometry ($100 \times (\text{BMT}/\text{TotalMT} - \text{BWT}/\text{TotalWT})$) where * indicates a significant (p-value < 0.05, n=3) difference from WT binding and error bars represent standard deviation. (D) SDS-PAGE of liposome co-floitation of COQ9 α 10 mutants in PC/PE/CL/NBD-PE (68.9/9/22/0.1%) liposomes where Quint S is V290S, L297S, V298S, L301S, M302S, $\Delta\alpha$ 10 is COQ9^{N Δ 79, C Δ 31}. (E) Quantification of liposome bound COQ9 mutants in panel D relative to WT by densitometry ($100 \times (\text{BMT}/\text{TotalMT} - \text{BWT}/\text{TotalWT})$) where * indicates a significant (p-value < 0.05, n=3) difference from WT binding and error bars represent standard deviation. (F) Liposome co-floitation schematic where liposome bound (B) and unbound (UB) fractions quantified by SDS-PAGE densitometry are shown. (G) Thermal melt curve derivative plots ($\Delta\text{fluorescence}/\Delta\text{temperature}$ from 40–65 $^{\circ}\text{C}$) for proteins assayed by liposome co-floitation as measured by differential scanning fluorimetry. All plots are scaled as shown in WT and the thermal melting temperature (T_m) is at the apex of the melt curve derivative plot.

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

Similar structural features are used by other lipid binding proteins to drive their interactions with membranes,^{136–138} leading us to hypothesize that $\alpha 10$ has a role outside the canonical TetR fold, potentially to localize COQ9 to membrane-bound cargo in the IMM. We determined the structure of the $\alpha 10$ -containing COQ9^{N Δ 79} to 2 Å resolution and discovered that $\alpha 10$ extends away from the rest of COQ9, forming a crystal contact at the interface of two adjacent COQ9 chains in the crystal lattice (Figure 4.5D; Table 4.2, PDB: 6AWL). Furthermore, coarse-grained molecular dynamics (CG-MD) simulations of COQ9 in solution revealed a dynamic 200-degree movement of $\alpha 10$ where $\alpha 10$ collapsed on the hydrophobic surface of COQ9 within 3 μ s, moved across this surface for 7 μ s, and eventually re-extended (Figure 4.5E).

To examine the mechanism of COQ9 membrane association, including a potential role for $\alpha 10$, we performed CG-MD simulations of COQ9 with membrane models comprised of phosphatidylcholine (PC), phosphatidylethanolamine (PE), and cardiolipin (CL) to resemble the IMM (Daum, 1985). In these simulations, COQ9 invariably associated with the membrane in a multi-step fashion (Figure 4.6A). COQ9 first diffused in the aqueous environment until two lysines (K288, K291 in Step 1) in the $\alpha 9$ - $\alpha 10$ loop interacted electrostatically with phosphate head groups of negatively charged cardiolipin. Second, hydrophobic residues of $\alpha 10$ (V290, L297, V298, L301, M302 in Step 2) interacted with the membrane, reaching the glycerol backbones of membrane phospholipids. Next, the protein core ($\alpha 1$ - $\alpha 9$) approached the membrane, first with $\alpha 8$ (W240, Y241 in Step 3) and then with $\alpha 7$ (P210, H211 in Step 4). Finally, residues along the hydrophobic surface of COQ9 were exposed to the membrane (P201, L204, I206, M208, L209, I213, L217, L256 in Step 5).

To investigate the step-wise membrane binding mechanism biochemically, we tested whether COQ9 alanine mutants of residues contributing to Steps 1-4 affected liposome co-floitation (Figures 4.6A-4.6C, 4.6F). Consistent with simulations, neutralizing the electrostatic interaction in Step 1 decreased membrane association, while mutations to residues mediating Steps 3 and 4 did not (Figures 4.6B, 4.6C). Alanine mutants of $\alpha 10$ hydrophobic residues in Step 2 reduced liposome association the most (Figures 4.6B, 4.6C). Similarly, mutating $\alpha 10$ hydrophobic residues to serine or fully deleting $\alpha 10$ reduced liposome association by more

than 60% (Figures 4.6D, 4.6E). Importantly, all tested mutants had robust thermal melt curves, indicating stably folded protein (Figure 4.6G).

4.3.3 COQ9 Association with Cardiolipin-Rich Membranes Warps the Lipid Bilayer

COQ9 membrane association was strongly influenced by CL, a quintessential mitochondrial lipid with key roles in cristae organization and respiratory supercomplex formation.^{12, 139–141} In CG-MD simulations featuring a simple PC-based membrane model, COQ9, although approaching the membrane several times, was not able to establish a specific and stable interaction with the membrane (Figures 4.7A, 4.7B). Consistently, cardiolipin specificity was observed biochemically in liposome co-floitation assays (Figures 4.7C-4.7H).

Our simulations also revealed striking features of membrane lipid organization upon COQ9 association. In Step 5, the hydrophobic surface of COQ9 displaced lipid head groups, creating a deformation in the membrane, an area void of phosphate head groups, and a higher concentration of CL surrounding $\alpha 7$, $\alpha 8$ and $\alpha 10$ (Figure 4.8A). Lipids interacting with COQ9 had low diffusion coefficients, where recurrently binding CL molecules persisted for over 5 μ s and established favorable interactions with residues K288 and W240 (Figure 4.8B). The local deformation of the membrane surface caused lipids surrounding COQ9 to be orthogonally displaced from the membrane surface by 5 Å while their hydrophobic tails entered the hydrophobic cavity of COQ9 (Figures 4.8C, 4.8D).

Such lifting of lipids off the membrane by a peripheral membrane protein is reminiscent of, although distinct from, the mechanism observed in a previous MD study reporting the partial extraction of a single PC molecule into the hydrophobic pocket of phosphatidylinositol transfer protein- α (PITP α).¹⁴² These observations in CG-MD simulations provide a potential molecular mechanism for how COQ9 extracted lipid cargo in the adventitious co-crystallization with phospholipids and in the specific co-purification with aromatic isoprenes. Collectively, our computational and biochemical findings support a model whereby the amphipathic $\alpha 10$ drives membrane association in a multi-step fashion and poises the hydrophobic surface of COQ9 to deform the IMM in order to access membrane-bound lipids.

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

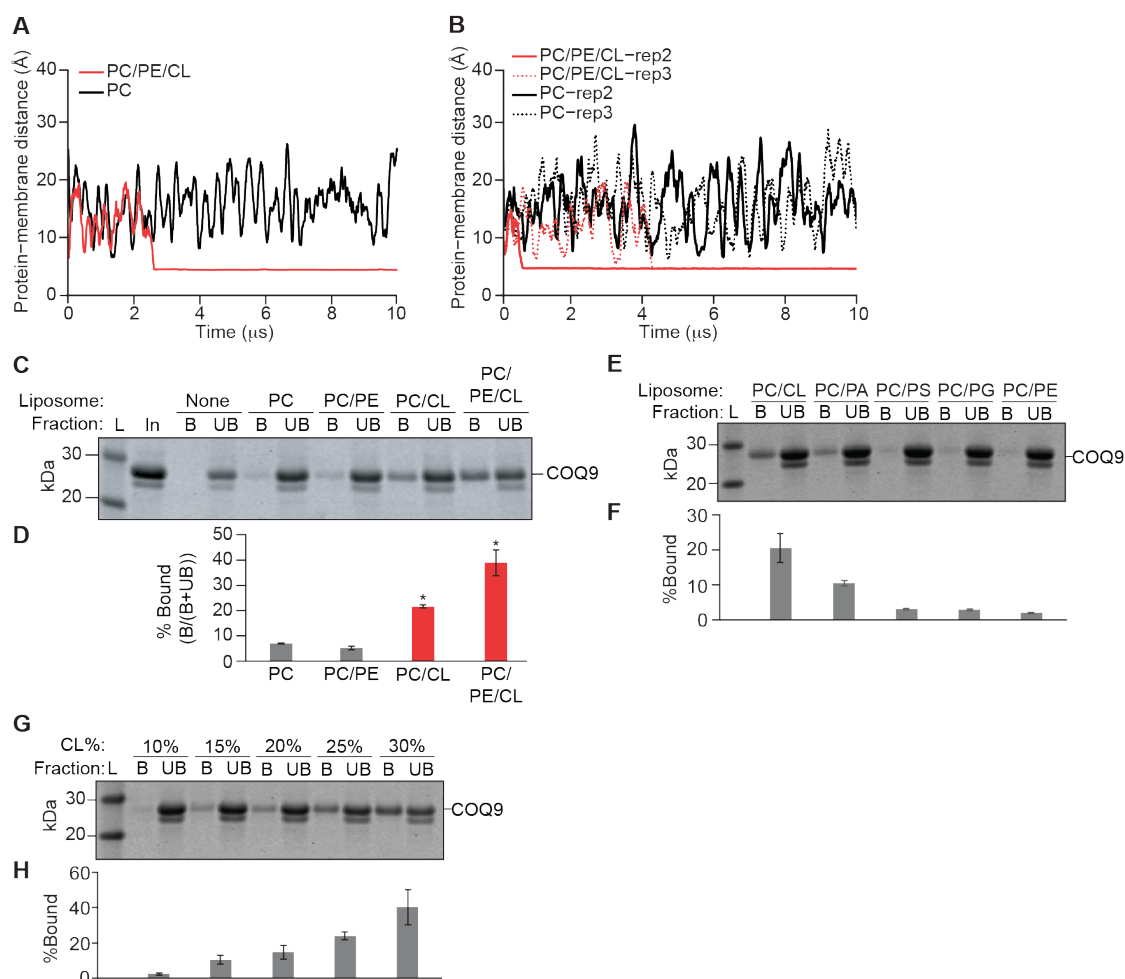


Figure 4.7 – COQ9 Membrane Association is Strongly Influenced by Cardiolipin. (A) Time evolution of the minimum pairwise distance between COQ9 and the membrane (PC (black) or PC/PE/CL (red)) in CG-MD simulation replica 1. (B) Time evolution of the minimum pairwise distance between COQ9 and the membrane (PC (black) or PC/PE/CL (red)) in CG-MD simulation replicas 2 (solid) and 3 (dotted). (C) SDS-PAGE of liposome co-floitation of COQ9^{NΔ79} with PC/NBD-PE (99.9/0.1 molar %), PC/PE/NBD-PE (90.9/9/0.1%), PC/CL/NBD-PE (77.9/22/0.1%), or PC/PE/CL/NBD-PE (68.9/9/22/0.1%) liposomes, as described in Figure 4.6. (D) Quantification of liposome bound COQ9 in panel C by densitometry (100×B/Total) where * indicates a significant (p-value < 0.05, n=2) difference from PC liposomes and error bars represent standard deviation. (E) SDS-PAGE of liposome co-floitation of COQ9^{NΔ79} in PC/CL/NBD-PE, PC/PA/NBD-PE, PC/PS/NBD-PE, PC/PG/NBD-PE, or PC/PE/NBD-PE (all 77.9/22/0.1 molar %) liposomes. (F) Quantification of liposome bound COQ9 in panel E by densitometry (100×B/Total) where error bars represent standard deviation (n=2). (G) SDS-PAGE of liposome co-floitation of COQ9^{NΔ79} with PC/CL/NBD-PE liposomes, with increasing CL molar % (10,15,20,25,30) where NBD-PE is at a constant 0.1 molar %, as described in Figure 4.6. (H) Quantification of liposome bound COQ9 in panel G by densitometry (100×B/Total) where error bars represent standard deviation (n=2).

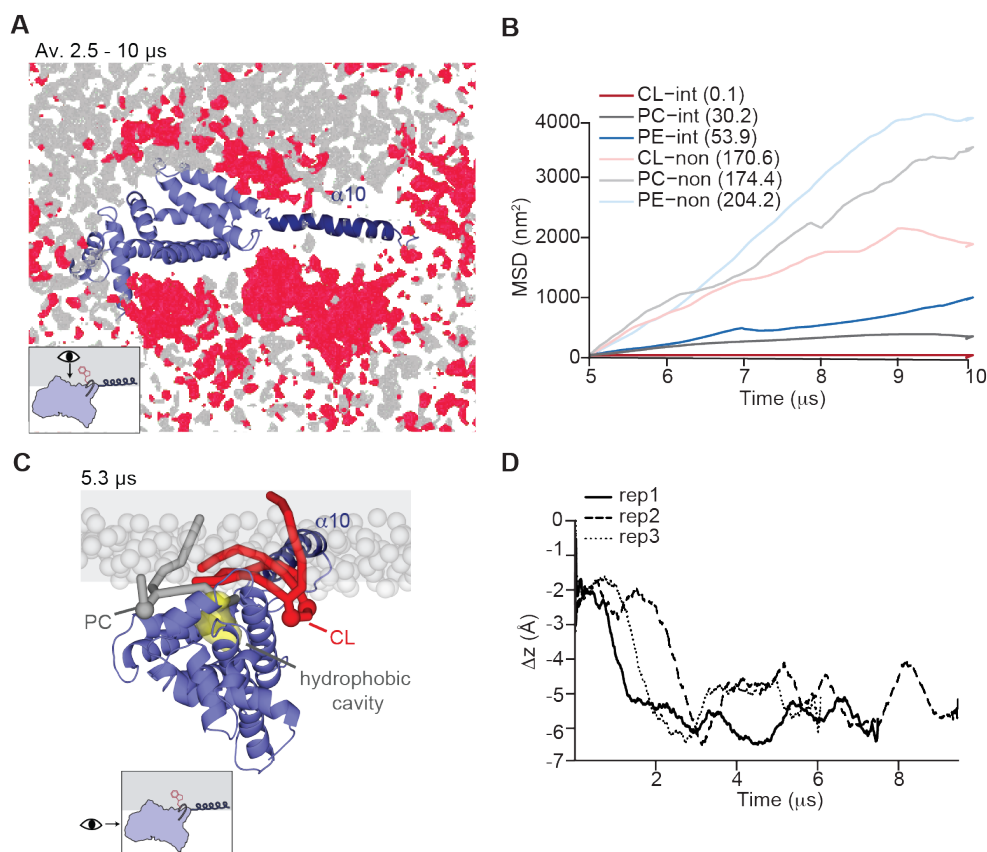


Figure 4.8 – COQ9 Warps Cardiolipin-Containing Membranes. (A) The occupancy of lipid phosphate head groups averaged throughout a 7.5 μs CG-MD simulation during membrane interaction, with phosphate head groups of CL in red and PC/PE in gray, viewed from inside the membrane. (B) Mean square displacement of representative CL (red), PC (gray), PE (blue) lipids, for those that establish a stable interaction with COQ9 (“int,” dark colors) and never interact with COQ9 (“non,” light colors) for a PC/PE/CL membrane (replica 1). Lateral diffusion coefficients are given in $\text{nm}^2/\mu\text{s}$ in parentheses in the legend for replica 1. Lateral diffusion coefficients in $\text{nm}^2/\mu\text{s}$ for replica 2: CL-int: 0.4, PC-int: 22.4, PE-int: 7.5, CL-non: 90.5, PC-non: 157.6, PE-non: 136.5, and for replica 3: CL-int: 2.3, PC-int: 9.1, PE-int: 7.2, CL-non: 112.3, PC-non: 131.0, PE-non: 123.9. (C) Snapshot from a CG-MD simulation after Step 5 where CL (red) and PC (gray) acyl tails enter the COQ9 hydrophobic cavity (yellow), viewed from the N-terminus of COQ9. (D) The z-shift of lipid phosphate head groups that are within 12 \AA of membrane-bound COQ9 relative to the base plane of membrane lipid phosphate head groups, for replicas 1 (solid), 2 (dashed), and 3 (dotted). For each replica, $t=0 \mu\text{s}$ refers to the beginning of a stable interaction of the protein with the membrane (Step 4 in Fig. 4.6A); therefore, each replica has a different total membrane interaction time.

4.3.4 COQ9 Interaction with Aromatic Isoprenes Illuminates Mechanisms of Specificity

Investigating the interaction of COQ9 with membrane-embedded isoprene molecules could illuminate mechanisms of specificity. To investigate how CoQ interacts with COQ9, we per-

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

formed CG-MD simulations of COQ9 with a PC/PE/CL/CoQ₁₀ membrane.

In this new simulation, as observed previously with PC/PE/CL membranes, COQ9 created an area void of lipid phosphate heads and all four CoQ molecules migrated to this "tightly sealed" hydrophobic region created by COQ9 (Figure 4.9A). Once COQ9 associated with the membrane, one CoQ molecule quickly occupied the hydrophobic pocket of COQ9 and established a very stable interaction (residence time > 10 μ s, Figure 4.9B). Other CoQ molecules interacted more transiently with the hydrophobic surface of COQ9. The interaction was always through the CoQ head group, whereas the isoprene tail extended towards the membrane (Figure 4.9A). This simulation supports a model whereby, COQ9 acts as an attractor or a hub for CoQ molecules, creating a favorable environment for them. Our simulation also highlights possible modes of entry and mechanisms of specificity for isoprene lipids into the COQ9 pocket. Namely, the simulations revealed aromatic stacking interactions between W240 and multiple CoQ molecule headgroups (Figure 4.9B).

The multiple isoprene positions in the geranylgeraniol-bound COQ9 crystal structure could serve as an independent check of the accuracy of our simulations. We therefore compared the CoQ poses from our simulation to the geranylgeraniol poses in the crystal structure. Our reasoning for not using geranylgeraniol in our simulations is that (1) geranylgeraniol is not an endogenous ligand of COQ9 (it was used in co-crystallization experiments because of its higher solubility) and (2) geranylgeraniol is likely to behave differently in a simulated membrane than in co-crystallization experiments, which lack a membrane and the corresponding COQ9:membrane interactions. Generally, our crystallography and simulations were highly consistent with the COQ9 hydrophobic surface and cavity binding CoQ-like molecules (Figures 4.9C-E).

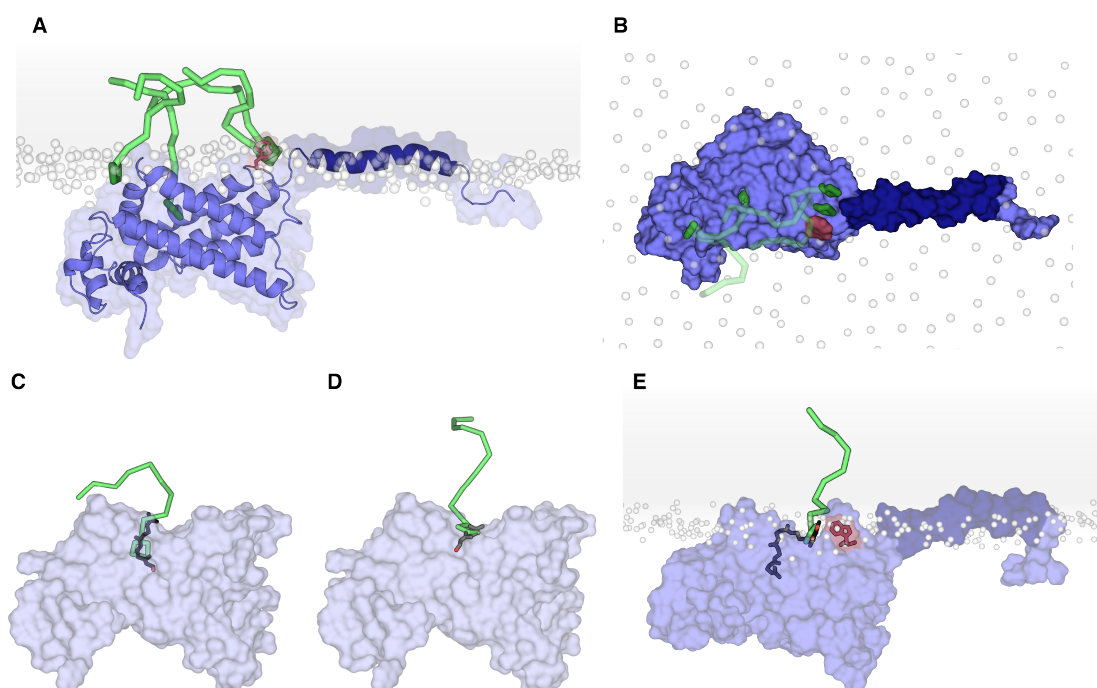
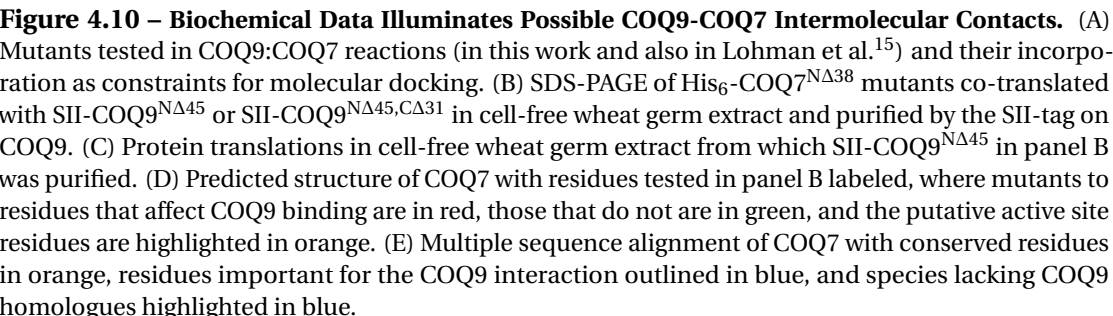


Figure 4.9 – COQ9 Creates a Hub for CoQ Molecules. (A) Side view of all four CoQ molecules (head groups shown in dark green and tails in light green) interacting with COQ9. (B) Top view of all four CoQ molecules (head groups shown in solid green and tails in transparent green) interacting with COQ9, showing one CoQ molecule bound to the hydrophobic pocket (yellow), and aromatic stacking interactions between two CoQ molecules and W240 (magenta). Lipid phosphate head groups are shown in white. (C-E) Comparison of CoQ poses from CG-MD simulations with isoprene poses from crystallography (PDB: 6DEW) or with the docking pose of DMQ₅. (C) Overlay of a CoQ pose from the CG-MD simulation (green) with the farnesol molecule (black) in the hydrophobic cavity of COQ9. (D) Overlay of a CoQ pose from the CG-MD simulation with the geraniol molecule (gray) on the hydrophobic surface of COQ9. (E) Overlay of a CoQ pose from the CG-MD simulation with the docking pose of DMQ₅ (black) with W240 in red and lipid phosphate head groups in white.

4.3.5 COQ9 Interaction and Interface with COQ7 Suggest a Lipid Presentation Role

Prior studies in human patients,^{112,113,117} mice,^{15,114,116,143} and yeast^{115,144,145} have suggested that COQ9 may be especially important for the C6-hydroxylation step of CoQ biosynthesis performed by the di-iron carboxylate hydroxylase COQ7.^{125,146–148} Previously, an *in vitro* interaction between COQ9 and COQ7 was identified, revealing through mutagenesis that a conserved surface patch of COQ9 (D237, W240, Y241 in the flexible α 7- α 8 loop, Figure 4.4C) mediates COQ7 binding.¹⁵

A

B

Here, we identified residues of COQ7 that, reciprocally, are required for COQ9 binding (Figures 4.10A-4.10D). We designed 13 COQ7 point mutants based on evolutionary conservation and predicted solvent accessibility, and we discovered that mutants within a predicted loop of COQ7 and mutants at the C-terminus fail to bind COQ9 (Figures 4.10B-4.10D). Interestingly, some of these residues, although generally conserved, are absent in certain nematodes that lack COQ9 homologs (Figure 4.10E).

We used these biochemical data reporting on possible COQ9-COQ7 intermolecular contacts combined with the orientation of COQ7 and COQ9 on the IMM identified by our individual simulations (Figure 4.11A) to determine a potential molecular interface of the complex. The modeled conformation that fulfills these spatial constraints places the COQ9 hydrophobic surface (where isoprenes bind) adjacent to the COQ7 active site (Figures 4.11B, 4.11E). The protein-protein interface comprises $\alpha 7$, $\alpha 8$, and the $\alpha 7$ - $\alpha 8$ loop of COQ9 with $\alpha 1$, $\alpha 2$, and a loop of COQ7, where key contacting interfacial residues include W240 and Y241 of COQ9, and P108 and W115 of COQ7.

To visualize this complex in a more endogenous context, we incorporated a partial COQ7 substrate (demethoxy CoQ with 5 isoprene units, DMQ₅) into our COQ9 model associated with the IMM (Figures 4.11B, 4.11E, 4.11G).^{109, 145} In this conformation, molecular docking predicted that two isoprene units of DMQ₅ fit into the COQ9 hydrophobic cavity in a pose consistent with a co-crystallized isoprene (RMSD=1.5 Å, Figures 4.10C, 4.11C). The rest of the DMQ₅ tail lay on the hydrophobic surface such that the aromatic head group extends towards W240—the residue that conferred aromatic isoprene binding specificity above (see Figure 4.3) (Figures 4.11C-4.11E, 4.11G). The predicted architecture of the membrane-bound COQ9-COQ7 complex suggests that COQ9 could lift the COQ7 substrate from the membrane and present it to the COQ7 active site, similar to lipid metabolic nanoreactors.^{149, 150} Moreover, we again observed in simulations an increased CL occupancy around COQ7 (Figure 4.11F), as was observed for COQ9 simulations (Figure 4.8A), suggesting that this lipid species can be important for mediating the interaction of these proteins with the IMM and thus facilitating COQ9-COQ7 complex formation.

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

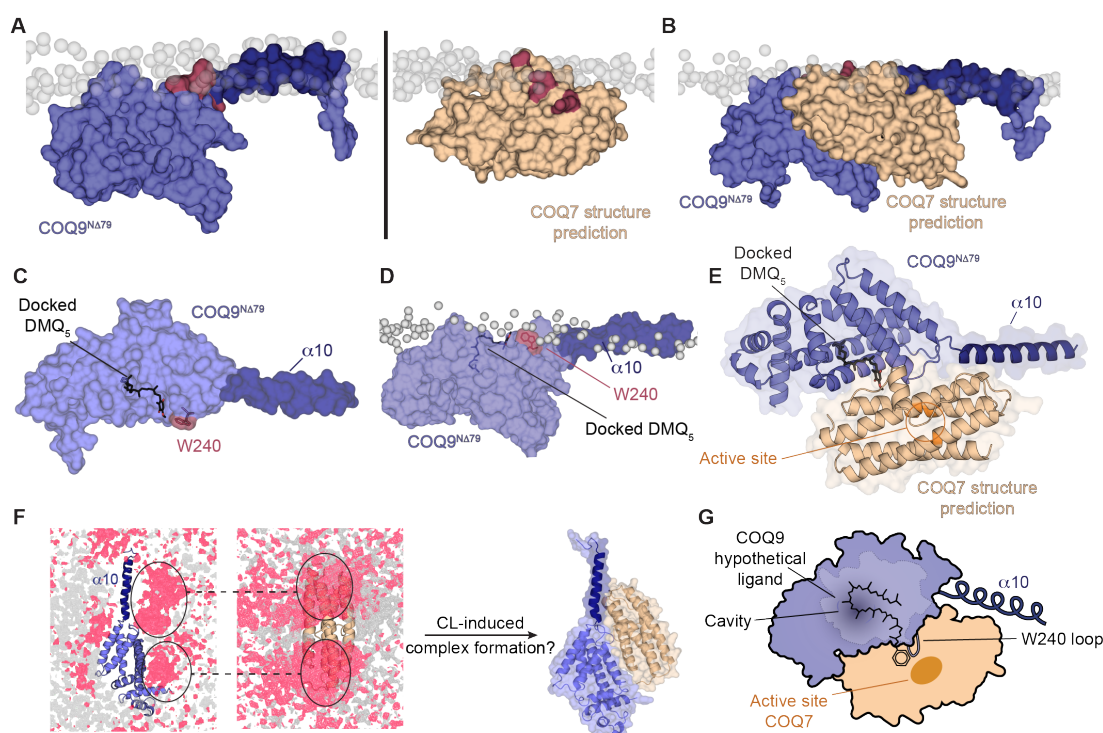


Figure 4.11 – COQ9 and COQ7 Predicted Molecular Interface Suggests Substrate Presentation. (A) Membrane binding CG-MD simulations of COQ9 and COQ7 with protein-protein interface residues in red. (B) Predicted COQ9-COQ7 complex with the IMM overlaid. (C) Molecular docking of DMQ₅ (stick) on COQ9^{NΔ79} (PDB: 6AWL, Chain A) where $\alpha 10$ (dark blue) and W240 (red) are labeled. (D) Molecular docking of DMQ₅ (stick) superimposed on the membrane-bound conformation of COQ9^{NΔ79} where W240 is shown in red, membrane phospholipid head groups in gray, and $\alpha 10$ in dark blue. (E) Predicted model of the COQ9-COQ7 complex with DMQ₅ docked to COQ9 overlaid. (F) Occupancy of lipid phosphate head groups in CG-MD simulations of COQ9 (left) and COQ7 (middle) with CL in red and PC/PE in gray, where CL-rich regions are circled and COQ9 and COQ7 are oriented as in the modeled COQ9:COQ7 complex (right). (G) Model for how COQ9 could potentially present substrate to COQ7, viewed from inside the membrane.

4.3.6 COQ9 Lipid- and Membrane-Binding Residues are Critical for CoQ Production *In Vivo*

The work above identifies structural features of COQ9 predicted to facilitate CoQ intermediate selection from the IMM and delivery to COQ7. To test the importance of these features *in vivo*, we engineered two Coq9p-FLAG mutants in *S. cerevisiae*: W188K (analogous to W240K in humans, to disrupt aromatic isoprene and COQ7 binding) and $\Delta\alpha 10$ (to disrupt membrane binding) (Figure 4.12A). Both mutants failed to rescue respiratory growth in $\Delta coq9$ *S. cerevisiae* (Figures 4.12B, 4.13A), indicating their importance for core COQ9 function.¹⁵ Additionally,

Coq9p chimeras possessing other amphipathic or transmembrane domains in lieu of its $\alpha 10$ could not rescue $\Delta coq9$ respiratory growth defects (Figures 4.14A-D).

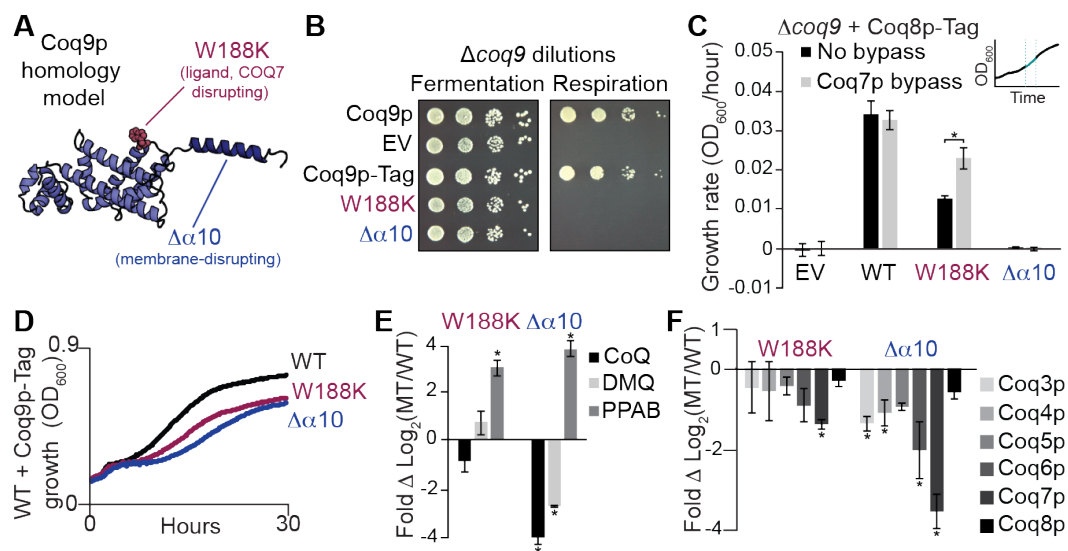


Figure 4.12 – Coq9p Protein-, Lipid- and Membrane-Disrupting Mutants Perturb Endogenous CoQ Biosynthesis. (A) Homology model of *S. cerevisiae* Coq9p with W188K (lipid- and COQ7-disrupting mutant that is analogous to W240K in human) in red and $\Delta\alpha 10$ (Coq9pC Δ 28, membrane-disrupting mutant) in blue. (B) Genetic complementation of $\Delta coq9$ *S. cerevisiae* in fermentation (CoQ-independent) and respiration (CoQ-dependent) conditions where EV is empty vector and mutants have a C-terminal FLAG tag. (C) Respiration growth rates (OD_{600}/hr , linear slope post-diauxic shift as shown in inset) of $\Delta coq9$ *S. cerevisiae* overexpressing Coq8p-FLAG along with Coq9p-FLAG mutants, and the Coq7p chemical bypass 2,4-dihydroxybenzoate (gray bars) where * indicates a significant difference (p -value < 0.05 , $n=3$) and error bars represent standard deviation. (D) Growth curves of WT *S. cerevisiae* overexpressing Coq9p-FLAG mutants in respiration conditions, plotted as an average of monoclonal transformants ($n=3$). (E) Fold changes of CoQ intermediates from monoclonal WT *S. cerevisiae* overexpressing Coq9p-FLAG mutants compared to WT Coq9p-FLAG. PPAB is 3-hexaprenyl-4-aminobenzoate, DMQ is demethoxy CoQ, and * indicates a significant difference (absolute value $\Delta\log_2 > 1$, p -value < 0.05 , $n=3$) from WT Coq9p-FLAG. (F) Fold changes of complex Q protein levels from monoclonal WT *S. cerevisiae* overexpressing Coq9p-FLAG mutants compared to WT Coq9p-FLAG, where * indicates a significant change (absolute value $\Delta\log_2 > 1$, p -value < 0.05 , $n=3$) from WT Coq9p-FLAG. See also Figures 4.13 and fig:q9-S7.

Previous work has shown that certain defects in CoQ biosynthesis can be bypassed by providing modified CoQ precursors that obviate the need for a specific enzyme^{116,143,151} (Figure 4.2A), or by overexpressing Coq8p,^{152,153} which confers stability to complex Q by an unclear mechanism. Overexpression of Coq8p-FLAG partially rescued $\Delta coq9$ yeast expressing W188K, but had no effect on $\Delta\alpha 10$, consistent with these mutants disrupting CoQ biosynthesis through distinct means (i.e., isoprene binding and membrane binding, respectively) (Figures 4.12C,

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

4.13B). Additionally, levels of the Coq7p substrate (DMQ) were particularly elevated in the Coq8p-FLAG + Coq9pW188K strain (Figure 4.13D), and 2,4-dihydroxybenzoate—a compound capable of specifically bypassing a Coq7p enzymatic defect—augmented the Coq8p-FLAG rescue of the W188K growth defect. Together, these results clearly indicate compromised Coq7p activity in this strain, as would be predicted from our experiments above. Notably, both strains contained CoQ intermediates that resembled those in $\Delta coq9$ + Coq8p-FLAG (including those associated with Coq6p defects, Figure 4.13D),^{144,153,154} which indicates Coq9p is important for other biosynthesis steps, possibly through a role in complex Q formation or stabilization. Interestingly, both mutants inhibited respiratory growth in WT yeast, demonstrating that they are both non-functional and also confer dominant-negative inhibition (Figures 4.12D, 4.13C). Mass spectrometry-based lipidomics and proteomics analyses revealed that these strains had decreased levels of CoQ, enhanced levels of an early CoQ biosynthetic intermediate (PPAB), and decreased levels of complex Q proteins, consistent with the predicted disruption in CoQ biosynthesis (Figures 4.12D-F). Together, these *in vivo* data further support the functional importance of COQ9 isoprene-, membrane-, and COQ7-binding, and suggest a model whereby COQ9 acts as a ‘lipid presenter’ to overcome hydrophobic barriers associated with CoQ biosynthesis.

4.4 Discussion

The spatiotemporal organization of lipids in metabolism, signaling, and membrane structure poses significant biophysical barriers within the largely aqueous cellular environment. Many strategies—from lipid transport proteins¹⁵⁵ to organelle contact site complexes^{156–158}—serve to minimize these barriers, and each begins with recognition of membrane-embedded cargo. CoQ biosynthesis, the bulk of which occurs within complex Q or the CoQ metabolon (a supramolecular complex of enzymes and structural elements¹⁵⁹), shares these hydrophobic barriers. Several additional features make CoQ biosynthesis a notable case study in lipid biology: the discovery of a key lipid-binding protein, the tethering of complex Q to the inner mitochondrial membrane, the proposed dynamic nature of complex Q, and the exceptional hydrophobicity of CoQ intermediates.^{16,109,152} We sought to further define the mechanism of COQ9 lipid binding and its role in enabling the CoQ biosynthetic process.

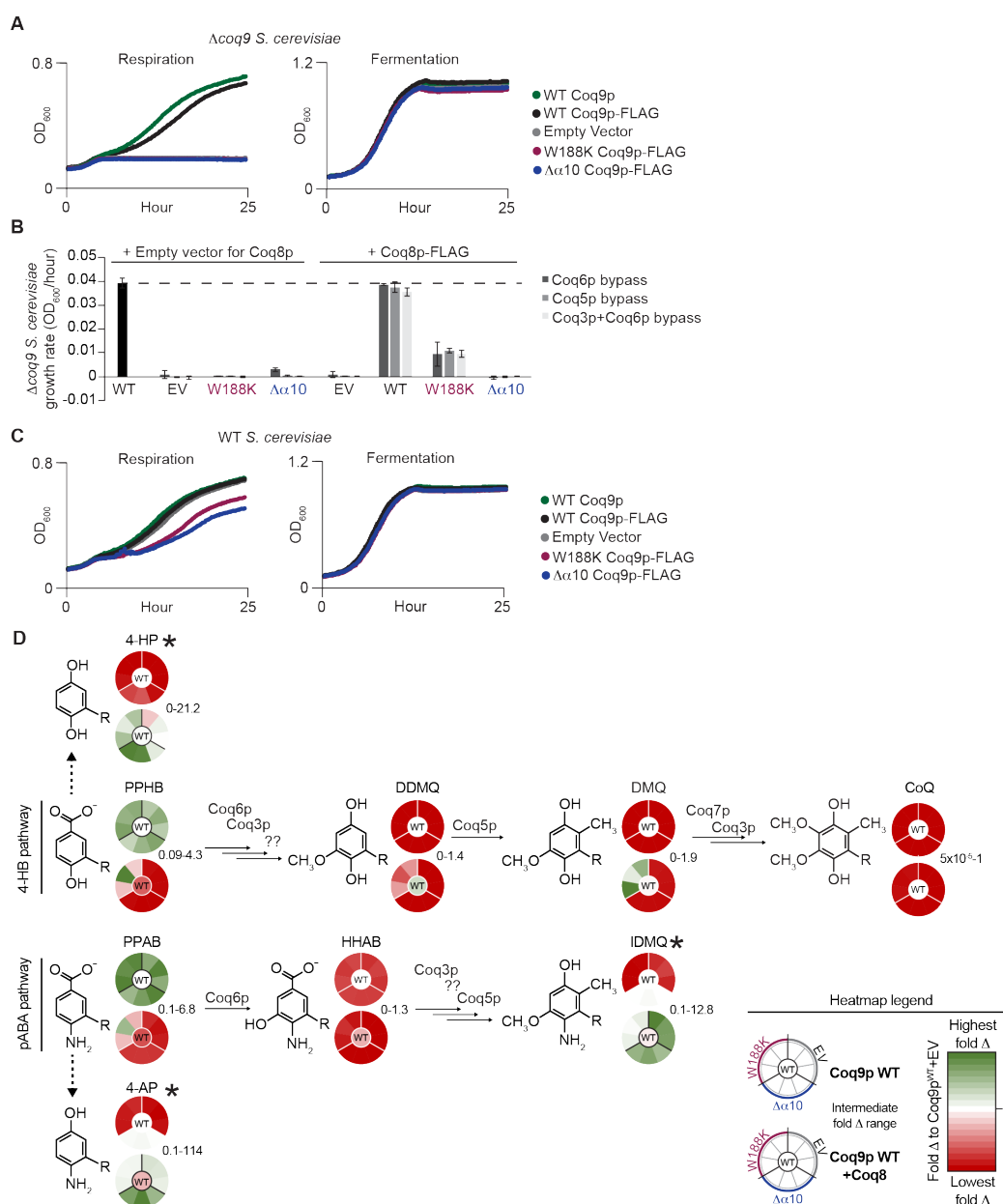


Figure 4.13 – Coq9p Mutant Growth Inhibition Profile, Related to Figure 4.12. (A) Growth curves of *Δcoq9 S. cerevisiae* overexpressing Coq9p mutants in respiration (CoQ-dependent) and fermentation (CoQ-independent) conditions. (B) Respiration growth rates (OD₆₀₀/hr post-diauxic shift) of *Δcoq9 S. cerevisiae* overexpressing Coq9p-FLAG mutants and combinations of Coq8p-FLAG, empty vector, and chemical bypasses for Coq6p, Coq5p, and Coq3p+Coq6p (see Figure 4.3). Overexpressed WT Coq9p-FLAG (black bar) is a reference for complete rescue (dotted line) and EV refers to empty vector for Coq9p expression (p416GPD). Error bars represent standard deviations of growth rates (n=3). (C) Growth curves of WT *S. cerevisiae* overexpressing Coq9p mutants, as described in panel A. (D) *S. cerevisiae* CoQ intermediates in Coq9p mutants. Levels of CoQ intermediates in *Δcoq9* yeast overexpressing either Coq9p^{WT}, empty vector (EV), Coq9p ^{$\Delta\alpha10$} , or Coq9p^{W188K} with and without Coq8p-FLAG co-expressed (n=3). Heat maps correspond to individual intermediate levels relative to *Δcoq9* + Coq9p^{WT} + EV (white) without Coq8p, where red shows a decrease, green shows an increase, and the layout for each heatmap is shown in the legend. Intermediates associated with Coq6p defects (4-HP, 4-AP, and IDMQ) are annotated with an asterisk.

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis Protein COQ9

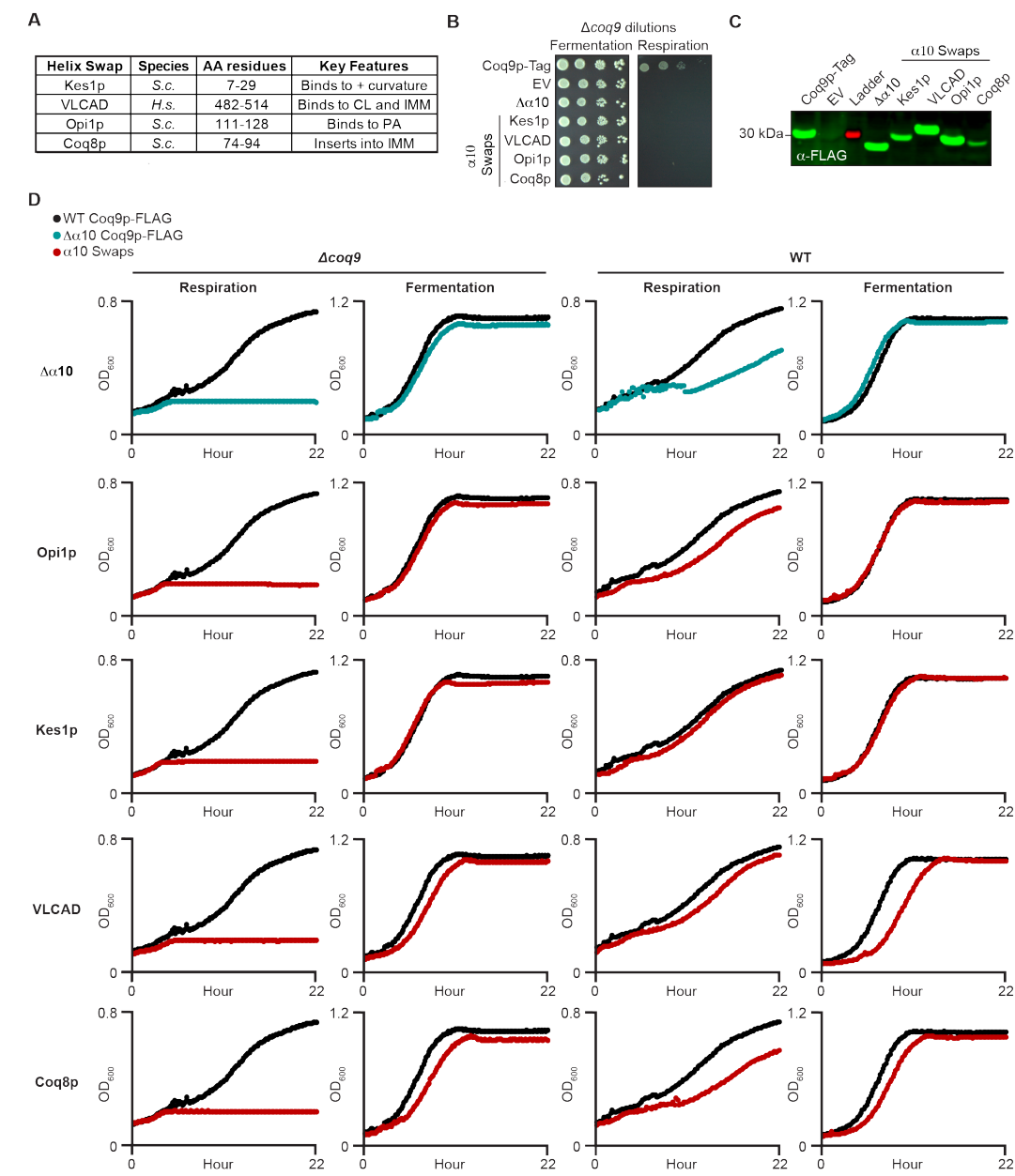


Figure 4.14 – Phenotypes of $\alpha 10$ Swap Chimeric Proteins, Related to Figure 4.12. (A) Table of amphipathic and transmembrane helix swaps: Kes1p,¹⁶⁰ VLCAD,^{161,162} Opi1p,¹⁶³ Coq8p,^{164,52} (B) Drop assay of $\Delta coq9$ yeast expressing helix swap chimeras or controls (WT Coq9p^{FLAG}, empty vector, $\Delta \alpha 10$ Coq9p^{FLAG}) on fermentation or respiration media. (C) Western blot (anti-FLAG) of whole cell WT yeast containing FLAG-tagged helix swap constructs. (D) Representative growth curves for helix swap chimeras (red dots) expressed in $\Delta coq9$ or WT yeast, in fermentation or respiration media. WT Coq9p^{FLAG} curves are shown in black dots, $\alpha 10$ Coq9p^{FLAG} in teal dots.

Recent advances in structural biology, bioinformatics, and *in vitro* biochemistry have equipped researchers with powerful tools to study how cells manage lipid metabolism and transport.^{155, 165, 166} Yet, defining the biochemical mechanisms that control lipid cargo selection faces the substantial challenge of studying protein function at the membrane-water barrier. Here, we combined molecular dynamics simulations with traditional biochemical and structural approaches to describe the stepwise mechanism for cargo selection by COQ9. We discovered that COQ9 uses a distinct amphipathic helix connected to a foundational ligand binding domain to associate with membranes. In our molecular simulations, COQ9 deforms the membrane bilayer by extruding polar head groups at the protein-IMM interface. During the simulation, this warping effect of the membrane surface allows lipid tails to be drawn into the COQ9 hydrophobic cavity at sites consistent with isoprene moieties found in our co-crystallization experiments. The creation of such a peculiar environment in the membrane—a continuous and tightly sealed hydrophobic region encompassing both the IMM and COQ9—could be key to solving the hydrophobicity problem of extracting CoQ intermediates from the membrane. This mode of action shares parallels with the “liftase” model proposed for how lipid transfer proteins (e.g. saposins,^{167, 168} GM2-AP,¹⁶⁹ and PlsC¹⁰⁶) overcome the topological challenge of reaching substrates located inside the membrane. Upon membrane binding, these proteins are envisaged to improve lipid accessibility and processing through modifying membrane organization (e.g. lifting lipids out of the membrane, distorting and destabilizing the membrane). We showed here, for the case of COQ9, that molecular simulations and biochemical approaches can be used together to molecularly dissect the mechanism of membrane reshaping and lipid trafficking adopted by lipid binding proteins.

This proposed role for COQ9 to retrieve CoQ intermediates from the bilayer is supported by its binding specificity for aromatic isoprenes. This specificity is guided by a highly conserved and mobile tryptophan (W240 in human or W188 in yeast), and the proximity of this residue to DMQ₅ in molecular docking experiments suggests that π - π stacking interactions could contribute to substrate specificity, to dynamic sensing, and/or to lipid presentation and transfer (Figure 4.3D). Importantly, CoQ intermediates are thought to be protected from water—whether they are embedded in the lipid bilayer, channeled through complex Q, or

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

passed along the periphery of the IMM by a lipid chaperone. Correspondingly, COQ9 could act as a lipid presenter at the membrane surface and/or between enzymes in complex Q.

Lipid-presenting proteins are notoriously challenging to characterize because of their dynamic nature, their dependence on other proteins, and their function at the membrane-water barrier. The case of Coq9p is further complicated by its concurrent role in complex Q stability, as dysfunctional CoQ proteins can cause the degradation of complex Q. Nonetheless, perturbations to Coq9p membrane ($\alpha 10$) and lipid binding (W188) residues severely disrupt *in vivo* CoQ biosynthesis by compromising the integrity of complex Q stability and by hindering Coq7p enzymatic activity. The specific connection between Coq9p function and Coq7p activity is reflected in patients with primary CoQ deficiency caused by mutations to COQ9.^{112,113} To date, six reports of COQ9 patient mutations result in decreased levels of COQ7 and the apparent accumulation of the COQ7 substrate, DMQ.^{112,113,117} Interestingly, one of the two reported patient mutations in COQ7 (L111P) is a point mutant at the predicted COQ9:COQ7 molecular interface.¹⁴³ Our predicted architecture of the membrane-bound COQ9-COQ7 complex suggests that COQ9 could lift the COQ7 substrate from the membrane and present it to the COQ7 active site—a model reminiscent of lipid metabolic nanoreactors,^{149,150} but one which will require further experimentation and structural analyses to validate.

In addition to mechanistic insights into how a peripheral membrane protein can extract membrane-embedded cargo, our work introduces a potential mechanism for the spatial organization of membrane lipids and, possibly, for complex Q seeding. In both our simulations and liposome co-flotation studies, CL had a role in the binding and dynamics of COQ9-membrane interactions. Indeed, we recently demonstrated that CL likewise enhances the membrane association and ATPase activity of COQ8, a second poorly characterized protein in the CoQ pathway.¹⁶⁴ If complex Q is localized to specific sites on the mitochondrial cristae, these findings suggest that its position and assembly could be influenced by local CL concentration. Cardiolipin is known to have a pivotal role for the formation and stability of several inner mitochondrial membrane protein complexes.^{12,139,140} Cardiolipin occupancy and persistency around COQ9 and COQ7 observed in our membrane binding simulations suggests that CL is

important for establishing and maintaining the specific interaction of these proteins with the IMM, and thus for facilitating COQ7-COQ9 complex formation. These observations suggest that cardiolipin-rich areas might template the membrane with a preferential surface pattern able to favor COQ9-COQ7 complex formation, as well as the recruitment of other members of complex Q. Further study of the localization, components, and connectivity of complex Q will inform these models.

Finally, there is speculation that complex Q is not a static metabolon, but that it might instead form upon substrate availability, and that CoQ or a pathway intermediate is part of the complex.^{109,170} Work from us and others suggest Coq8p could assist with sensing and delivering substrates based on its lipid interactions, ATPase activity, and apparent ability to stabilize complex Q.^{19,20,153,164} The structural homology of COQ9 with ligand-sensing transcription factors also makes COQ9 a plausible “sensor” candidate.¹⁵ Here, we introduce the flexible $\alpha 7$ - $\alpha 8$ and $\alpha 9$ - $\alpha 10$ loops that guide ligand interactions, COQ7-binding, and membrane interactions as possible key features in metabolon formation and activity.

Overall, our analyses present COQ9 as a new model for how nature employs peripheral lipid binding proteins to surmount hydrophobic challenges. These insights into how a protein can access, bind, and present a hydrophobic ligand may also inform new strategies to treat lipid deficiency disorders that are recalcitrant to supplementation, including many CoQ-related diseases. Moreover, our integrative approach combining structural, biochemical, and computational studies was effective at providing sub-nanometer resolution of COQ9 function at the protein-membrane interface and may thus serve as a model for interrogating other lipid binding proteins or lipid metabolism or signaling pathways whose investigations present similar challenges.

4.5 Supplementary Information

This Section details the work done by our experimental collaborators to support our computational findings. The experimental work included in this Chapter was led by Dr. Danielle C. Lohman and Prof. David J. Pagliarini in University of Wisconsin-Madison.

Saccharomyces cerevisiae strains (WT and $\Delta coq9$) were the haploid MATalpha BY4742 (his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0) from the gene deletion consortium (Thermo #YSC1054). *Escherichia coli* strains for protein overexpression were BL21[DE3]-RIPL from Agilent and for cloning applications, DH5 α from NEB were used. Unless specified otherwise, *E. coli* were grown at 37 °C in LB media with antibiotics and *S. cerevisiae* were grown at 30 °C in drop-out minimal media.

4.5.1 DNA Constructs and Cloning

Point mutants were generated by standard site-directed mutagenesis while insertions and deletions were generated by standard PIPE⁹² cloning methods.

4.5.2 Protein Expression and Purification

COQ9^{N Δ 79} and COQ9^{N Δ 79, C Δ 31} Expression and Purification

His₈-MBP-tev-COQ9^{N Δ 79} and mutants were expressed in *E. coli* (BL21[DE3]-RIPL strain) by autoinduction as previously described.¹⁵ Cells were isolated and resuspended in buffer A supplemented with 1 mg/mL Lysozyme (Sigma): 50 mM HEPES pH 7.5, 400 mM NaCl, 0.3 mM TCEP, 0.25 mM PMSE, and peptide protease inhibitors: 500 ng/mL leupeptin hemisulfate, pepstatin A, chymostatin, aprotinin, and antipain dihydrochloride at pH 7.2. Cells were lysed by sonication (4 °C), clarified by centrifugation (15,000 x g, 30 min, 4 °C), and bound to cobalt IMAC resin (TALON, Clontech, 1 hr, 4 °C). Resin was washed with Buffer A and his-tagged protein was eluted with Buffer A supplemented with 100 mM imidazole. Eluted protein was concentrated (50-kDa MW-cutoff spin filter, Merck Millipore Ltd.) and exchanged into Storage Buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.3 mM TCEP). For experiments with untagged

protein, His₈-MBP-tev-COQ9^{NΔ79} was incubated with TEV protease (prepared in-house, 1:50 TEV/fusion protein, mass:mass using $\epsilon_{280} = 102,130 \text{ M}^{-1}\text{cm}^{-1}$ and MW = 71.3 kDa, 2 hrs, 25 °C), then bound to cobalt IMAC resin (1 hr, 4 °C). Cleaved COQ9^{NΔ79} ($\epsilon_{280} = 36,130 \text{ M}^{-1}\text{cm}^{-1}$, MW = 27.4 kDa) was collected, concentrated (10-kDa MW-cutoff spin filter), and exchanged into Storage Buffer. Protein was aliquoted, frozen in N₂(l), and stored at -80 °C.

SII-COQ9^{NΔ45} and His₆-COQ7^{NΔ38} Expression and Purification

SII-COQ9^{NΔ45} and His₆-COQ7^{NΔ38} were expressed and purified in wheat germ extract (WE-PRO2240, CellFree Sciences, Matsuyama, Japan) as previously described.¹⁵ Transcription reactions (0.2 mg/mL DNA, 20 mM magnesium acetate, 2 mM spermidine trihydrochloride, 10 mM DTT, 80 mM HEPES-KOH pH 7.8, 4 mM each NTP pH 7.0, 1.6 U/ μ L SP6 RNA polymerase (Promega, Madison, WI), and 1 U/ μ L RNasin (Promega)) were incubated for 4 h at 37 °C. Single or mixed 1:1 COQ9:COQ7 transcription products were added to wheat germ extract in a dialysis cup reaction: 60 O.D. wheat germ extract, 24 mM HEPES-KOH, pH 7.8, 100 mM potassium acetate, 6.25 mM magnesium acetate, 0.4 mM spermidine trihydrochloride, 4 mM DTT, 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 0.0005% sodium azide, 0.04 mg/mL creatine kinase, 0.3 mM each amino acid at pH 7.0, and 5 μ L of RNA. Translation reactions were incubated (18 hr, 22 °C) in 32-fold excess dialysis buffer (without wheat germ extract, creatine kinase, or RNA). Duplicate translations were pooled and centrifuged (5 min, 20,000 \times g, 10 °C). Soluble protein was incubated with StrepTactin resin, washed in 25 mM HEPES pH 7.8, 150 mM NaCl, 1 mM DTT using a 96-well filter plate (HTS Multiscreen, Millipore), and eluted with 2.5 mM desthiobiotin.

4.5.3 Biochemical Assays

Differential Scanning Fluorimetry

DSF was performed as previously described.¹⁵ Reactions (20 μ L) of protein (2 μ M), SYPRO Orange dye (Invitrogen) (5X), NaCl (150 mM), HEPES (100 mM, pH 7.5) were prepared in Micro Amp Optical 96 (0.2 mL) Well Reaction Plates and sealed with Optical Adhesive Covers (Applied Biosystems). Fluorescence was measured (excitation filter λ 470 \pm 15 nm and emission filter

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

m4 623 ± 14 nm) during heating (15 to 95 °C, 0.017 °C/s) in Applied Biosystems QuantStudio 6 Flex Real Time PCR System. T_m values were calculated in Protein Thermal Shift software (Applied Biosystems).

Western Blot

Western blots of yeast whole cell lysate (4×10^7 cells) were probed with primary α -FLAG (Sigma-Aldrich F1804), secondary α -mouse IgG (RDye® 800CW), then visualized on a LI-COR imager.

Liposome Co-Flotation Assays

Liposomes were made with the following lipids from Avanti Polar Lipids reference: PC(#840051C); NBD-PE(#810145C); PE(#850725C); CL(#710335C). Unless otherwise specified, liposome compositions were PC/PE/CL/NBD-PE 68.9/9/22/0.1 molar percent. Lipid films were prepared by drying well-mixed lipids in organic solvent first under Ar(g) then by vacuum chamber (25-30 inHg, 17 hrs, 20 °C). The lipid film was reconstituted in HBS buffer (20 mM HEPES pH 7.5, 150 mM NaCl) to 10 mM total lipids by incubation (30 min, 30–35 °C) followed by 3 freeze/thaw cycles in N₂(l). Liposomes were extruded through 100 nm membranes (Avanti Polar Lipids, 11 passes, 30–35 °C) and used within 4 hours of preparation.

Liposome co-flotation was adapted from Connerth et al.⁹⁷ Liposomes (100 μ L) were incubated with protein (50 μ L) (10 min, 20 °C) and final reactions contained: 2.5 μ M protein, 6.66 mM liposomes, 20 mM HEPES pH 7.5, 150 mM NaCl. Sucrose (2.72 M, 110 μ L) was added to the protein liposome mixture and transferred (250 μ L) to an ultracentrifuge tube (Beckman #343776). A sucrose gradient in HBS was layered: 1.15 M sucrose with reaction (250 μ L), 0.86 M sucrose (300 μ L), 0.29 M sucrose (250 μ L), and 150 μ L HBS. After centrifugation (240,000 \times g, 1 h, 4 °C, Sorvall MX 120 Plus Micro-ultracentrifuge), fractions were removed from top (450 μ L) to bottom (450 μ L). Liposome content was quantified by NBD-PE fluorescence (excitation: 460 nm, emission: 535 nm) to validate flotation. Proteins in each fraction were CHCl₃:MeOH precipitated, following a method adapted from Wessel and Flugge⁹⁸ with methanol (1800 μ L) and chloroform (450 μ L). After vortexing, water (1350 μ L) was added and samples were vortexed and centrifuged (5 min, 4,000 \times g, 20 °C). The upper aqueous layer was discarded and

methanol (1000 μ L) was added to the protein disc. After mixing by inversion, samples were centrifuged (5 min, 20,000 \times g, 20 °C) and all liquid removed. The precipitated protein pellet was vacuum dried (25-30 inHg, 30 min, 20 °C), resuspended in denaturing loading buffer with 10 mM DTT, and analyzed by SDS-PAGE. Protein bands were quantified by densitometry on a LiCOR Odyssey CLx (700 nm) using Image Studio v5.2 software. Error bars represent standard deviation of co-floatations performed in parallel (n=2 or 3) and Student's t-test was used to determine statistical significance assuming two tailed distributions and equal variance.

4.5.4 *S. cerevisiae* Growth Assays

Δ coq9 or WT *S. cerevisiae* (BY4742) were transformed as previously described with p416GPD (Ura selection) plasmids encoding Coq9p variants and grown on Ura drop-out synthetic media plates containing glucose (2%, w/v).⁹⁶ Individual colonies were used to inoculate starter cultures, which were incubated (30 °C, ~18 h, 230 rpm). Serial dilutions of yeast (10^4 , 10^3 , 10^2 , or 10 yeast cells) were dropped onto Ura⁻ agar media plates containing either fermentation (2% glucose, w/v) or respiration (3% glycerol, w/v) media and incubated (30 °C, 2-4 days). To assay yeast growth in liquid media, yeast from a starter culture were swapped into selective synthetic media with in fermentation (2% glucose, w/v) or respiration (3% glycerol 0.1% glucose, w/v) media at an initial density of 5×10^6 cells/mL.⁹¹ The cultures were incubated in a sterile 96 well plate with an optical, breathable cover seal (1140 rpm) and optical density readings were obtained every 10 min. For co-overexpression experiments, yeast were co-transformed with p413GPD (His selection) plasmids encoding Coq8p-FLAG and subsequent plating and growth assays performed in Ura⁻His⁻ media. For small molecule bypass experiments, small molecules were added to 1 mM final concentration to the growth assay media: 2,4-dihydroxybenzoic acid (Coq7p bypass, Sigma D109401); 3,4-dihydroxybenzoic acid (Coq6p bypass, Sigma P5630); vanillic acid (Coq6p+Coq3p bypass, sigma H36001); 4-hydroxy-2-methylbenzoic acid (Coq5p bypass, Sigma 653160).

4.5.5 Crystallography

COQ9 was purified as described in “COQ9^{NA79} and COQ9^{NA79,CA31} expression and purification” and then subjected to size exclusion chromatography on HiLoad 16/600 Superdex 75 pg (GE

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

Healthcare) in Storage buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.3 mM TCEP). Fractions containing COQ9, assessed by SDS-PAGE, were pooled to >10 mg/mL, concentrated with a MW-cutoff spin filter, frozen in N₂(l), and stored at -80 °C.

COQ9^{NΔ79} was screened for crystallization using IndexHT (Hampton Research) and JCSG+ (Molecular Dimensions) screens and optimized using a TTP Labtech Mosquito crystallization robot using MRC SD-2 sitting drop plates. Crystals were imaged using a UVEX-P UV/Visible microscope. 200 nL protein at 4.89 mg/mL in 10 mM HEPES buffer at pH 7.5 and 100 mM NaCl was mixed with 200 nL crystallization reagent. Optimized crystals were obtained from a reservoir of 21% w/v PEG3350, 0.25 M NaCl, 0.1 M bistris, pH 6.5. The crystal was cryoprotected by soaking the crystal in reservoir solution supplemented to 35% w/v PEG3350. Diffraction data was collected at the Advanced Photon Source on GM/CA@APS 23-ID-B beamline, using an Eiger 16M detector. Data was collected at 12keV (1.0332Å) 350 mm 0.2 degrees per frame, 0.2 seconds per frame, 70x50 micron slits, 40x attenuation. Data was reduced using XDS, scaled with XSCALE.¹⁷¹ The structure was solved by molecular replacement with Phaser in the Phenix suite¹⁷² using PDB 4RHP as a search model.¹⁷³ The structure was refined using alternating rounds phenix.refine¹⁷⁴ and model building in Coot.¹⁷⁵

COQ9^{NΔ79,CA31} was screened for crystallization using the same workflow as COQ9^{NΔ79}. Reasonable crystals were obtained from ammonium sulfate and tris buffer. This crystallization condition was favored for isoprene co-crystallization trials, since polymeric precipitants like polyethylene glycol might be mistaken for isoprene alcohols. Initial optimization experiments using farnesol and geranylgeraniol in plastic crystallization plates yielded no obvious electron density for bound isoprenols. Optimal crystals were obtained by microseeding into droplets hanging from glass coverslips, hanging over glass concavity plates, sealed with perfluoroether oil. The reservoir solution was 50 μL of 1.2 M ammonium sulfate, 0.1 M Tris buffer, pH 8.5, mixed with 4 μL of neat geranylgeraniol, the majority of which formed a second phase. Two microliters of protein at 5.8 mg/mL, diluted from a 21.7 mg/mL protein stock in 160 mM NaCl, 20 mM HEPES buffer, pH 7.5 and 0.3 mM TCEP was mixed with 1 μL of seed stock derived from a previous batch crystallization experiment conducted at 1.5 M ammonium sulfate, 0.1 M Tris pH 8.5. Crystals grew over a week. They were cryoprotected by supplementing the droplet with 2 μL of reservoir solution, and then replacing it with 3 M ammonium sulfate, 0.1

M Tris pH 8.5, supplemented with neat geranylgeraniol. The droplet was allowed to equilibrate overnight, and crystals were harvested by drawing them through a layer of perfluoroether oil in a Mitegen microcrystal mount, wicking away excess mother liquor, and plunging then into liquid nitrogen. Data was collected at the Advanced Photon Source on LS-CAT 21-ID-D beamline, using a Eiger 9M detector. Data was collected at 100K, using an 11 keV (1.12723Å) beam, 170 mm, 0.2 degrees per frame, 50 micron beam. Data reduction, structure solution and refinement used the same workflow as COQ9^{NΔ79}, except that a version of 6AWL with the 10th helix truncated was used as the starting model for molecular replacement. Long tubes of electron density were observed at consistent sites on all of the monomers. Using constituents present in the crystallization mixture, the extent and shape of these features could only be successfully modeled using isoprenols. Features at these sites were less clear or absent in previous electron density maps from preceding crystallization attempts with isoprenols conducted in plastic 96 well plates. Isoprenols were docked into the electron density using phenix.ligandfit using geranylgeraniol, farnesol and geraniol as probes. Three farnesol (three isoprene units) and three geraniol (2 isoprene units) were successfully modeled into the most orderly sites.

4.5.6 Liquid Chromatography Mass Spectrometry Lipidomics

LC-MS samples contained either lipid extractions of *S. cerevisiae* pellets (either 1×10^8 cells or 50 mL of OD₆₀₀=3), *E. coli* pellets (1×10^{10} cells), His₈-MBP-tagged protein (40 nmols), or Coq9p-FLAG protein (normalized to an equal amount by gel quantitation) prepared as described elsewhere. LC-MS analysis was performed on an Acquity CSH C18 column held at 50 °C (100 mm × 2.1 mm × 1.7 μm particle size; Waters) using a Vanquish Binary Pump (400 μL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate in ACN/H₂O (70:30, v/v) containing 250 μL/L acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in IPA/ACN (90:10, v/v) with the same additives. Mobile phase B was initially held at 2% for 2 min and then increased to 30% over 3 min. Mobile phase B was further increased to 50% over 1 min and 85% over 14 min and then raised to 99% over 1 min and held for 7 min. The column was reequilibrated for 2 min before the next injection. Ten

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

microliters of sample were injected by a Vanquish Split Sampler HT autosampler (Thermo Scientific). The LC system was coupled to a Q Exactive mass spectrometer by a HESI II heated ESI source kept at 325 °C (Thermo Scientific). The inlet capillary was kept at 300 °C, sheath gas was set to 25 units, auxiliary gas to 10 units, and the spray voltage was set to 4,000/5,000 V. For targeted analysis the MS was operated in positive parallel reaction monitoring (PRM) mode and negative targeted single ion monitoring (t-SIM) acquiring scheduled, targeted scans to CoQ intermediates: PPAB, DMQ, CoQ. For discovery analysis, the MS was operated in polarity switching dd-MS2 mode. MS acquisition parameters were 17,500 resolving power, 3×10^6 automatic gain control (AGC) target for MS¹ and 5×10^5 AGC target for MS² scans, 100-ms MS¹ and 35-ms MS² ion accumulation time, 240- to 1,600-Th MS1 scan range, 1.4-Th isolation width for fragmentation, stepped HCD collision energy (20, 25/30 units), 1.0% under fill ratio, and 10 s dynamic exclusion.

For extended CoQ intermediate detection (Figure 4.13D) the LC-MS was performed as described above, with the following differences: Acquity CSH C18 column (150 mm \times 2.1 mm \times 1.7 μ m particle size); Vanquish Binary Pump (250 μ L/min flow rate); ACN/H₂O (90:10, v/v) in Mobile phases A and B; Mobile phase B increased from 65-99% over 9 min, held at 99% (2 min), then returned to 65% over 0.5 min, column re-equilibration (2.5 min); Q Exactive HF mass spectrometer; HESI II heated ESI source (300 °C); S-lens RF level to 50.0 and spray voltage set to 4,000 V for positive and 3,500 V for negative modes. MS was operated in positive and negative PRM mode, respectively, acquiring targeted scans to detect CoQ intermediates: 4-HP, PPHB, PPAB, 4-AP, HHAB, DDMQ, DMQ, IDMQ, CoQ₆, CoQ₁₀. MS acquisition parameters were 60,000 resolving power for MS² scans, 5×10^5 AGC, 200-ms MS² ion accumulation time, 2.2-Th isolation width for fragmentation, and stepped HCD collision energy of 15, 25, and 35 units.

Data Analysis

The resulting CoQ intermediate data were processed using TraceFinder 4.0 (Thermo Fisher Scientific). The resulting discovery lipidomics raw files were converted to mgf files via MSConvertGUI (ProteoWizard, Dr. Parag Mallick, Stanford University) and processed using Compound Discoverer 2.0 (Thermo Fisher Scientific) and an in-house developed software suite,

Lipidex.¹⁷⁶ All raw files were loaded into Compound Discoverer with blanks marked as such to generate two result files using the following Workflow Processing Nodes: Input Files, Select Spectra, Align Retention Times, Detect Unknown Compounds, Group Unknown Compounds, Fill Gaps and Mark Background Compounds for the so called “Aligned” result and solely Input Files, Select Spectra, and Detect Unknown Compounds for an “Unaligned” Result. Under Select Spectra, scan type was set to Full and unrecognized mass analyzer replacement to FTMS, the retention time limits were set between 0.4 and 21 min, MS order as well as unrecognized MS order replacements were set to MS1. Under Align Retention Times the mass tolerance was set to 10 ppm and the maximum shift according to the dataset to either 0.8 min. Under Detect Unknown Compounds, the mass tolerance was also set to 10 ppm, with an S/N threshold of 3, and a minimum peak intensity of 5E4. Further, [M+H]⁺+1 and [M-H]⁻-1 were selected as ions and a maximum peak width of 0.75 min as well as a minimum number of scans per peak equaling 5 were set. Lastly, for Group Unknown Compounds as well as Fill Gaps, mass tolerance was set to 10 ppm and retention time tolerance to 0.2 minutes. For best compound selection rules #1 and #2 were set to “unspecified,” while MS1 was selected for preferred MS order and [M+H]⁺+1 as the preferred ion. For everything else, the default settings were used. Resulting peak tables were exported as excel files in three levels of Compounds, Compound per File and Features (just Features for the “Unaligned”) and later saved as csvs. Normalization and enrichments calculations were performed and p-values were calculated via Student’s T Test assuming two-tailed distributions and equal variance.

4.5.7 Sample Preparation for *S. cerevisiae* Lipids Enriched in Coq9p-FLAG Pull-Downs

Coq9p-FLAG Expression and Immunoprecipitation

S. cerevisiae (BY4742) expressing p416GPD Coq9p-FLAG were grown in Ura⁻ synthetic complete media with 1% glucose (1L inoculated with 1×10⁸ cells in triplicate, 30 °C, 220 rpm.), harvested by centrifugation (4,000×g, 5 min, 25 °C) 4 hours post diauxic shift, frozen in N₂(l), and stored at -80 °C. Yeast pellets (1×10⁸ cells in triplicate) containing overexpressed protein were harvested in parallel.

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

Yeast spheroplasts were prepared as previously described.¹⁷⁷ Yeast pellets were resuspended in 35 mL of pretreatment buffer (0.1 M TrisSO₄, pH 9.4, 10 mM DTT), incubated (30 °C, 220 rpm, 15 min), and pelleted by centrifugation (1,500×g, 5 min). Pellets were resuspended in SP buffer (1.2 M Sorbitol, 20 mM KPi, pH 7.4), zymolyase (7.5 mg zymolyase per g pellet, MP Biomedicals 0832092) was added, and incubated (30 °C, 220 rpm, 40 min). Yeast spheroplasts were pelleted by centrifugation (4,500×g, 5 min, 4 °C), washed with 40 mL SP buffer, and pelleted by centrifugation (4,500×g, 5 min, 4 °C). Yeast spheroplasts were resuspended in 25 mL lysis buffer with protease inhibitors (10 mM TrisCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, 1 mM PMSF and peptide protease inhibitors: 500 ng/mL leupeptin hemisulfate, pepstatin A, chymostatin, aprotinin, and antipain dihydrochloride at pH 7.2) and lysed by sonication (3 pulses of 15 sec). Digitonin in lysis buffer (0.4% final digitonin by weight) was added and lysate was incubated (4 °C, 30 min, nutating) then clarified by centrifugation (15,000×g for 30 min at 4 °C). Clarified lysate was added to 300 µL FLAG-conjugated magnetic bead slurry (Sigma M8823) equilibrated with lysis buffer and incubated (4 °C, 2.5 hrs, end-over-end). FLAG beads were pelleted by centrifugation (700×g, 3 min, 4 °C), resuspended in lysis buffer (1 mL), washed 5X in lysis buffer with 0.4% digitonin, and washed 2X in lysis buffer without digitonin with magnetization. Protein was eluted into 1 mL of elution buffer (10 mM TrisCl pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, and 0.2 mg/mL FLAG peptide, Sigma F3290) with gentle agitation (1 hr, 25 °C).

Lipid Extraction

Immuno-purified Coq9p-FLAG (triplicates normalized to the lowest concentration by gel quantification) was thawed and transferred to 5 mL Eppendorf tubes. Buffer (10 mM TrisHCl pH 7.8, 140 mM NaCl) was added to 580 µL and vortexed (30 s). Extraction was initiated by CH₃Cl:MeOH (1:1, v/v, 3 mL) and vortexing (2×30 s). Samples were acidified to pH 0.5-2 (100 µL 1 M HCl) and mixed by vortexing (2×30 s). Phase separation was completed by centrifugation (3,220×g, 2 min, 4 °C), and the organic layer (1.5 mL) was transferred to a new 1.5 tubes and dried under Ar(g). The dried lipids were reconstituted in ACN/IPA/H₂O (65:35:5, v/v/v, 50 µL) by vortexing (2×30 s), transferred to a sample vial and stored under Ar(g) at -80 C. Frozen yeast pellets (1×10⁸ cells by OD₆₀₀) were extracted similarly with the following exceptions: glass

beads (0.5 mm diameter, 100 μ L) added upon thawing, 900 μ L of CH₃Cl:MeOH, 200 μ L 1 M HCl and 400 μ L of organic transferred.

4.5.8 Sample Preparation for *E. coli* Lipids Co-Purifying with His₈-MBP-COQ9^{NA79}

Sample Preparation

His₈-MBP-[TEV]-COQ9^{NA79}, His₈-MBP-[TEV]-COQ9^{NA79,W240K}, His₆-MBP proteins were expressed and purified as described in “COQ9^{NA79} and COQ9^{NA79, CA31} expression and purification” from 1L of *E. coli* in triplicate. Upon *E. coli* harvest, cell pellets (1×10¹⁰ cells by OD₆₀₀) were collected in triplicate. Upon protein collection, the elution was performed in a column, collected in 1 mL aliquots by gravity flow, and frozen in N₂(l), and stored at -80 °C for lipid extraction at the first IMAC elution.

Lipid Extraction

Cell pellets were thawed on ice, resuspended in 400 μ L PBS by vortexing (30 s). Purified protein (40 nmol by A₂₈₀) was brought to 456 μ L in 50 mM Tris pH 7.4, 200 mM NaCl, 100 mM imidazole. The aqueous samples were transferred to clean glass tubes and lipid extraction was initiated by addition of CH₃Cl:MeOH (1:1, v/v, 5 mL) and vortexing (30 s) to form a single phase. Samples were acidified to pH 0.5-2 (250 μ L 1 M HCl) and mixed by vortexing (1×30 s, 1×15 s). After inducing phase separation by addition of 500 μ L of brine, vortexing (15 s), and centrifugation (1,800×g, 4 min, 4 °C), the aqueous layer was aspirated, and organic layer dried under Ar(g). The dried lipids were reconstituted in ACN/IPA/H₂O (65:35:5, v/v/v, 100 μ L) by vortexing (2×30 s), transferred to a sample vial, and stored under Ar(g) at -80 C.

4.5.9 Sample Preparation for CoQ Intermediates in *S. cerevisiae*

Sample Preparation

(A) For PPAB, CoQ, DMQ detection (Figure 4.12E): *S. cerevisiae* (WT BY4742) expressing plasmids encoding Coq9p-FLAG variants in p416GPD were cultured as described previously.⁹¹ Yeast were grown in synthetic media (Ura⁻) respiration media (0.1% glucose, 3% glycerol w/v)

Chapter 4. Membrane, Lipid, and Protein Interactions of Coenzyme Q Biosynthesis

Protein COQ9

to early respiration growth. Specifically, 100 mL was inoculated with 2.5×10^6 cells, from 12 h overnight cultures grown in Ura⁻ 2% glucose, and incubated at 30 °C for 25 h at 230 rpm. Cells were grown and harvested in triplicate (1×10^8 cells by OD₆₀₀) by centrifugation (3,000 g, 3 min, 4 °C).¹⁷⁸ The supernatant was aspirated and pellets frozen in N₂(l) and stored at -80 °C. (B) For extended CoQ intermediate detection (Figure 4.13D): *S. cerevisiae* (WT BY4742) expressing plasmids encoding Coq9p-FLAG variants in p416GPD or Coq8p-FLAG in p413GPD were cultured as described previously.^{153,179} Yeast were grown in synthetic media (Ura⁻His⁻) fermentation media (2% galactose, 0.1% dextrose, 100 μM pABA). Specifically, 50 mL was inoculated with 6 mL pABA-starved yeast in stationary phase and incubated at 30 °C at 230 rpm until OD ~3. Cells were grown and harvested in triplicate (50 mL of OD₆₀₀=3 cells) by centrifugation (3,000 g, 3 min, 4 °C).

Lipid Extraction

(A) For PPAB, CoQ, DMQ detection (Figure 4.12E): Frozen yeast pellets (1×10^8 cells) were thawed on ice and mixed with glass beads (0.5 mm diameter, 100 μL). CHCl₃/MeOH (1:1, v/v, 4 °C, 900 μL) was added and samples were vortexed (2×30 s). HCl (1 M, 200 μL, 4 °C) was added and samples vortexed (2×30 s). The samples were centrifuged (5,000 g, 2 min, 4 °C) to complete phase separation. 400 μL of the organic phase was transferred to a clean tube and dried under Ar(g). The organic residue was reconstituted in ACN/IPA/H₂O (65:30:5, v/v/v) (100 μL) for LC–MS analysis. (B) For extended CoQ intermediate detection (Figure 4.13D): Yeast pellets were resuspended in water (100 μL) and submitted to a Soxhlet extraction with added internal standard (CoQ₁₀, 1 μM in final reconstitution) as described previously.^{153,179} Aqueous sample was submitted to two iterative extractions: vortexed (30s) with methanol (1 mL) and petroleum ether (2 mL). The organic phases were transferred to a clean tube and dried under Ar(g). The organic residue was reconstituted in ACN/IPA/H₂O (65:30:5, v/v/v) (200 μL) for LC–MS analysis.

4.5.10 LC-MS/MS Proteomics Analysis of *S. cerevisiae***Sample Preparation**

Frozen yeast pellets (1×10^8 cells) were re-suspended in 100 μ L guanidine-HCl (6M) and boiled for five minutes at 100 °C. Methanol was added to a concentration of 90% to precipitate protein and the sample was centrifuged for 10 min at 14,000 \times g. The resulting pellet was re-suspended in 100 μ L lysis buffer (8 M Urea, 100 mM Tris, 20 mM TCEP, 80 mM Chloroacetamide) and diluted with 1 mL 50 mM Tris. Protein digestion was performed overnight with trypsin (4 μ g) and centrifuged the following morning for 5 min at 14,000 \times g prior to de-salting resulting supernatant with Strata C18 solid phase extraction cartridges. Peptides were dried in a vacuum centrifuge before re-suspending in 0.2% formic acid and quantified using Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific).

LC-MS/MS Label-Free Proteomics

Samples were analyzed using an LC-MS instrument comprising an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) and an Ultimate 3000 RSLCnano chromatography system. Mobile phase A consisted of 0.2% formic acid in water and mobile phase B consisted of 0.2% formic acid in acetonitrile. A 75-min gradient ranging from 0% to 50% B was employed spanning a total runtime of 90 min. 2 μ g analytes were injected onto a 1.7 micron C18 column packed in-house to a length of 35 cm and heated to 55 °C. Survey scans of peptide precursors were collected from 350-1350 Th with an AGC target of 1,000,000 and a resolution of 60,000 in the Orbitrap followed by HCD MS/MS scans taken at top speed in the ion trap. The resulting LC-MS proteomics data were processed using Maxquant software version 1.5.2.8 and searched against a *S. cerevisiae* database downloaded from Uniprot on 8/10/16.^{180–182} The digestion enzyme was set to trypsin with up to two missing cleavages, carbamidomethylation of cysteine as a fixed modification, and oxidation of methionine and protein N-terminal acetylation as variable modifications. The match between runs feature was utilized to decrease missing data values within the data set. Precursor mass tolerance was 20 ppm and product ions were searched at 0.5 Da tolerances. Peptides were filtered to a 1% FDR and combined to protein groups based on the rules of parsimony.

4.5.11 Quantification and Statistical Analysis

See each individual method for the associated statistical analysis. The majority of p-values in this report were calculated using an unpaired, two-tailed, Student's t-test. MS data analysis was performed as described in each MS analysis section. In all cases, n represents independent replicates of an experiment and error bars represent standard deviation.

4.5.12 Experimental Data Availability and Accession Codes

The crystallography coordinates have been deposited in the Protein Data Bank under ID codes 6AWL and 6DEW. Raw image data has been deposited in Mendeley:

<http://dx.doi.org/10.17632/63ztbyhx5h.1>

Raw mass spectrometry files have been deposited in Chorus under accession numbers 1532 (proteomics) and 1533 (lipidomics).

5 Conclusions and Perspectives

The biological membrane provides a physical and regulatory barrier between the cell and its environment. It is a dynamic and complex system containing a variety of biological molecules, notably lipids, and proteins. Integral and peripheral membrane proteins account for more than one-third of the human proteome, and are estimated to represent the target for over 50% of modern drugs; however, there is a great disparity between our understanding of soluble and membrane proteins.

New and powerful techniques including cryo-electron microscopy and tomography, state-of-the-art lipidomics, mass spectrometry, membrane imaging by high-speed atomic force microscopy, are nowadays revolutionizing the way protein-membrane interplay is studied, but still cannot provide a detailed description of the molecular events that govern it. This has spearheaded the emergence of integrative modeling techniques that blend information from various experimental sources with theoretical and computational descriptions of the underlying biophysics into structural models.

Lipid metabolism and transport rely on proteins that operate at the membrane-water barrier and have dynamic interactions with membranes, lipids, and other proteins. In this thesis, we elaborate on the use of combined computational and experimental approaches to reveal structure-function relationships of membrane proteins and their close interplay with their respective membrane environments. How this integrative approach helps gaining robust

Chapter 5. Conclusions and Perspectives

insights into the function of membrane-associated protein systems is illustrated through the interrogation of two poorly characterized membrane-associated proteins, namely coenzyme Q (CoQ) biosynthesis proteins COQ8 and COQ9.

In mitochondria, CoQ precursors are likely buried between the leaflets of the inner mitochondrial membrane due to their extreme hydrophobicity. In Chapter 3 of this thesis, the computational analysis regarding COQ8 helped us discover that a conserved domain on COQ8 drives the electrostatic membrane association, where specific interactions with cardiolipins (the signature mitochondrial lipids) are observed. A diverse array of experimental input provided by experimental collaborators validated our computational findings, and demonstrated that COQ8 exhibits a conserved ATPase activity that is activated by COQ8 binding to cardiolipin-containing membranes. Our findings suggest that COQ8 might sense CoQ precursors within the membrane and then couple ATP hydrolysis to the partial extraction of these precursors into the aqueous matrix environment where they could be modified by CoQ biosynthesis enzymes.

The synthesis and transport of lipids in an aqueous cellular environment poses a long-standing biological question. In Chapter 4 of this thesis, through molecular simulations, we discovered that COQ9 uses a unique, amphipathic C-terminal helix connected to a foundational ligand-binding domain to peripherally associate with membranes, and we described the molecular mechanism of membrane interaction and cargo selection, which were confirmed by biochemical experiments. Our simulations also revealed a striking mechanism of how peripheral membrane proteins can lift lipids off the membrane, which helped answer the question of how proteins overcome the topological challenge of reaching substrates located inside the membrane. In this sense, our findings suggest that COQ9 acts as a bridge between two worlds: positioned at the membrane-water barrier to bring hydrophobic CoQ precursors close to the water to enable their interaction with CoQ pathway enzymes.

Overall, our results emphasize the importance of membrane composition, and the effects of protein interaction on membrane dynamics. Our work also highlights how the combination of experimental and computational techniques provides a unique and valuable perspective that neither of them could have provided alone. A full integration of these

techniques is the key to open the doors to a detailed understanding of protein-membrane interplay, which still remains elusive for many systems nowadays despite its pharmacological and biological relevance. The work presented in this thesis may therefore serve as a model for investigating other biosynthesis pathways, or other peripheral membrane proteins whose analyses involve similar challenges.

The most immediate follow-up study on this work is the structural determination of the COQ9:COQ7 complex, and visualization of the liposome-bound COQ proteins and protein complexes through cryo-EM studies. We already initiated an effort in this area through our experimental collaborators, and believe that these findings would pave the way for new hypotheses that would be investigated through molecular modeling and simulations. Another interesting idea that we would like to pursue is the effect of membrane curvature on the membrane binding and activity of the COQ proteins, as these proteins are located on the highly curved inner mitochondrial membrane. Membrane curvature can be induced by using different cardiolipin concentrations on the two leaflets of a finite (noncontinuous) membrane patch in coarse-grained MD simulations. The local curvature in the simulations can be quantified, and liposome co-floitation assays can be performed with matching curvatures, using liposomes of different sizes dictating their curvature.

These studies will also stimulate new work on approaching the remaining knowledge gaps on the biosynthesis of CoQ and the CoQ synthome. What is the three-dimensional structure of the CoQ synthome? Does it interact with any of the mitochondrial respiratory chain complexes? How are the hydrophobic CoQ and its intermediates transported within the membrane and also to other organelles? As the known functions of CoQ in human health and disease expand, filling these gaps through a full integration of experimental and computational techniques will help understand mitochondrial diseases, and will guide treatment strategies.

A Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

This Chapter presents a study related to the methods used in this thesis. The Chapter is formatted as the following manuscript to be submitted for publication:

"Using Energy Grids to Constrain the Search Space of Small Molecules During Docking."

Tamò G*, **Aydin D***, Träger S, Dal Peraro M. (*Contributed Equally)

Molecular docking is a foundational tool used in structure-based drug design to predict the preferred binding mode of a ligand in the binding site of a protein. Docking protocols generally use a scoring function to estimate the binding affinity of the ligand, which linearly combines quantitative features describing the binding pose. The weight constants assigned to balance the relative contribution of these features require careful and tedious computations, hindering the addition of new feature terms inside an already established scoring function. In this Chapter, we present a new docking procedure that makes the addition of new features possible and effortless by switching from an unconstrained to a constrained optimization protocol. To achieve this, the well-established AutoDock Vina scoring function is used as objective while energy grid-based parameters are introduced as inequality constraints to guide the assembly towards favorable areas of the optimization search space. The constrained optimization method is applied to flexible small molecule docking and achieves a success rate of 77% for the Astex diverse set containing 85 high-resolution protein-ligand structures.

A.1 Introduction

The field of computer-aided drug discovery has emerged over four decades ago with the ultimate goal of providing the most accurate predictions that would enable the cost- and time-efficient delivery of a new drug. Molecular docking is an indispensable computational tool that is now routinely used in structure-based drug design.¹⁸³ Given the three-dimensional structures of the unbound protein target and a database of compounds, molecular docking programs aim to predict the most probable bound conformation of a ligand in the receptor binding site. Practically, the molecular docking process can be decomposed into two steps: ligand pose generation by the sampling algorithm, and ranking of the poses using a scoring function.

To date, more than 60 molecular docking methods have been implemented, the strength and limitations of which have been extensively reviewed in the literature.^{184–188} Amongst the most popular molecular docking methods one can find AutoDock Vina,³⁷ which is currently one of the most cited open source docking methods in the scientific literature^{189,190} with over 7,000 citations as of March 2019. Importantly, AutoDock Vina has displayed great accuracy when tested in its ability to predict the bound conformation of ligands and their binding free energies. To achieve such performance, it relies on a carefully calibrated global scoring function composed of linearly added energy terms.³⁷ Coefficients in the form of weight constants are assigned to balance the contribution of these uncorrelated terms, which are obtained from regression analyses by iteratively fitting the scoring function to experimentally determined binding affinities.

One major limitation of such an approach is that, added to the costly calculation of these weights, the regression coefficients can be heavily dependent on the dataset used to compute them.¹⁹¹ More importantly, the addition of new terms to an already optimized scoring function is nearly impossible without rebalancing their contribution. This limitation comes from the fact that the docking problem is treated as an unconstrained single objective optimization.

In an attempt to optimize several scoring terms without balancing them, Gu et al.¹⁹¹ converted the docking problem to a multi-objective optimization where multiple scoring

functions were minimized simultaneously by a genetic algorithm. This approach was tested against the GOLD test set and showed remarkable docking accuracies compared to other popular molecular docking methods including GOLD,³⁵ Glide,³⁸ Surflex¹⁹² and Dock.¹⁹³

Another way to circumvent this problem would be to use a constrained optimization scheme, which aims at minimizing a single objective function while ensuring the satisfaction of pre-defined constraints.⁴⁵ Constrained optimization schemes have been applied in various fields including medicine,¹⁹⁴ optics,¹⁹⁵ engineering.¹⁹⁶ Recently, we implemented a new constrained optimizer named *mViE* (memetic viability evolution)⁴⁵ within our open source optimization framework named *pow^{er}* (parallel optimization workbench to enhance resolution, available at <http://lbm.epfl.ch/resources>).^{42,43} This framework was used in the *in silico* prediction of macromolecular assemblies and found to successfully dock symmetric and heterodimeric assemblies using experimental data to guide the assembly process.⁴⁶ The advantage of this optimization scheme is that it facilitates the addition of new constraints to guide and accelerate the optimization process, without the need of rebalancing the contribution of the different components.

In the same vein, we sought to design a constrained optimization method suitable for docking small molecules. To our knowledge, no constrained optimization application has yet been reported to predict the docking of small molecules. The reason behind this could be related either to the difficulty to find features to be used as constraints or to the general complexity of the optimization problem.

In this study, we extended our previously implemented constrained optimization protocol *pow^{er}-mViE* to support the docking of small molecules. We inherited the accurate scoring function of AutoDock Vina³⁷ as an objective function, and introduced inequality constraints in the form of receptor-ligand physicochemical properties, extracted from energy grids, which were used to guide the ligand docking. In order to test and validate the method, we used high-resolution receptor-ligand complexes of the PDBbind core set⁴⁸ to extract and calibrate the constraints, and the Astex diverse set⁴⁷ as the validation set. For rigid and flexible docking settings, our constrained molecular docking approach showed accuracies comparable to that of state-of-the-art software including the AutoDock Vina program. We envision that the use of

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

constraints extracted from energy grids would be critical for docking ligands more accurately when also the native dynamics of the target protein is considered during docking, for instance, as accounted for in molecular dynamics simulations.

A.2 Methods

Our constrained molecular docking protocol requires as input the atomistic structures of both the receptor and the ligand (Figure A.1). From these structures, atom types are assigned based on the X-Score nomenclature.^{37,39} Then, local and global constraints that assist the docking are computed from atom-type-based pre-computed energy grid maps, using the procedure outlined in this section. These local and global constraints are simultaneously used during optimization to guide the docking of the ligand into favorable regions of the receptor binding site. In order to quantify the binding free energy of the poses during docking, the scoring function inherited from AutoDock Vina is minimized. At the end of the optimization, the best receptor-ligand pose, i.e. the pose with the lowest energy which satisfies all the constraints, is returned.

A.2.1 The Constrained Optimization Strategy

The constrained optimization procedure that minimizes the energy function while satisfying the constraints was encoded within the *pow^{er}-mViE* framework (available at <http://lbm.epfl.ch/resources>)^{42,43,46} coupled with the constrained optimizer *mViE*.⁴⁵

Briefly, *mViE* advances a population of solutions based on (1+1)-CMAES strategy to locally explore the search space and recombines these local units using a differential evolution operator to perform a global search.⁴⁵ In this context, *mViE* attempts to find candidate solutions that satisfy pre-defined inequality constraints using the concept of viability evolution,^{44,45} which is an abstraction of artificial evolution and aims at selecting the promising candidate solution by adapting boundaries, termed “viability boundaries”, set around the inequality constraints. Initially relaxed around the constraints, these viability boundaries are gradually tightened during optimization. Only the solutions not violating these boundaries are termed “viable” and are selected for next iterations. This has the effect of gradually driving

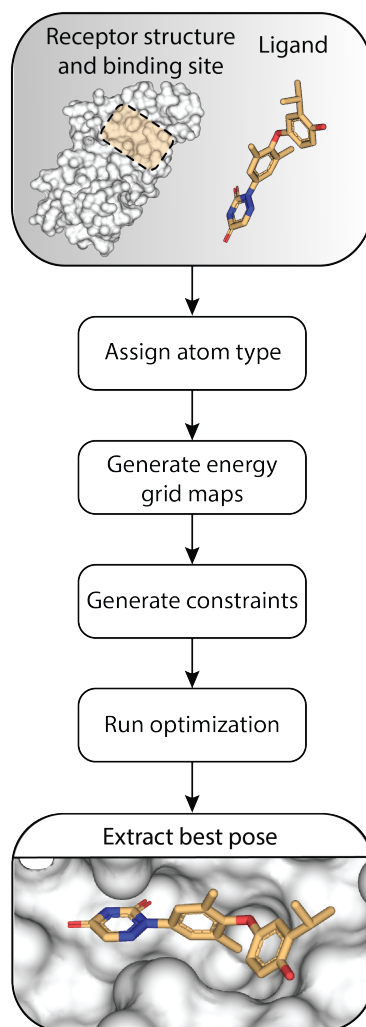


Figure A.1 – Constrained Ligand Docking Workflow. The *pow^{er}-mViE* molecular docking method takes as input the structures of the receptor and the ligand, and a docking box used to define a particular binding site. Atom types are assigned to both receptor and ligand structures based on the X-Score nomenclature using MGLTools (<http://mgltools.scripps.edu/>). Then, grid maps recapitulating the energy of interaction between each ligand atom type and receptor atoms are computed and constraints used to guide the ligand docking into favorable regions of the binding site are extracted. A constrained optimization is then run with the aim of minimizing an objective function estimating binding energy similar to that of AutoDock Vina, while ensuring the satisfaction of the constraints computed in the previous step. At the end of the optimization run, the receptor-ligand complex associated with the lowest estimated binding energy that satisfies all the constraints is returned.

the solutions towards “feasible” areas of the search space where the constraints are satisfied and the objective is minimal.⁴⁴ The minute workings of *mViE* can be found in the work of Maesani et al.⁴⁵ and thus will not be discussed further in this work.

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

A.2.2 General Formulation of the Docking Problem

In the context of molecular docking, *power-mViE* was used to sample the position, orientation and flexibility of the ligand encoded in a vector of design variables $X = [x, y, z, \alpha, \beta, \gamma, ens]$, where x , y and z correspond to the three translations, α , β and γ to the three Eulerian angles defining the ligand orientation and *ens* the ligand conformational ensemble generated by sampling torsional angles of the ligand rotatable bonds, in order to solve the docking problem generally formulated as:

$$\min f(X), \text{ s.t. } \begin{cases} l_i \leq X_i \leq u_i, i = 1, 2, \dots, n \\ g_j(X) \leq 0, j = 1, 2, \dots, m \end{cases} \quad (\text{A.1})$$

where $f(X)$ is the objective function to be minimized, l_i and u_i the lower and upper boundary ranges defining the search space of the variable X_i , and $g_j(X)$ the inequality constraints defined on each solution X .

A.2.3 The Objective Function

The objective function featured in AutoDock Vina³⁷ was integrated in the new *power-mViE* molecular docking method due its proven robustness, and detailed description. Given the atomistic pair-wise interaction term $d_{ij} = r_{ij} - R_{ti} - R_{tj}$, where r_{ij} is the measured Euclidian distance between the i^{th} and j^{th} atoms and R_t the van der Waals radius of these atoms of type t , the objective function $f(X)$ is defined as:

$$\text{AutoDockVina } f(X) = \begin{cases} w_1 \cdot \text{Gauss}_1 + \\ w_2 \cdot \text{Gauss}_2 + \\ w_3 \cdot \text{Repulsion} + \\ w_4 \cdot \text{Hydrophobic} + \\ w_5 \cdot \text{HydrogenBonds} \end{cases} \quad (\text{A.2})$$

where

$$\text{Gauss}_1(d) = e^{-\left(\frac{d}{0.5}\right)^2} \quad (\text{A.3})$$

$$Gauss_2(d) = e^{-\left(\frac{d-3}{2}\right)^2} \quad (A.4)$$

$$Repulsion(d) = \begin{cases} d^2, & \text{if } d < 0 \\ 0, & \text{if } d \geq 0 \end{cases} \quad (A.5)$$

$$Hydrophobic(d) = \begin{cases} 1, & \text{if } d < 0.5 \\ 1.5 - d, & \text{if } 0.5 < d < 1.5 \\ 0, & \text{if } d > 1.5 \end{cases} \quad (A.6)$$

$$HydrogenBonds(d) = \begin{cases} 1, & \text{if } d < -0.7 \\ -\frac{d}{-0.7}, & \text{if } -0.7 < d < 0 \\ 0, & \text{if } d > 0 \end{cases} \quad (A.7)$$

In the AutoDock Vina scoring function (eq. 2), the steric terms from eq. 3, eq. 4 and eq. 5 essentially reproduce the attractive and repulsive parts found in a standard Lennard-Jones potential and are applied to every atom irrespective of their type. Instead, the hydrophobic term depicted in eq. 6 is applicable only when both atoms involved in the interaction are hydrophobic. Similarly, in eq. 7, the hydrogen bonding term is only applied whenever one of the atoms participating in the interaction is a hydrogen bond acceptor and the other a hydrogen bond donor. The coefficients w_1 , w_2 , w_3 , w_4 , w_5 were calibrated respectively at -0.0356, -0.00516, 0.840, -0.0351 and -0.587 in the original study by Trott and Olson³⁷ to balance the contribution of the steric terms.

The choice for the AutoDock Vina energy function as the objective function implemented in *pow^{er}-mViE* was based on the fact it was open-source, well-documented and showed remarkable results when integrated into other molecular docking programs.^{191, 197, 198} Nevertheless, given the flexibility of the *pow^{er}-mViE* protocol, in principle any objective

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

function can be integrated or developed.

A.2.4 Constraint Terms

In the *pow^{er}-mViE* protocol, the information encoded in the physicochemical interactions between the receptor and the ligand is used as constraints to guide the docking procedure. Precisely, inequality constraints indicating the ideal ligand position and conformation inside the binding site are extracted as spatial distances from pre-computed grid maps of interaction energies for different atom types (Figure A.1).

Grid maps have been used previously to speed up the energy calculation by pre-computing the interaction energies between the ligand and the receptor.^{199–201} They have also been used to detect the location of the binding pocket.^{202,203} In this work, the pre-computed grid maps were not used to accelerate the receptor-ligand energy calculation nor to find potential binding sites. Instead, for the first time, they were used to derive a set of local and global constraints to guide the docking process.

Assuming the knowledge of the receptor binding site location, a cubic lattice is first generated at that location with evenly spaced voxels (Figure A.2, panel 1). Voxels found at a Euclidian distance too close to any receptor atom are removed, therefore leaving only non-overlapping voxels (Figure A.2, panel 1), from which two types of constraints are extracted: local and global (Figure A.2, panel 2).

From the atomistic structure of the ligand, atom types are defined based on the X-Score nomenclature³⁹ and include five atom types: hydrophobic (H), hydrogen bond acceptor (A), hydrogen bond donor (D), polar (P) and metal (M). Probes, each with the property of the ligand atom types, are iterated over each non-overlapping voxel, and the energy of interaction between the atom-type-probe and the receptor atoms is evaluated using eq. 2 (Figure A.2, panel a).

The non-overlapping voxels associated with the lowest interaction energies are extracted and then subjected to clustering. The resulting cluster centroids are then used in the form of local inequality constraints of the type $LocalAnchorDist_j \leq Ldist(j = 1, 2, \dots, m)$

where the *LocalAnchorDist* is the smallest Euclidian distance measured between one of the ligand atom and a centroid with a similar atom type, for all the atom types m of the ligand, and *Ldist* is a cutoff distance expressed in angstroms (Å) (Figure A.2, panel b). These local constraints assist the docking procedure by locally anchoring the ligand atoms to predicted grid points of the same atom type (Figure A.2, panel 3). Local constraints are considered satisfied whenever at least one ligand atom of each atom type is found below the cutoff distance *Ldist* of any anchoring centroid with the same atom type (Figure A.2, panel 3).

Global constraints are computed by combining the lowest energy voxels based on their 3D coordinates to form a cloud of grid points located inside the binding site (Figure A.2, panel c). The global constraint is used as a global anchor during docking through the introduction of spatial distances that constrain the ligand inside the binding site where the energies are minimal (Figure A.2, panel 3). Practically, this is done with the global inequality constraint $GlobalAnchorDist \leq Gdist$, where the *GlobalAnchorDist* is the average Euclidian distance computed from all ligand atoms and their respective nearest voxels located in the cloud of grid points, and *Gdist* the cutoff distance in angstroms (Å) below which the constraint is considered satisfied.

This leads us to the formal definition of the molecular docking problem as:

$$\min AutoDockVina f(X), \text{ s.t. } \begin{cases} l_i \leq X_i \leq u_i, i = 1, 2, \dots, n \\ LocalAnchorDist_j \leq Ldist, j = 1, 2, \dots, m \\ GlobalAnchorDist \leq Gdist \end{cases} \quad (A.8)$$

A.2.5 Training the Energy Grid Map Parameters To Obtain Optimal Constraints

The cutoff distance values associated with *Ldist* and *Gdist* were determinant for the quality of the local and global constraints as defined in eq. 8 because cutoff distances that are too high would not be specific enough to constrain the ligand in a desirable area of the binding site, and conversely cutoff values that are too low might be too specific and be satisfied only by a subset of receptor-ligand complexes. In turn, the values of *Ldist* were dependent on the parameter values related to the construction and extraction of grid voxels.

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

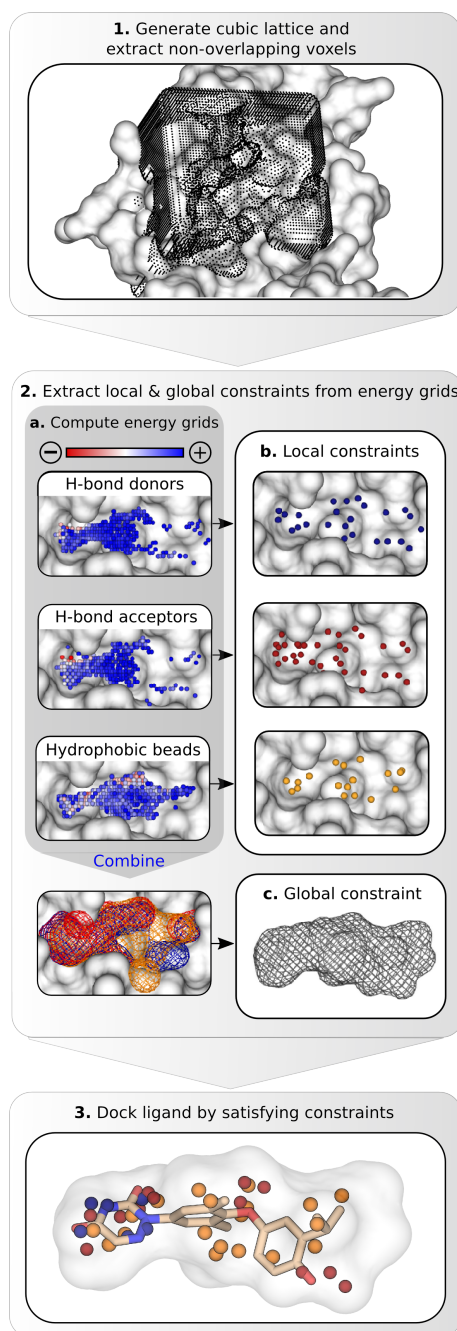


Figure A.2 – Constraint Generation from Energy Grid Maps. (1.) A cubic lattice is generated at the location of the receptor binding site. Voxels found too close to the receptor atoms are removed, leaving non-overlapping voxels in the lattice. (2.) For each atom type, the interaction energy between the lattice and the receptor atoms is computed by eq. A.2. Voxels are then ranked based on their interaction energy. The lowest energy voxels are extracted and used in two different ways (a-b). (b.) After clustering, the cluster centers are used as local constraints during docking by acting as docking anchors. (c.) Alternatively, the lowest-energy voxels for each atom type are combined based on their coordinates to create a spatial envelope used as the global constraint. (4.) The ligand is docked to the binding site by minimizing the objective function estimating the binding energy (eq. A.2) while simultaneously satisfying the global and local constraints. For the global constraint satisfaction, the ligand is spatially constrained into the cloud of voxels. Instead, for the local constraint, each ligand atom is spatially anchored to the voxel of the same atom type (same color in figure).

Thus, a comprehensive training benchmark was undertaken to find a set of grid parameter values that minimized *Ldist* and *Gdist* whilst ensuring they were satisfied in most receptor-ligand complexes. A simple combinatorial approach was performed to find the optimal grid parameter values, which were: the distance cutoff that defined an overlap between grid voxels and receptor atom, the number of low-energy voxels used to compute the constraints, and the distance separating each voxel of the grid map. Precisely, we combined values ranging from 2.6 to 3.3 Å for the distance cutoff defining an overlap between grid voxels and receptor atom, values ranging from 500 to 2000 voxels for the number of low-energy voxels recorded; and values ranging from 0.375 Å to 1.5 Å defining the voxel spacing^{37, 199, 200} of the grid map.

For this benchmark, the PDBbind core set⁴⁸ containing 195 high-resolution receptor-ligand complexes was used as the training set. For each combination of these three parameter values, a grid map was generated on bound receptor-ligand complexes, the value of *LocalAnchorDist* and *GlobalAnchorDist* were recorded and a rigid docking was performed to evaluate the quality of the constraints. The cloud of voxels defining the global constraint required a voxel spacing of 0.375 Å, a cutoff distance defining an overlap between voxel and receptor atoms of 2.8 Å and fixed number 500 low-energy voxels to be combined based on their 3D coordinates, for each atom type. In this respect, *Gdist* was calibrated at 1.5 Å. Conversely, the extraction of the centroid voxels defining the local constraints required a voxel spacing of 1.0 Å, a cutoff distance defining an overlap between voxel and receptor atoms of 3.0 Å and fixed number 600 low-energy voxels to be clustered, for each atom type. In this case, the cutoff distance *Ldist*, applied to all atom type voxels, was calibrated at 2.0 Å.

A.2.6 Step-by-Step Docking Protocol

As a first assessment and in effort to validate the improvement provided by the constraints, docking was performed on the training set (PDBbind core set⁴⁸) considering both the ligand and the receptor as rigid structures. In this case, the location of the binding site was known. Only the binding pocket residues (protein residues within 10 Å from the ligand center of mass) were used as the receptor instead of the whole protein to speed up the calculation. In order to simulate a realistic docking case, the geometric center of the binding site (represented by the

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

center of mass of the bound ligand) was randomized. The ligand was then removed and its orientation and position were randomized. For both the ligand and the receptor, atom types were assigned using the MGLTools (available at <http://mgltools.scripps.edu/>). The boundaries delimiting the ligand search space were represented as a cubic box of 23 Å side length, with a center corresponding to the randomized binding site geometric center. For comparative purposes, the number of function evaluations attributed to each receptor-ligand docking was consistent between docking methods and approximated the number function evaluation attributed to a standard rigid docking performed by AutoDock Vina, which was set to 120,000 function evaluations.

A second, more unbiased and exhaustive, round of assessment was undertaken on the Astex diverse set,⁴⁷ which contained 85 docking cases of high-resolution receptor-ligand complexes. As the grid parameters and constraints were not trained on this set, it served as the basis for an unbiased test. Moreover, unlike the previous assessment where both the ligand and the receptor were treated as rigid bodies, here the flexibility of the ligand was included in the search space in the form of a conformational ensemble. To account for the larger search space, the side length of the cubic box defining the search space and the number of function evaluations were increased to 25 Å and 200,000 respectively. Eventually for all optimization methods, the time required to compute a docking on a single-threaded execution (For AutoDock Vina: AutoDock Vina parameter `-cpu 1`) was recorded on a Xeon E3-1200 3.6 GHz processor.

A.2.7 Generation of the Ligand Conformational Ensemble

Here, the flexibility of the ligand was generated in isolation from the target binding site and added to the ligand search space during optimization. To address this task, we implemented a systematic rotatable bond sampler to generate ligand conformers.

The aim of the suggested implementation was to systematically generate all possible favorable energy conformations of the ligand and eventually add this ensemble of poses, treated as a conformational database, to the optimization search space. Specifically, for a given ligand, the number of active rotatable bonds was extracted using the MGLTools. Then,

discrete angle increments of 120° were performed for each rotatable bond. In order to filter out physically implausible ligand conformations, the internal ligand energy was computed on each atom pair separated by at least 3 covalent bonds (1-4 Lennard-Jones contributions) using a simple 12-6 Lennard-Jones potential, at each step of the permutation. In this case, ligand conformers with energies > 0.0 kcal/mol were not included in the conformational ensemble. The conformers were added as an extra dimension in the optimization space and used for the assessment of the Astex diverse set.

A.2.8 Metrics to Evaluate Accuracy

At end of the docking protocol, the accuracy of the docking was evaluated. For each receptor-ligand complex, the docked ligand pose associated with the lowest interaction energy (rank₁) was extracted. In the case of *pow^{er}-mViE*, an additional selection criterion was that the best pose should also satisfy all the pre-defined constraints of equation (8). The position and conformation of the rank₁ ligand pose were compared to that of the original crystal structure using the standard structural similarity metric $RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \delta_i^2}$, where averaging is done over the N pairs of equivalent heavy atoms and δ_i is the distance between the two atoms in the i -th pair. A docking was considered successful if the structure of the docked ligand was < 2.0 Å from the original crystal structure. Eventually, a success rate was issued by simply dividing the number of successful docking cases over all docking cases. This standard evaluation protocol has been used in several studies^{37, 187, 199, 204} in order to benchmark and validate newly developed protein-small molecule docking methods.

A.3 Results and Discussion

Given a receptor molecule that is held fixed in space, and a ligand that is moved around the receptor, our docking protocol *pow^{er}-mViE* attempts to minimize a scoring function describing the quality of candidate poses by sampling the rotation, orientation and conformation of the ligand, while simultaneously satisfying pre-defined local and global constraints describing the ideal position of the ligand inside the binding site (see Methods). Briefly, these local and global constraints aim to anchor and restrain the ligand to desirable regions of the search space, and

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

are extracted from atom-type-specific energy grid maps. The local constraints consist of voxel centroids extracted by clustering the lowest energy voxels according to their atom types, and the global constraint is obtained by combining the coordinates of the lowest energy voxels into one global cloud of grid voxels. Local constraints are considered satisfied whenever at least one ligand atom of each atom type is found below a cutoff distance $Ldist$ from any anchor of the same atom type. The satisfaction of global constraints is computed by making sure the average distance between all ligand atoms and their nearest cloud voxel is smaller than the global cutoff value $Gdist$.

A.3.1 Constrained Rigid and Flexible Docking

The optimal set of parameters was obtained through an exhaustive rigid docking benchmark on the PDBbind core set consisting in 195 receptor-ligand complexes where the docking accuracy (86.7%) was on par with the accuracy obtained with AutoDock Vina (84.6%). Noteworthy, the AutoDock Vina docking program was trained on receptor-ligand complexes from the PDBbind refined set.³⁷

The constrained molecular docking approach was further extended to flexibly dock ligands inside their respective binding sites. In this case, the receptor was still kept rigid while ligand flexibility was sampled. We tested this approach on the Astex diverse set,⁴⁷ which contains 85 high-resolution receptor-ligand complex structures. In the previously described rigid docking benchmark, a single randomly chosen binding site geometric center was used to train the grid parameter values and the constraint cutoff values $Ldist$ and $Gdist$. In order to ensure that the benchmark is not biased towards these specific geometric centers, and to assess how the location of the geometric center affected the outcome of the $pow^{\epsilon r}$ - $mViE$ molecular docking, 5 different sets of the 85 complexes were created, each with different binding site geometric centers. For each of the 5 sets, geometric centers were randomized from the ligand center of mass with continuous values within ± 2.0 Å in the x , y and z plane. For each of the 5 sets, 5 trials were performed to increase the statistics, giving a total of 25 x 85 docking trials. The $Ldist$, $Gdist$ and grid parameters were the same as those extracted from the rigid docking benchmark.

Ligand flexibility was encoded as a pre-generated ensemble of ligand conformers obtained through incrementing the torsional angles in a combinatorial approach (see Methods). This conformer ensemble was added as an extra search space dimension and contained the crystal conformation of the ligand. The accuracy computed from this flexible docking method was compared to the one obtained using AutoDock Vina in exactly the same conditions (i.e. geometric center location, starting from the bound ligand structure, equal number of function evaluations) and are shown in Table A.1.

Table A.1 – Flexible docking of the Astex diverse set and reported accuracies for *pow^{er}-mViE* and AutoDock Vina.

		<i>pow^{er}-mViE</i>		<i>AutoDock Vina</i>	
	Trial Number	Accuracy (%)	Mean Accuracy (%)	Accuracy (%)	Mean Accuracy (%)
Random Center 1	1	72.6	72.3 ± 3.0	73.8	71.7 ± 1.6
	2	66.7		71.4	
	3	73.8		70.2	
	4	73.8		72.6	
	5	70.2		70.2	
Random Center 2	1	69.0	71.2 ± 4.1	73.8	72.6 ± 1.2
	2	76.2		73.8	
	3	75.0		72.6	
	4	67.9		71.4	
	5	67.9		71.4	
Random Center 3	1	69.0	68.6 ± 2.0	71.4	71.9 ± 2.0
	2	66.7		71.4	
	3	71.4		73.8	
	4	66.7		69.0	
	5	69.0		73.8	
Random Center 4	1	77.4	72.4 ± 4.2	70.2	72.1 ± 1.4
	2	76.2		72.6	
	3	67.9		73.8	
	4	70.2		71.4	
	5	70.2		72.6	
Random Center 5	1	69.0	71.4 ± 1.9	69.0	70.5 ± 1.0
	2	71.4		70.2	
	3	70.2		70.2	
	4	73.8		71.4	
	5	72.6		71.4	
Overall Mean Accuracy (%)		71.0 ± 3.2		71.8 ± 1.5	
Overall Max Accuracy (%)		77.4		73.8	
Overall Min Accuracy (%)		66.7		69.0	

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

In this evaluation, *pow^{er}-mViE* and AutoDock Vina had similar accuracies (Table A.1). Importantly, the location of the geometric center used to compute the constraints did not influence the accuracy of the flexible docking (Table A.1) with an average accuracy estimated at $71.0\% \pm 3.2\%$ for all the 25 docking trials. These results were found comparable with those obtained using AutoDock Vina with respective geometric centers (mean accuracy: $71.8\% \pm 1.5\%$). Notably, the maximal accuracies obtained both for AutoDock Vina (73.8%) and *pow^{er}-mViE* (77.4%) flexible docking protocols (Table A.1) were comparable to the AutoDock Vina accuracies (76.5%) described in the study by Zoete et al.²⁰⁴ obtained with similar conditions. Both AutoDock Vina and *pow^{er}-mViE* show lower accuracies in flexible docking when compared to rigid docking. Docking accuracies from rigid-body docking methods are known to be higher compared to flexible docking ones,¹⁸⁶ since rigid docking problems have a smaller search space with only six degrees of freedom consisting of three translations and three rotations.

The Astex diverse set, used to validate the method, includes 5 receptor-ligand complexes (i.e., 1GPK, 1HNN, 1N1M, 1N2V and 1OYT) that are also present in the PDBbind core set, used to train the grid parameters. Removing these 5 docking cases when computing the overall accuracy did not lead to a different overall mean accuracy ($70.2\% \pm 3.2\%$).

The results presented herein have to be reflected with care since the strategy to sample ligand flexibility was different between AutoDock Vina and *pow^{er}-mViE*. Precisely, AutoDock Vina samples the angles of each ligand torsion during optimization, while *pow^{er}-mViE* assigns a structure from a pre-computed conformer ensemble by systematically incrementing each torsional angle by 120° .³⁷ The strategy to sample from a pre-computed set of ligand conformers through a systematic search is a conceptually simpler method compared to sampling each of the ligand torsions.^{184, 185} An alternative strategy includes incrementally growing the ligand into the binding site.²⁰⁵ This method in particular was applied by the Dock³⁴ and FlexX²⁰⁶ programs and involves iteratively docking ligand fragments into the binding site then covalently linking them. Developing a similar sampling strategy for the next molecular docking implementation of *pow^{er}-mViE* would certainly increase the robustness of the approach by alleviating the conformer search space.

A.3.2 Evaluation of Successes and Failures

For each of the 84 flexible docking cases featured in the Astex diverse set, the number of successful trials (i.e. where the best ligand pose is found at RMSD < 2.0 Å of the bound ligand) was computed over the 25 trials. In 44 receptor-ligand docking cases, both AutoDock Vina and *pow^{er}-mViE* consistently found the native ligand binding mode with success rates $\geq 76\%$ (20/25 trials).

Conversely, in 10 docking cases which comprises complexes 1GM8, 1HVV, 1JD0, 1JJE, 1Q41, 1R58, 1SQ5, 1TZ8, 1W1P and 1UVE, neither AutoDock Vina nor *pow^{er}-mViE* could consistently find the ligand binding mode as found in the native receptor-ligand complex (success rates $\leq 20\%$, 5/25 trials). Upon visual inspection of the receptor-ligand complex structures and searching through the literature, it appeared that some of these cases were in fact “usual suspects” previously labeled as difficult docking cases.^{199,204} The reason behind such difficulty lies in the presence of explicit water molecules mediating critical receptor-ligand hydrogen bonding interactions in the native complex structure, which are removed prior to docking. This was attributed as the root cause for the docking failures of 1GM8, 1HVV, 1JD0, 1SQ5 and 1W1P. Another reason for failure is due to the ligand interaction with metal ions as in complexes 1JD0 or 1JJE. The reason for the failure of other docking cases could not be detected through the visual inspection of the complex crystal structure and thus was attributed to limitations in the scoring function accuracy.

More interesting docking cases consisted of those where AutoDock Vina and *pow^{er}-mViE* showed conflicting success rates. For complexes 1MMV and 1UML in particular (Figure A.3A), AutoDock Vina succeeded with a rate of 100% in 25 trials, while *pow^{er}-mViE* showed a success rate of $\leq 20\%$ (5/25 trials). Visual inspection of the native complex structures did not reveal the presence of water molecules or metal ions mediating the interactions between the ligand and receptors. In order to check whether the local and global constraints may have wrongly anchored the ligand into undesirable areas of the binding site, we computed the local and global constraint satisfaction as well as the energy score, according to eq. 2, for the best (lowest estimated binding energy) receptor-ligand poses returned by AutoDock and *pow^{er}-mViE*. Such analysis showed that both the correctly docked lowest-energy poses computed

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

by AutoDock Vina and the wrongly docked best poses of *pow^{er}-mViE* were found to satisfy the local and global constraints. However, the lowest energies recorded for the best poses of 1MMV (-10.9 kcal/mol) and 1UML (-16.3 kcal/mol) returned by AutoDock Vina were lower than the ones computed on the crystal structures of 1MMV (-9.6 kcal/mol) and 1UML (-14.9 kcal/mol) respectively, which in turn were lower than the best pose computed by *pow^{er}-mViE* (1MMV: -8.67 kcal/mol, 1UML: -11.7 kcal/mol). This showed in these two cases that while the constraints were satisfied in the best poses returned by both *pow^{er}-mViE* and AutoDock Vina, failure to compute better poses by *pow^{er}-mViE* was probably due to the fact that it was unable to overcome a local energy minimum. A possible solution to better find the correct binding pose could be either to increase the number of function evaluations for these two cases or refine the constraints further so as to more precisely anchor the ligand into the binding site.

Conversely, *pow^{er}-mViE* correctly docked the ligand of complexes 1MEH, 1N2V, 1R55 with success rates of $\geq 84\%$ (21/25 trials) while AutoDock Vina consistently failed to find the native binding mode (0% success rate) for these complexes. Interestingly, visual inspection of these complexes revealed that both 1MEH and 1N2V contained water molecules, and 1R55 a Zn ion bridging the interaction between the ligand and the receptor. These have been previously noted as being particularly difficult docking cases.^{199,204} Comparisons between the estimated binding energy associated with the *pow^{er}-mViE* best poses (1MEH: -10.2 kcal/mol, 1N2V: -8.4 kcal/mol, 1R55: -9.3 kcal/mol), the energies from the best AutoDock Vina poses (1MEH: -10.4 kcal/mol, 1N2V: -8.6 kcal/mol, 1R55: -10.2 kcal/mol) and crystal structures (1MEH: -9.7 kcal/mol, 1N2V: -8.0 kcal/mol, 1R55: -8.0 kcal/mol) indicated that both *pow^{er}-mViE* and AutoDock Vina found lower energies than that of the complex structure. Moreover, the local and global constraints were satisfied in the best poses returned by both *pow^{er}-mViE* and AutoDock Vina, and in the crystal structures. Thus, we suggest that the reason for the success of *pow^{er}-mViE* in this case might be due to a combined effect of constraints anchoring the ligand to a suitable region of the binding mode and to the coarse angular step of 120° which might have prevented the ligand to adopt a conformation leading to a non-native binding mode.

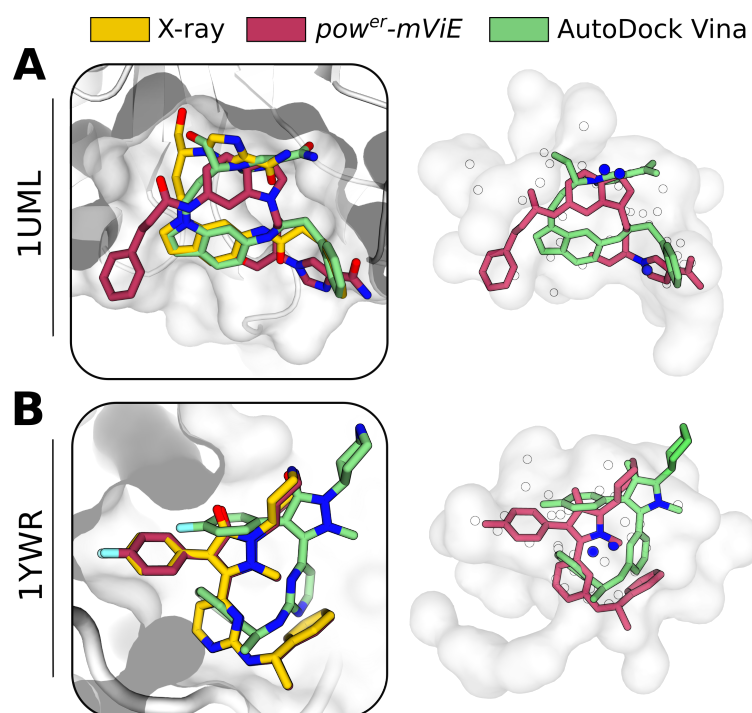


Figure A.3 – Docking Success and Failures of pow^{er} - $mViE$ and AutoDock Vina. (A) Comparison of the crystal structure and the best ligand poses returned by the pow^{er} - $mViE$ and AutoDock Vina for 1UML (left panel) and associated local and global constraints satisfaction (right panel). (B) Comparison of the crystal structure and the best ligand poses returned by the pow^{er} - $mViE$ and AutoDock Vina for 1YWR (left panel) and associated local and global constraints satisfaction (right panel). For the representation of constraint satisfaction for both 1UML and 1YWR (right panels), the white cloud surface represents the global constraint, the full blue circles represent the satisfied local polar centroids ($Ldist < 2 \text{ \AA}$), which are located close to polar ligand atoms also colored in blue, and the empty white circles represent unsatisfied local polar centroids.

The docking cases of 1YWR (pow^{er} - $mViE$: 72%, AutoDock Vina: 0%) and 1G9V (pow^{er} - $mViE$: 52%, AutoDock Vina: 0%) were found to also be worth mentioning since they illustrated how the use of constraints helped find the correct binding mode of the ligand. Visual inspection of the 1YWR crystal structure did not reveal the presence of water molecules or metal ions inside the binding site. Evaluating the level of constraint satisfaction on the best poses returned by AutoDock Vina over the 25 trials showed that a local constraint specifying the minimum distance between a polar centroid voxel and a polar atom of the ligand was always $> 2.0 \text{ \AA}$, which is above the $Ldist$ cutoff value, and thus violates that local constraint (Figure A.3B). Unlike the best AutoDock Vina poses, both the crystal structure 1YWR and the best pow^{er} - $mViE$ poses were found to satisfy all the local and global constraints. Additionally, the estimated binding energy associated with the best pose returned by pow^{er} - $mViE$ (-13.7

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

kcal/mol) was lower than the ones of the crystal structure (-12.4 kcal/mol), which was in turn lower than the best AutoDock Vina pose (-12.0 kcal/mol). This remarkable example served as a showcase illustrating how the use of constraints by *pow^{er}-mViE* was useful in anchoring the ligand in favorable regions of the binding site where the minimal binding energy was found, leading to a pose resembling that of the native binding mode.

Anecdotic cases like those of 1MEH, 1N2V, 1R55, and particularly 1YWR put forward the usefulness of using pre-computed constraints to assist the docking of ligands inside their binding site. Noteworthy, 1MEH and 1N2V featured water-mediated ligand-receptor interactions and 1R55 contained Zn ions in the binding site. Such cases are considered difficult to solve. Nevertheless, similar water- and metal ion-mediated docking cases such as 1GM8, 1HVY, 1JD0, 1JJE, 1Q41, 1R58, 1SQ5, 1TZ8, 1W1P and 1UVF still remain elusive to most small molecule docking algorithms including *pow^{er}-mViE* and AutoDock Vina. If present in the receptor crystal structure, a suggested strategy to improve the docking performance is to keep and optimize the orientation and position of the water hydrogen atoms, using specific energy minimizations.^{207,208} If not present in the receptor crystal structure, water molecules can be added to the binding site using grid-based or molecular dynamics simulations and can be selected based on their energetic stability for docking.²⁰⁹ The presence of metal ions can hinder the performance of molecular docking programs because it may affect the hydration and protonation states of charged residues,²¹⁰ which are factors still difficult to predict using available methods as they often require more expensive calculations at the quantum level.²¹¹

A.4 Conclusions

In the work presented here, we applied the principles of constrained optimization to assist the docking of small molecules inside the binding site of their respective targets. Precisely, local and global constraints were extracted from pre-computed energy grid maps at the location of the binding site. These constraints were essentially used during docking to anchor ligands inside desirable binding site areas by constraining their global position inside low-energy areas, and by locally matching ligand atoms to areas favoring their associated atom types.

Using *pow^{er}-mViE* on a constrained rigid docking setting, we combined these local

and global constraints together with the AutoDock Vina scoring function and witnessed an almost fourfold improvement in docking accuracy compared to a rigid docking setting where an unconstrained optimization was performed. Moreover, in a flexible docking setting, we obtained accuracies comparable to state-of-the-art small molecule docking methods.

Despite the encouraging results obtained in this work, we suggest necessary improvements for the future implementation of the method. To increase the robustness of our ligand conformer sampling approach, we suggest a sampling method similar to Dock³⁴ and FlexX²⁰⁶ programs, which consists in first dividing the ligand into fragments and then individually docking the fragments by respecting their covalent bonding.

Similar to other small molecule docking methods,^{199,204} we faced difficulties in docking cases where water or metal ions were mediating the interaction between the ligand and the receptor. A suggested approach to improve docking accuracies when water mediates ligand-receptor interactions could be to first to refine the hydrogen atom positions of the water molecules found in the complex crystal structure, then to use these waters as a part of the receptor structure during docking.^{207,208} When not present in the complex crystal structure, water molecules can be added using molecular dynamics-based techniques, which assess their relative positional retention inside the binding site.²⁰⁹ As for docking cases where protein-ligand interactions are mediated through a metal ion, more refined techniques are suggested which include calculation at the quantum level.²¹¹

Finally, we envision this approach to be advantageous in more difficult, but more realistic docking cases such as those where both the ligand and receptor flexibility are sampled during optimization. Instead of using a single receptor conformation, the inclusion of receptor flexibility in small molecule docking holds promise for increasing docking accuracy and capturing a better physical representation of the ligand-receptor interactions.^{190,212} In the best-case scenario, receptor flexibility can be accounted for by including different experimentally solved structures of the protein target in the docking process, determined for instance via X-ray crystallography or NMR. However in more realistic docking cases where only a single conformation of the receptor is available, receptor flexibility can be sampled using *in silico* methods. A popular way to treat receptor flexibility using *in silico* methods consists in using

Appendix A. Using Energy Grids to Constrain the Search Space of Small Molecules During Docking

amino-acid side-chain rotamer libraries,²⁰⁵ which locally samples the binding site flexibility based on experimentally observed amino-acid side-chain conformations. This approach however has limited use in docking cases where the binding site is expected to undergo substantial backbone conformational change in order to accommodate the ligand,¹⁸⁴ e.g. in loopforming active sites²¹³ or cryptic pockets.²¹⁴ For such flexible cases, more expensive techniques, including molecular dynamics simulations, can be used on the receptor in its apo form to generate different snapshots of binding site conformations. This ensemble of binding site conformations can be used directly during docking, as in ICM (4D docking)^{33,215} and is expected to provide not only significant enrichment in virtual screening but also more chemically diverse hits.

Bibliography

- ¹ G. van Meer, D. R. Voelker, and G. W. Feigenson. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Bio*, 9(2):112–124, February 2008.
- ² A. S. B. Olsen and N. J. Faergeman. Sphingolipids: membrane microdomains in brain development, function and neurological diseases. *Open Biol*, 7(5), 2017.
- ³ E. J. Dufourc. Sterols and membrane dynamics. *J Chem Biol*, 1(1-4):63–77, November 2008.
- ⁴ A. Doerr. Membrane protein structures. *Nature Methods*, 6:35, December 2008.
- ⁵ J. P. Overington, B. Al-Lazikani, and A. L. Hopkins. How many drug targets are there? *Nature Reviews Drug Discovery*, 5(12):993, December 2006.
- ⁶ V. Monje-Galvan and J. B. Klauda. Peripheral membrane proteins: Tying the knot between experiment and computation. *Biochim. Biophys. Acta*, 1858(7 Pt B):1584–1593, July 2016.
- ⁷ S. Vanni, L. Vamparys, R. Gautier, G. Drin, C. Etchebest, P F J. Fuchs, and B. Antonny. Amphipathic lipid packing sensor motifs: probing bilayer defects with hydrophobic residues. *Biophys. J.*, 104(3):575–584, February 2013.
- ⁸ T. B. Blum, A. Hahn, T. Meier, K. M. Davies, and W. Kühlbrandt. Dimers of mitochondrial ATP synthase induce membrane curvature and self-assemble into rows. *PNAS*, 116(10):4250–4255, March 2019.
- ⁹ N. A. Brazhe, A. B. Evlyukhin, E. A. Goodilin, A. A. Semenova, S. M. Novikov, S. I. Bozhevolnyi, B. N. Chichkov, A. S. Sarycheva, A. A. Baizhumanov, E. I. Nikelshparg, L. I. Deev, E. G. Maksi-

Bibliography

- mov, G. V. Maksimov, and O. Sosnovtseva. Probing cytochrome c in living mitochondria with surface-enhanced Raman spectroscopy. *Sci Rep*, 5:13793, September 2015.
- ¹⁰ C. D. M. Sperka-Gottlieb, A. Hermetter, F. Paltauf, and G. Daum. Lipid topology and physical properties of the outer mitochondrial membrane of the yeast, *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 946(2):227–234, December 1988.
- ¹¹ J. Dudek. Role of Cardiolipin in Mitochondrial Signaling Pathways. *Front Cell Dev Biol*, 5:90, 2017.
- ¹² M. Zhang, E. Mileykovskaya, and W. Dowhan. Cardiolipin is essential for organization of complexes III and IV into a supercomplex in intact yeast mitochondria. *J Biol Chem*, 280(33):29403–8, August 2005.
- ¹³ C. Arnarez, S. J. Marrink, and X. Periole. Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes. *Chem Sci*, 7(7):4435–4443, July 2016.
- ¹⁴ A. Y. Hsu, T. Q. Do, P. T. Lee, and C. F. Clarke. Genetic evidence for a multi-subunit complex in the O-methyltransferase steps of coenzyme Q biosynthesis. *Biochim. Biophys. Acta*, 1484(2-3):287–297, April 2000.
- ¹⁵ D. C. Lohman, F. Forouhar, E. T. Beebe, M. S. Stefely, C. E. Minogue, A. Ulbrich, J. A. Stefely, S. Sukumar, M. Luna-Sanchez, A. Jochem, S. Lew, J. Seetharaman, R. Xiao, H. Wang, M. S. Westphall, R. L. Wrobel, J. K. Everett, J. C. Mitchell, L. C. Lopez, J. J. Coon, L. Tong, and D. J. Pagliarini. Mitochondrial COQ9 is a lipid-binding protein that associates with COQ7 to enable coenzyme Q biosynthesis. *Proc Natl Acad Sci U S A*, 111(44):E4697–705, November 2014.
- ¹⁶ J. A. Stefely and D. J. Pagliarini. Biochemistry of Mitochondrial Coenzyme Q Biosynthesis. *Trends Biochem Sci*, 42(10):824–843, October 2017.
- ¹⁷ A. M. James, H. M. Cochemé, M. Murai, H. Miyoshi, and M. P. Murphy. Complementation of coenzyme Q-deficient yeast by coenzyme Q analogues requires the isoprenoid side chain. *FEBS J.*, 277(9):2067–2082, May 2010.
- ¹⁸ Y.-N. Dai, K. Zhou, D.-D. Cao, Y.-L. Jiang, F. Meng, C.-B. Chi, Y.-M. Ren, Y. Chen, and C.-Z. Zhou. 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. D Biol. Crystallogr.*, 70(Pt 8):2085–2092, August 2014.
- ¹⁹ J. A. Stefely, A. G. Reidenbach, A. Ulbrich, K. Oruganty, B. J. Floyd, A. Jochem, J. M. Saunders, I. E. Johnson, C. E. Minogue, R. L. Wrobel, G. E. Barber, D. Lee, S. Li, N. Kannan, J. J. Coon, C. A. Bingman, and D. J. Pagliarini. Mitochondrial ADCK3 employs an atypical protein kinase-like fold to enable coenzyme Q biosynthesis. *Mol Cell*, 57(1):83–94, January 2015.
- ²⁰ J. A. Stefely, F. Licitra, L. Laredj, A. G. Reidenbach, Z. A. Kemmerer, A. Grangeray, T. Jaeg-Ehret, C. E. Minogue, A. Ulbrich, P. D. Hutchins, E. M. Wilkerson, Z. Ruan, D. Aydin, A. S. Hebert, X. Guo, E. C. Freiburger, L. Reutenauer, A. Jochem, M. Chergova, I. E. Johnson, D. C. Lohman, M. J. Rush, N. W. Kwiecien, P. K. Singh, A. I. Schlagowski, B. J. Floyd, U. Forsman, P. J. Sindelar, M. S. Westphall, F. Pierrel, J. Zoll, M. Dal Peraro, N. Kannan, C. A. Bingman, J. J. Coon, P. Isope, H. Puccio, and D. J. Pagliarini. Cerebellar Ataxia and Coenzyme Q Deficiency through Loss of Unorthodox Kinase Activity. *Mol Cell*, 63(4):608–20, August 2016.
- ²¹ D. C. Lohman, D. Aydin, H. C. Von Bank, R. W. Smith, V. Linke, E. Weisenhorn, M. T. McDevitt, P. Hutchins, E. M. Wilkerson, B. Wancewicz, J. Russell, M. S. Stefely, E. T. Beebe, A. Jochem, J. J. Coon, C. A. Bingman, M. Dal Peraro, and D. J. Pagliarini. An Isoprene Lipid-Binding Protein Promotes Eukaryotic Coenzyme Q Biosynthesis. *Mol. Cell*, 73(4):763–774.e10, February 2019.
- ²² S. Jo, T. Kim, V. G. Iyer, and W. Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*, 29(11):1859–1865, August 2008.
- ²³ C. Bovigny, G. Tamò, T. Lemmin, N. Maïno, and M. Dal Peraro. LipidBuilder: A Framework To Build Realistic Models for Biological Membranes. *J Chem Inf Model*, 55(12):2491–2499, December 2015.
- ²⁴ S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B*, 111(27):7812–7824, July 2007.
- ²⁵ C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, and R. C. Walker. Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput*, 10(2):865–879, February 2014.

Bibliography

- ²⁶ S. E. Feller and A. D. MacKerell. An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. *J. Phys. Chem. B*, 104(31):7510–7515, August 2000.
- ²⁷ C. Sagui and T. A. Darden. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu Rev Biophys Biomol Struct*, 28:155–179, 1999.
- ²⁸ T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, June 1993.
- ²⁹ H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation*, 6(1-3):121–142, March 1991.
- ³⁰ H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci*, 4(3):225–248, May 2014.
- ³¹ L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput*, 4(5):819–834, May 2008.
- ³² S. J. Marrink and D. P. Tieleman. Perspective on the Martini model. *Chem Soc Rev*, 42(16):6801–6822, August 2013.
- ³³ R. Abagyan, M. Totrov, and D. Kuznetsov. Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. *J Comput Chem*, 15(5):488–506, May 1994.
- ³⁴ T. J. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, 15(5):411–428, May 2001.
- ³⁵ G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267(3):727–748, April 1997.

- ³⁶ A. Grosdidier, V. Zoete, and O. Michielin. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.*, 39(Web Server issue):W270–277, July 2011.
- ³⁷ O. Trott and A. J. Olson. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, 31(2):455–461, January 2010.
- ³⁸ R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 47(7):1739–1749, March 2004.
- ³⁹ R. Wang, L. Lai, and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, 16(1):11–26, January 2002.
- ⁴⁰ G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, 19(14):1639–1662, November 1998.
- ⁴¹ V. Namasivayam and R. Günther. pso@autodock: a fast flexible molecular docking program based on Swarm intelligence. *Chem Biol Drug Des*, 70(6):475–484, December 2007.
- ⁴² M. T. Degiacomi, I. Iacovache, L. Pernot, M. Chami, M. Kudryashev, H. Stahlberg, F. G. van der Goot, and M. Dal Peraro. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat. Chem. Biol.*, 9(10):623–629, October 2013.
- ⁴³ M. T. Degiacomi and M. Dal Peraro. Macromolecular symmetric assembly prediction using swarm intelligence dynamic modeling. *Structure*, 21(7):1097–1106, July 2013.
- ⁴⁴ A. Maesani, P. R. Fernando, and D. Floreano. Artificial evolution by viability rather than competition. *PLoS ONE*, 9(1):e86831, 2014.
- ⁴⁵ A. Maesani, G. Iacca, and D. Floreano. Memetic Viability Evolution for Constrained Optimization. *IEEE Transactions on Evolutionary Computation*, 20:125–144, 2016.

Bibliography

- ⁴⁶ G. Tamò, A. Maesani, S. Träger, M. T. Degiacomi, D. Floreano, and M. Dal Peraro. Disentangling constraints using viability evolution principles in integrative modeling of macromolecular assemblies. *Sci Rep*, 7(1):235, 2017.
- ⁴⁷ M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortenson, and C. W. Murray. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.*, 50(4):726–741, February 2007.
- ⁴⁸ R. Wang, X. Fang, Y. Lu, and S. Wang. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.*, 47(12):2977–2980, June 2004.
- ⁴⁹ C. J. Leonard, L. Aravind, and E. V. Koonin. Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. *Genome Res.*, 8(10):1038–1047, October 1998.
- ⁵⁰ N. Kannan, S. S. Taylor, Y. Zhai, J. C. Venter, and G. Manning. Structural and functional diversity of the microbial kinome. *PLoS Biol.*, 5(3):e17, March 2007.
- ⁵¹ W. W. Poon, D. E. Davis, H. T. Ha, T. Jonassen, P. N. Rather, and C. F. Clarke. Identification of *Escherichia coli* ubiB, a gene required for the first monooxygenase step in ubiquinone biosynthesis. *J. Bacteriol.*, 182(18):5139–5146, September 2000.
- ⁵² L. X. Xie, E. J. Hsieh, S. Watanabe, C. M. Allan, J. Y. Chen, U. C. Tran, and C. F. Clarke. Expression of the human atypical kinase ADCK3 rescues coenzyme Q biosynthesis and phosphorylation of Coq polypeptides in yeast coq8 mutants. *Biochim. Biophys. Acta*, 1811(5):348–360, May 2011.
- ⁵³ J. Zheng, E. A. Trafny, D. R. Knighton, N. H. Xuong, S. S. Taylor, L. F. Ten Eyck, and J. M. Sowadski. 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr. D Biol. Crystallogr.*, 49(Pt 3):362–365, May 1993.
- ⁵⁴ A. S. Khadria, B. K. Mueller, J. A. Stefely, C. H. Tan, D. J. Pagliarini, and A. Senes. A Gly-zipper motif mediates homodimerization of the transmembrane domain of the mitochondrial kinase ADCK3. *J. Am. Chem. Soc.*, 136(40):14068–14077, October 2014.

- ⁵⁵ S. Kim, T.-J. Jeon, A. Oberai, D. Yang, J. J. Schmidt, and J. U. Bowie. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 102(40):14278–14283, October 2005.
- ⁵⁶ A. Ferrer, C. Caelles, N. Massot, and F. G. Hegardt. Activation of rat liver cytosolic 3-hydroxy-3-methylglutaryl coenzyme A reductase kinase by adenosine 5'-monophosphate. *Biochem. Biophys. Res. Commun.*, 132(2):497–504, October 1985.
- ⁵⁷ Y. Takai, A. Kishimoto, Y. Iwasa, Y. Kawahara, T. Mori, and Y. Nishizuka. Calcium-dependent activation of a multifunctional protein kinase by membrane phospholipids. *J. Biol. Chem.*, 254(10):3692–3695, May 1979.
- ⁵⁸ Y. Takai, A. Kishimoto, U. Kikkawa, T. Mori, and Y. Nishizuka. Unsaturated diacylglycerol as a possible messenger for the activation of calcium-activated, phospholipid-dependent protein kinase system. *Biochem. Biophys. Res. Commun.*, 91(4):1218–1224, December 1979.
- ⁵⁹ B. M. Burgering and P. J. Coffey. Protein kinase B (c-Akt) in phosphatidylinositol-3-OH kinase signal transduction. *Nature*, 376(6541):599–602, August 1995.
- ⁶⁰ T. Halgren. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des*, 69(2):146–148, February 2007.
- ⁶¹ M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, September 2015.
- ⁶² D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput*, 9(1):687–697, January 2013.
- ⁶³ X. Periole, M. Cavalli, S. J. Marrink, and M. A. Ceruso. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J Chem Theory Comput*, 5(9):2531–2543, September 2009.
- ⁶⁴ X. Periole, T. Huber, S. J. Marrink, and T. P. Sakmar. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *J Am Chem Soc*, 129(33):10126–10132, August 2007.

Bibliography

- ⁶⁵ L. V. Schafer, D. H. de Jong, A. Holt, A. J. Rzepiela, A. H. de Vries, B. Poolman, J. A. Killian, and S. J. Marrink. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *P Natl Acad Sci USA*, 108(4):1343–1348, January 2011.
- ⁶⁶ P. J. Stansfeld, R. Hopkinson, F. M. Ashcroft, and M. S. P. Sansom. PIP2-Binding Site in Kir Channels: Definition by Multiscale Biomolecular Simulations. *Biochemistry-Uk*, 48(46):10926–10933, November 2009.
- ⁶⁷ G. van den Bogaart, K. Meyenberg, H. J. Risselada, H. Amin, K. I. Willig, B. E. Hubrich, M. Dier, S. W. Hell, H. Grubmuller, U. Diederichsen, and R. Jahn. Membrane protein sequestering by ionic protein-lipid interactions. *Nature*, 479(7374):552–555, November 2011.
- ⁶⁸ G. Daum. Lipids of Mitochondria. *Biochim Biophys Acta*, 822(1):1–42, 1985.
- ⁶⁹ T. A. Wassenaar, H. I. Ingolfsson, R. A. Bockmann, D. P. Tieleman, and S. J. Marrink. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J Chem Theory Comput*, 11(5):2144–2155, May 2015.
- ⁷⁰ W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *J Mol Graph Model*, 14(1):33–38, February 1996.
- ⁷¹ W. L. DeLano. *The PyMOL Molecular Graphics System (2002)* DeLano Scientific, Palo Alto, CA, USA. <http://www.pymol.org>. 2002.
- ⁷² K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi, and D. T. Jones. Protein structure prediction servers at University College London. *Nucleic Acids Res.*, 33(Web Server issue):W36–38, July 2005.
- ⁷³ S. M. Claypool, P. Boontheung, J. M. McCaffery, J. A. Loo, and C. M. Koehler. The cardiolipin transacylase, tafazzin, associates with two distinct respiratory components providing insight into Barth syndrome. *Mol. Biol. Cell*, 19(12):5143–5155, December 2008.
- ⁷⁴ N. Gebert, A. S. Joshi, S. Kutik, T. Becker, M. McKenzie, X. L. Guan, V. P. Mooga, D. A. Stroud, G. Kulkarni, M. R. Wenk, P. Rehling, C. Meisinger, M. T. Ryan, N. Wiedemann, M. L.

- Greenberg, and N. Pfanner. Mitochondrial cardiolipin involved in outer-membrane protein biogenesis: implications for Barth syndrome. *Curr. Biol.*, 19(24):2133–2139, December 2009.
- ⁷⁵ S. M. Claypool. Cardiolipin, a critical determinant of mitochondrial carrier protein assembly and function. *Biochim. Biophys. Acta*, 1788(10):2059–2068, October 2009.
- ⁷⁶ J. K. Cullen, N. Abdul Murad, A. Yeo, M. McKenzie, M. Ward, K. L. Chong, N. L. Schieber, R. G. Parton, Y. C. Lim, E. Wolvetang, G. J. Maghzal, R. Stocker, and M. F. Lavin. AarF Domain Containing Kinase 3 (ADCK3) Mutant Cells Display Signs of Oxidative Stress, Defects in Mitochondrial Homeostasis and Lysosomal Accumulation. *PLoS ONE*, 11(2):e0148213, 2016.
- ⁷⁷ A. Tauche, U. Krause-Buchholz, and G. Rödel. Ubiquinone biosynthesis in *Saccharomyces cerevisiae*: the molecular organization of O-methylase Coq3p depends on Abc1p/Coq8p. *FEMS Yeast Res.*, 8(8):1263–1275, December 2008.
- ⁷⁸ E.-N. Vögtle, J. M. Burkhart, H. Gonczarowska-Jorge, C. Kücükköse, A. A. Taskin, D. Kopczynski, R. Ahrends, D. Mossmann, A. Sickmann, R. P. Zahedi, and C. Meisinger. Landscape of submitochondrial protein distribution. *Nat Commun*, 8(1):290, 2017.
- ⁷⁹ N. Jura, N. F. Endres, K. Engel, S. Deindl, R. Das, M. H. Lamers, D. E. Wemmer, X. Zhang, and J. Kuriyan. Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell*, 137(7):1293–1307, June 2009.
- ⁸⁰ T. F. Franke, S. I. Yang, T. O. Chan, K. Datta, A. Kazlauskas, D. K. Morrison, D. R. Kaplan, and P. N. Tsichlis. The protein kinase encoded by the Akt proto-oncogene is a target of the PDGF-activated phosphatidylinositol 3-kinase. *Cell*, 81(5):727–736, June 1995.
- ⁸¹ Y. Nishizuka. Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. *Science*, 258(5082):607–614, October 1992.
- ⁸² K. Gupta, J. A. C. Donlan, J. T. S. Hopper, P. Uzdaviny, M. Landreh, W. B. Struwe, D. Drew, A. J. Baldwin, P. J. Stansfeld, and C. V. Robinson. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature*, 541(7637):421–424, 2017.

Bibliography

- ⁸³ K. Moravcevic, C. L. Oxley, and M. A. Lemmon. Conditional peripheral membrane proteins: facing up to limited specificity. *Structure*, 20(1):15–27, January 2012.
- ⁸⁴ P. L. Yeagle. Non-covalent binding of membrane lipids to membrane proteins. *Biochim. Biophys. Acta*, 1838(6):1548–1559, June 2014.
- ⁸⁵ J. Planas-Iglesias, H. Dwarakanath, D. Mohammadyani, N. Yanamala, V. E. Kagan, and J. Klein-Seetharaman. Cardiolipin Interactions with Proteins. *Biophys. J.*, 109(6):1282–1294, September 2015.
- ⁸⁶ K. Pfeiffer, V. Gohil, R. A. Stuart, C. Hunte, U. Brandt, M. L. Greenberg, and H. Schägger. Cardiolipin stabilizes respiratory chain supercomplexes. *J. Biol. Chem.*, 278(52):52873–52880, December 2003.
- ⁸⁷ B. J. Floyd, E. M. Wilkerson, M. T. Veling, C. E. Minogue, C. Xia, E. T. Beebe, R. L. Wrobel, H. Cho, L. S. Kremer, C. L. Alston, K. A. Gromek, B. K. Dolan, A. Ulbrich, J. A. Stefely, S. L. Bohl, K. M. Werner, A. Jochem, M. S. Westphall, J. W. Rensvold, R. W. Taylor, H. Prokisch, J.-J. P. Kim, J. J. Coon, and D. J. Pagliarini. Mitochondrial Protein Interaction Mapping Identifies Regulators of Respiratory Chain Function. *Mol. Cell*, 63(4):621–632, 2016.
- ⁸⁸ E. Mileykovskaya and W. Dowhan. Cardiolipin membrane domains in prokaryotes and eukaryotes. *Biochim. Biophys. Acta*, 1788(10):2084–2091, October 2009.
- ⁸⁹ M. Fry and D. E. Green. Cardiolipin requirement for electron transfer in complex I and III of the mitochondrial respiratory chain. *J. Biol. Chem.*, 256(4):1874–1880, February 1981.
- ⁹⁰ M. Giustini, F. Castelli, I. Husu, M. Giomini, A. Mallardi, and G. Palazzo. Influence of cardiolipin on the functionality of the Q(a) site of the photosynthetic bacterial reaction center. *J Phys Chem B*, 109(44):21187–21196, November 2005.
- ⁹¹ J. A. Stefely, N. W. Kwiecien, E. C. Freiburger, A. L. Richards, A. Jochem, M. J. P. Rush, A. Ulbrich, K. P. Robinson, P. D. Hutchins, M. T. Veling, X. Guo, Z. A. Kemmerer, K. J. Connors, E. A. Trujillo, J. Sokol, H. Marx, M. S. Westphall, A. S. Hebert, D. J. Pagliarini, and J. J. Coon. Mitochondrial protein functions elucidated by multi-omic mass spectrometry profiling. *Nat Biotechnol*, 34(11):1191–1197, November 2016.

- ⁹² H. E. Klock, E. J. Koesema, M. W. Knuth, and S. A. Lesley. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins*, 71(2):982–994, May 2008.
- ⁹³ P. G. Blommel, P. A. Martin, K. D. Seder, R. L. Wrobel, and B. G. Fox. Flexi vector cloning. *Methods Mol. Biol.*, 498:55–73, 2009.
- ⁹⁴ D. Mumberg, R. Müller, and M. Funk. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, 156(1):119–122, April 1995.
- ⁹⁵ B. G. Fox and P. G. Blommel. Autoinduction of protein expression. *Curr Protoc Protein Sci*, Chapter 5:Unit 5.23, April 2009.
- ⁹⁶ R. D. Gietz and R. A. Woods. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Method Enzymol*, 350:87–96, 2002.
- ⁹⁷ M. Connerth, T. Tatsuta, M. Haag, T. Klecker, B. Westermann, and T. Langer. Intramitochondrial transport of phosphatidic acid in yeast by a lipid transfer protein. *Science*, 338(6108):815–8, November 2012.
- ⁹⁸ D. Wessel and U. I. Flugge. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem*, 138(1):141–3, April 1984.
- ⁹⁹ W. A. Prinz. Lipid Trafficking sans Vesicles: Where, Why, How? *Cell*, 143(6):870–874, December 2010.
- ¹⁰⁰ Y. Bai, J. G. McCoy, E. J. Levin, P. Sobrado, K. R. Rajashankar, B. G. Fox, and M. Zhou. X-ray structure of a mammalian stearyl-CoA desaturase. *Nature*, 524(7564):252–6, August 2015.
- ¹⁰¹ D. I. Chan and H. J. Vogel. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem J*, 430:1–19, August 2010.
- ¹⁰² J. Crosby and M. P. Crump. The structural role of the carrier protein - active controller or passive carrier. *Nat Prod Rep*, 29(10):1111–1137, 2012.
- ¹⁰³ Y. Watanabe, Y. Tamura, S. Kawano, and T. Endo. Structural and mechanistic insights into phospholipid transfer by Ups1-Mdm35 in mitochondria. *Nat Commun*, 6:7922, August 2015.

Bibliography

- ¹⁰⁴ F. Yu, F. He, H. Yao, C. Wang, J. Wang, J. Li, X. Qi, H. Xue, J. Ding, and P. Zhang. Structural basis of intramitochondrial phosphatidic acid transport mediated by Ups1-Mdm35 complex. *EMBO Rep*, 16(7):813–23, July 2015.
- ¹⁰⁵ M. Rossmann, R. Schultz-Heienbrok, J. Behlke, N. Remmel, C. Alings, K. Sandhoff, W. Saenger, and T. Maier. Crystal structures of human saposins C and D: implications for lipid recognition and membrane interactions. *Structure*, 16(5):809–17, May 2008.
- ¹⁰⁶ R. M. Robertson, J. W. Yao, S. Gajewski, G. Kumar, E. W. Martin, C. O. Rock, and S. W. White. A two-helix motif positions the lysophosphatidic acid acyltransferase active site for catalysis within the membrane bilayer. *Nat Struct Mol Biol*, 24(8):666–+, August 2017.
- ¹⁰⁷ D. J. Fazakerley, R. Chaudhuri, P. Yang, G. J. Maghzal, K. C. Thomas, J. R. Krycer, S. J. Humphrey, B. L. Parker, K. H. Fisher-Wellman, C. C. Meoli, N. J. Hoffman, C. Diskin, J. G. Burchfield, M. J. Cowley, W. Kaplan, Z. Modrusan, G. Kolumam, J. Y. Yang, D. L. Chen, D. Samocha-Bonet, J. R. Greenfield, K. L. Hoehn, R. Stocker, and D. E. James. Mitochondrial CoQ deficiency is a common driver of mitochondrial oxidants and insulin resistance. *Elife*, 7, February 2018.
- ¹⁰⁸ B. G. Hughes, P. M. Harrison, and S. Hekimi. Estimating the occurrence of primary ubiquinone deficiency by analysis of large-scale sequencing data. *Sci Rep*, 7(1):17744, December 2017.
- ¹⁰⁹ U. C. Tran and C. F. Clarke. Endogenous synthesis of coenzyme Q in eukaryotes. *Mitochondrion*, 7 Suppl:S62–71, June 2007.
- ¹¹⁰ G. Metz, K. P. Howard, W. B. S. Vanliemt, J. H. Prestegard, J. Lugtenburg, and S. O. Smith. Nmr-Studies of Ubiquinone Location in Oriented Model Membranes - Evidence for a Single Motionally-Averaged Population. *J Am Chem Soc*, 117(1):564–565, January 1995.
- ¹¹¹ E. L. Ulrich, M. E. Girvin, W. A. Cramer, and J. L. Markley. Location and mobility of ubiquinones of different chain lengths in artificial membrane vesicles. *Biochemistry*, 24(10):2501–8, May 1985.
- ¹¹² K. Danhauser, D. Herebian, T. B. Haack, R. J. Rodenburg, T. M. Strom, T. Meitinger, D. Klee, E. Mayatepek, H. Prokisch, and F. Distelmaier. Fatal neonatal encephalopathy and lactic

- acidosis caused by a homozygous loss-of-function variant in COQ9. *Eur J Hum Genet*, 24(3):450–4, March 2016.
- ¹¹³ A. J. Duncan, M. Bitner-Glindzicz, B. Meunier, H. Costello, I. P. Hargreaves, L. C. Lopez, M. Hirano, C. M. Quinzii, M. I. Sadowski, J. Hardy, A. Singleton, P. T. Clayton, and S. Rahman. A nonsense mutation in COQ9 causes autosomal-recessive neonatal-onset primary coenzyme Q10 deficiency: a potentially treatable form of mitochondrial disease. *Am J Hum Genet*, 84(5):558–66, May 2009.
- ¹¹⁴ L. Garcia-Corzo, M. Luna-Sanchez, C. Doerrier, J. A. Garcia, A. Guaras, R. Acin-Perez, J. Bullejos-Peregrin, A. Lopez, G. Escames, J. A. Enriquez, D. Acuna-Castroviejo, and L. C. Lopez. Dysfunctional Coq9 protein causes predominant encephalomyopathy associated with CoQ deficiency. *Hum Mol Genet*, 22(6):1233–48, March 2013.
- ¹¹⁵ A. Johnson, P. Gin, B. N. Marbois, E. J. Hsieh, M. Wu, M. H. Barros, C. F. Clarke, and A. Tzagoloff. COQ9, a new gene required for the biosynthesis of coenzyme Q in *Saccharomyces cerevisiae*. *J Biol Chem*, 280(36):31397–404, September 2005.
- ¹¹⁶ M. Luna-Sanchez, E. Diaz-Casado, E. Barca, M. A. Tejada, A. Montilla-Garcia, E. J. Cobos, G. Escames, D. Acuna-Castroviejo, C. M. Quinzii, and L. C. Lopez. The clinical heterogeneity of coenzyme Q10 deficiency results from genotypic differences in the Coq9 gene. *EMBO Mol Med*, 7(5):670–87, May 2015.
- ¹¹⁷ A. C. Smith, Y. Ito, A. Ahmed, J. A. Schwartzentruber, C. L. Beaulieu, E. Aberg, J. Majewski, D. E. Bulman, K. Horsting-Wethly, D. V. Koning, Consortium Care4Rare Canada, R. J. Rodenburg, K. M. Boycott, and L. S. Penney. A family segregating lethal neonatal coenzyme Q10 deficiency caused by mutations in COQ9. *J Inherit Metab Dis*, March 2018.
- ¹¹⁸ D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J Comput Chem*, 26(16):1701–1718, December 2005.
- ¹¹⁹ T. A. Wassenaar, K. Pluhackova, R. A. Bockmann, S. J. Marrink, and D. P. Tieleman. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J Chem Theory Comput*, 10(2):676–690, February 2014.

Bibliography

- ¹²⁰ B. Webb and A. Sali. Protein Structure Modeling with MODELLER. In Michael Kaufmann, Claudia Klinger, and Andreas Savelsbergh, editors, *Functional Genomics: Methods and Protocols*, Methods in Molecular Biology, pages 39–54. Springer New York, New York, NY, 2017.
- ¹²¹ Djurre H. de Jong, Nicoletta Liguori, Tom van den Berg, Clement Arnarez, Xavier Periole, and Siewert J. Marrink. Atomistic and Coarse Grain Topologies for the Cofactors Associated with the Photosystem II Core Complex. *J Phys Chem B*, 119(25):7791–7803, June 2015.
- ¹²² Z. Yu, S. E. Reichheld, A. Savchenko, J. Parkinson, and A. R. Davidson. A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *J Mol Biol*, 400(4):847–64, July 2010.
- ¹²³ R. Abagyan and M. Totrov. Biased Probability Monte-Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. *J Mol Biol*, 235(3):983–1002, January 1994.
- ¹²⁴ R. D. Finn, P. Coghill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, and A. Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44(D1):D279–D285, January 2016.
- ¹²⁵ P. Stenmark, J. Grunler, J. Mattsson, P. J. Sindelar, P. Nordlund, and D. A. Berthold. A new member of the family of di-iron carboxylate proteins. Coq7 (clk-1), a membrane-bound hydroxylase involved in ubiquinone biosynthesis. *J Biol Chem*, 276(36):33297–300, September 2001.
- ¹²⁶ C. Busso, J. R. Ferreira, J. A. Paulela, L. Bleicher, M. Demasi, and M. H. Barros. Coq7p relevant residues for protein activity and stability. *Biochimie*, 119:92–102, December 2015.
- ¹²⁷ S. Rea. CLK-1/Coq7p is a DMQ mono-oxygenase and a new member of the di-iron carboxylate protein family. *Febs Lett*, 509(3):389–394, December 2001.
- ¹²⁸ C. H. Wu, R. Apweiler, A. Bairoch, D. A. Natale, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Z. Huang, R. Lopez, M. Magrane, M. J. Martin, R. Mazumder, C. O'Donovan,

- N. Redaschi, and B. Suzek. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*, 34:D187–D191, January 2006.
- ¹²⁹ T. Schwede, J. Kopp, N. Guex, and M. C. Peitsch. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*, 31(13):3381–3385, July 2003.
- ¹³⁰ P. Benkert, M. Biasini, and T. Schwede. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27(3):343–350, February 2011.
- ¹³¹ F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Z. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, and D. G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, 7, October 2011.
- ¹³² B. G. Pierce, K. Wiehe, H. Hwang, B. H. Kim, T. Vreven, and Z. P. Weng. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, 30(12):1771–1773, June 2014.
- ¹³³ R. Chen, L. Li, and Z. P. Weng. ZDOCK: An initial-stage protein-docking algorithm. *Proteins*, 52(1):80–87, July 2003.
- ¹³⁴ L. Cuthbertson and J. R. Nodwell. The TetR family of regulators. *Microbiol Mol Biol Rev*, 77(3):440–75, September 2013.
- ¹³⁵ J. L. Ramos, M. Martínez-Bueno, A. J. Molina-Henares, W. Terán, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan, and R. Tobes. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev*, 69(2):326–56, June 2005.
- ¹³⁶ N. K. Chua, V. Howe, N. Jatana, L. Thukral, and A. J. Brown. A conserved degron containing an amphipathic helix regulates the cholesterol-mediated turnover of human squalene monooxygenase, a rate-limiting enzyme in cholesterol synthesis. *J Biol Chem*, 292(49):19959–19973, December 2017.
- ¹³⁷ R. B. Cornell. Membrane lipid compositional sensing by the inducible amphipathic helix of CCT. *Biochim Biophys Acta*, 1861(8 Pt B):847–861, August 2016.

Bibliography

- ¹³⁸ S. Suetsugu, S. Kurisu, and T. Takenawa. Dynamic shaping of cellular membranes by phospholipids and membrane-deforming proteins. *Physiol Rev*, 94(4):1219–48, October 2014.
- ¹³⁹ D. Acehan, A. Malhotra, Y. Xu, M. Ren, D. L. Stokes, and M. Schlame. Cardiolipin affects the supramolecular organization of ATP synthase in mitochondria. *Biophys J*, 100(9):2184–92, May 2011.
- ¹⁴⁰ V. M. Gohil, P. Hayes, S. Matsuyama, H. Schagger, M. Schlame, and M. L. Greenberg. Cardiolipin biosynthesis and mitochondrial respiratory chain function are interdependent. *J Biol Chem*, 279(41):42612–8, October 2004.
- ¹⁴¹ N. Ikon and R. O. Ryan. Cardiolipin and mitochondrial cristae organization. *Biochim Biophys Acta*, 1859(6):1156–1163, June 2017.
- ¹⁴² A. Grabon, A. Orlowski, A. Tripathi, J. Vuorio, M. Javanainen, T. Rog, M. Lonnfors, M. I. McDermott, G. Siebert, P. Somerharju, I. Vattulainen, and V. A. Bankaitis. Dynamics and energetics of the mammalian phosphatidylinositol transfer protein phospholipid exchange cycle. *J Biol Chem*, 292(35):14438–14455, September 2017.
- ¹⁴³ Y. Wang, C. Smith, J. S. Parboosingh, A. Khan, M. Innes, and S. Hekimi. Pathogenicity of two COQ7 mutations and responses to 2,4-dihydroxybenzoate bypass treatment. *J Cell Mol Med*, April 2017.
- ¹⁴⁴ C. H. He, D. S. Black, T. P. Nguyen, C. Wang, C. Srinivasan, and C. F. Clarke. Yeast Coq9 controls deamination of coenzyme Q intermediates that derive from para-aminobenzoic acid. *Biochim Biophys Acta*, 1851(9):1227–39, September 2015.
- ¹⁴⁵ E. J. Hsieh, P. Gin, M. Gulmezian, U. C. Tran, R. Saiki, B. N. Marbois, and C. F. Clarke. *Saccharomyces cerevisiae* Coq9 polypeptide is a subunit of the mitochondrial coenzyme Q biosynthetic complex. *Arch Biochem Biophys*, 463(1):19–26, July 2007.
- ¹⁴⁶ R. K. Behan and S. J. Lippard. The Aging-Associated Enzyme CLK-1 Is a Member of the Carboxylate-Bridged Diiron Family of Proteins. *Biochemistry-U S*, 49(45):9679–9681, November 2010.

- ¹⁴⁷ T. T. Lu, S. J. Lee, U. P. Apfel, and S. J. Lippard. Aging-associated enzyme human clock-1: substrate-mediated reduction of the diiron center for 5-demethoxyubiquinone hydroxylation. *Biochemistry*, 52(13):2236–44, April 2013.
- ¹⁴⁸ U. C. Tran, B. Marbois, P. Gin, M. Gulmezian, T. Jonassen, and C. F. Clarke. Complementation of *Saccharomyces cerevisiae* coq7 mutants by mitochondrial targeting of the *Escherichia coli* UbiF polypeptide: two functions of yeast Coq7 polypeptide in coenzyme Q biosynthesis. *J Biol Chem*, 281(24):16401–9, June 2006.
- ¹⁴⁹ V. A. Bankaitis, C. J. Mousley, and G. Schaaf. The Sec14 superfamily and mechanisms for crosstalk between lipid metabolism and lipid signaling. *Trends Biochem Sci*, 35(3):150–160, March 2010.
- ¹⁵⁰ K. E. Ile, G. Schaaf, and V. A. Bankaitis. Phosphatidylinositol transfer proteins and cellular nanoreactors for lipid signaling. *Nat Chem Biol*, 2(11):576–583, November 2006.
- ¹⁵¹ F. Pierrel. Impact of Chemical Analogs of 4-Hydroxybenzoic Acid on Coenzyme Q Biosynthesis: From Inhibition to Bypass of Coenzyme Q Deficiency. *Front Physiol*, 8:436, 2017.
- ¹⁵² C. H. He, L. X. Xie, C. M. Allan, U. C. Tran, and C. F. Clarke. Coenzyme Q supplementation or over-expression of the yeast Coq8 putative kinase stabilizes multi-subunit Coq polypeptide complexes in yeast coq null mutants. *Biochimica et biophysica acta*, 1841(4):630–44, April 2014.
- ¹⁵³ L. X. Xie, M. Ozeir, J. Y. Tang, J. Y. Chen, S. K. Jaquinod, M. Fontecave, C. F. Clarke, and F. Pierrel. Overexpression of the Coq8 kinase in *Saccharomyces cerevisiae* coq null mutants allows for accumulation of diagnostic intermediates of the coenzyme Q6 biosynthetic pathway. *J Biol Chem*, 287(28):23571–81, July 2012.
- ¹⁵⁴ M. Ozeir, L. Pelosi, A. Ismail, C. Mellot-Draznieks, M. Fontecave, and F. Pierrel. Coq6 is responsible for the C4-deamination reaction in coenzyme Q biosynthesis in *Saccharomyces cerevisiae*. *J Biol Chem*, 290(40):24140–51, October 2015.
- ¹⁵⁵ L. H. Wong, A. Copic, and T. P. Levine. Advances on the Transfer of Lipids by Lipid Transfer Proteins. *Trends Biochem Sci*, 42(7):516–530, July 2017.

Bibliography

- ¹⁵⁶ A. Murley and J. Nunnari. The Emerging Network of Mitochondria-Organellar Contacts. *Mol Cell*, 61(5):648–653, March 2016.
- ¹⁵⁷ Y. Tamura and T. Endo. Role of Intra- and Inter-mitochondrial Membrane Contact Sites in Yeast Phospholipid Biogenesis. *Adv Exp Med Biol*, 997:121–133, 2017.
- ¹⁵⁸ S. Lahiri, A. Toulmay, and W. A. Prinz. Membrane contact sites, gateways for lipid homeostasis. *Curr Opin Cell Biol*, 33:82–87, April 2015.
- ¹⁵⁹ Paul A. Srere. The metabolon. *Trends in Biochemical Sciences*, 10(3):109–110, March 1985.
- ¹⁶⁰ G. Drin, J.-F. Casella, R. Gautier, T. Boehmer, T. U. Schwartz, and B. Antonny. A general amphipathic alpha-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.*, 14(2):138–146, February 2007.
- ¹⁶¹ R. P. McAndrew, Y. Wang, A.-W. Mohsen, M. He, J. Vockley, and J.-J. P. Kim. Structural basis for substrate fatty acyl chain specificity: crystal structure of human very-long-chain acyl-CoA dehydrogenase. *J. Biol. Chem.*, 283(14):9435–9443, April 2008.
- ¹⁶² Y. Zhang, S. S. Bharathi, M. J. Rardin, R. Uppala, E. Verdin, B. W. Gibson, and E. S. Goetzman. SIRT3 and SIRT5 regulate the enzyme activity and cardiolipin binding of very long-chain acyl-CoA dehydrogenase. *PLoS ONE*, 10(3):e0122297, 2015.
- ¹⁶³ H. F. Hofbauer, M. Gecht, S. C. Fischer, A. Seybert, A. S. Frangakis, E. H. K. Stelzer, R. Covino, G. Hummer, and R. Ernst. The molecular recognition of phosphatidic acid by an amphipathic helix in Opi1. *J. Cell Biol.*, 217(9):3109–3126, September 2018.
- ¹⁶⁴ A. G. Reidenbach, Z. A. Kemmerer, D. Aydin, A. Jochem, M. T. McDevitt, P. D. Hutchins, J. L. Stark, J. A. Stefely, T. Reddy, A. S. Hebert, E. M. Wilkerson, I. E. Johnson, C. A. Bingman, J. L. Markley, J. J. Coon, M. Dal Peraro, and D. J. Pagliarini. Conserved Lipid and Small-Molecule Modulation of COQ8 Reveals Regulation of the Ancient Kinase-like UbiB Family. *Cell Chem Biol*, 25(2):154–165 e11, February 2018.
- ¹⁶⁵ L. Malinina, D. J. Patel, and R. E. Brown. How alpha-Helical Motifs Form Functionally Diverse Lipid-Binding Compartments. *Annu Rev Biochem*, 86:609–636, June 2017.

- ¹⁶⁶ T. Tatsuta, M. Scharwey, and T. Langer. Mitochondrial lipid trafficking. *Trends Cell Biol*, 24(1):44–52, January 2014.
- ¹⁶⁷ J. R. Alattia, J. E. Shaw, C. M. Yip, and G. G. Prive. Direct visualization of saposin remodelling of lipid bilayers. *J Mol Biol*, 362(5):943–53, October 2006.
- ¹⁶⁸ C. H. Hill, G. M. Cook, S. J. Spratley, S. Fawke, S. C. Graham, and J. E. Deane. The mechanism of glycosphingolipid degradation revealed by a GALC-SapA complex structure. *Nat Commun*, 9(1):151, January 2018.
- ¹⁶⁹ W. Furst and K. Sandhoff. Activator proteins and topology of lysosomal sphingolipid catabolism. *Biochim Biophys Acta*, 1126(1):1–16, June 1992.
- ¹⁷⁰ I. Gonzalez-Mariscal, E. Garcia-Teston, S. Padilla, A. Martin-Montalvo, T. Pomares-Viciano, L. Vazquez-Fonseca, P. Gandolfo-Dominguez, and C. Santos-Ocana. Regulation of coenzyme Q biosynthesis in yeast: a new complex in the block. *IUBMB life*, 66(2):63–70, February 2014.
- ¹⁷¹ W. Kabsch. Xds. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2):125–32, February 2010.
- ¹⁷² P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, 66:213–221, February 2010.
- ¹⁷³ A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *J Appl Crystallogr*, 40(Pt 4):658–674, August 2007.
- ¹⁷⁴ P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr*, 68(Pt 4):352–67, April 2012.
- ¹⁷⁵ P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4):486–501, April 2010.

Bibliography

- ¹⁷⁶ P. D. Hutchins. LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. *Cell Systems*, 6, 2018.
- ¹⁷⁷ I. R. Boldogh and L. A. Pon. Purification and subfractionation of mitochondria from the yeast *Saccharomyces cerevisiae*. *Mitochondria, 2nd Edition*, 80:45–64, 2007.
- ¹⁷⁸ A. S. Hebert, A. E. Merrill, J. A. Stefely, D. J. Bailey, C. D. Wenger, M. S. Westphall, D. J. Pagliarini, and J. J. Coon. Amine-reactive Neutron-encoded Labels for Highly Plexed Proteomic Quantitation. *Mol Cell Proteomics*, 12(11):3360–3369, November 2013.
- ¹⁷⁹ B. Marbois, L. X. Xie, S. Choi, K. Hirano, K. Hyman, and C. F. Clarke. para-Aminobenzoic acid is a precursor in coenzyme Q6 biosynthesis in *Saccharomyces cerevisiae*. *J Biol Chem*, 285(36):27827–38, September 2010.
- ¹⁸⁰ J. Cox, M. Y. Hein, C. A. Lubner, I. Paron, N. Nagaraj, and M. Mann. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, 13(9):2513–26, September 2014.
- ¹⁸¹ J. Cox and M. Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*, 26(12):1367–72, December 2008.
- ¹⁸² J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*, 10(4):1794–805, April 2011.
- ¹⁸³ H. Gohlke and G. Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed. Engl.*, 41(15):2644–2676, August 2002.
- ¹⁸⁴ X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, 7(2):146–157, June 2011.
- ¹⁸⁵ D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*, 3(11):935–949, November 2004.

- ¹⁸⁶ N. S. Pagadala, K. Syed, and J. Tuszynski. Software for molecular docking: a review. *Biophys Rev*, 9(2):91–102, April 2017.
- ¹⁸⁷ B. D. Bursulaya, M. Totrov, R. Abagyan, and C. L. Brooks. Comparative study of several algorithms for flexible ligand docking. *J. Comput. Aided Mol. Des.*, 17(11):755–763, November 2003.
- ¹⁸⁸ L. G. Ferreira, R. N. Dos Santos, G. Oliva, and A. D. Andricopulo. Molecular docking and structure-based drug design strategies. *Molecules*, 20(7):13384–13421, July 2015.
- ¹⁸⁹ S. F. Sousa, P. A. Fernandes, and M. J. Ramos. Protein-ligand docking: current status and future challenges. *Proteins*, 65(1):15–26, October 2006.
- ¹⁹⁰ E. Yuriev and P. A. Ramsland. Latest developments in molecular docking: 2010-2011 in review. *J. Mol. Recognit.*, 26(5):215–239, May 2013.
- ¹⁹¹ J. Gu, X. Yang, L. Kang, J. Wu, and X. Wang. MoDock: A multi-objective strategy improves the accuracy for molecular docking. *Algorithms Mol Biol*, 10:8, 2015.
- ¹⁹² R. Spitzer and A. N. Jain. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. Aided Mol. Des.*, 26(6):687–699, June 2012.
- ¹⁹³ C. M. Venkatachalam, X. Jiang, T. Oldfield, and M. Waldman. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graph. Model.*, 21(4):289–307, January 2003.
- ¹⁹⁴ O. A. Sauer, D. M. Shepard, and T. R. Mackie. Application of constrained optimization to radiotherapy planning. *Med Phys*, 26(11):2359–2366, November 1999.
- ¹⁹⁵ A. V. Tikhonravov, M. K. Trubetskov, and T. V. Amotchkina. Application of constrained optimization to the design of quasi-rugate optical coatings. *Appl Opt*, 47(28):5103–5109, October 2008.
- ¹⁹⁶ C. A. Coello Coello. Use of a self-adaptive penalty approach for engineering optimization problems. *Computers in Industry*, 41(2):113–127, March 2000.
- ¹⁹⁷ R. Quiroga and M. A. Villarreal. Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLoS ONE*, 11(5):e0155183, 2016.

Bibliography

- ¹⁹⁸ V. Y. Tanchuk, V. O. Tanin, A. I. Vovk, and G. Poda. A New, Improved Hybrid Scoring Function for Molecular Docking and Scoring Based on AutoDock and AutoDock Vina. *Chem Biol Drug Des*, 87(4):618–625, April 2016.
- ¹⁹⁹ M. A. C. Neves, M. Totrov, and R. Abagyan. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.*, 26(6):675–686, June 2012.
- ²⁰⁰ P. J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, 28(7):849–857, July 1985.
- ²⁰¹ S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell, and A. J. Olson. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat Protoc*, 11(5):905–919, May 2016.
- ²⁰² R. Harris, A. J. Olson, and D. S. Goodsell. Automated prediction of ligand-binding sites in proteins. *Proteins*, 70(4):1506–1517, March 2008.
- ²⁰³ M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.*, 15(6):359–363, 389, December 1997.
- ²⁰⁴ V. Zoete, T. Schuepbach, C. Bovigny, P. Chaskar, A. Daina, U. F. Röhrig, and O. Michielin. Attracting cavities for docking. Replacing the rough energy landscape of the protein by a smooth attracting landscape. *J Comput Chem*, 37(4):437–447, February 2016.
- ²⁰⁵ A. R. Leach, M. M. Hann, J. N. Burrows, and E. J. Griffen. Fragment screening: an introduction. *Mol Biosyst*, 2(9):430–446, September 2006.
- ²⁰⁶ M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 261(3):470–489, August 1996.
- ²⁰⁷ B. C. Roberts and R. L. Mancera. Ligand-protein docking with water molecules. *J Chem Inf Model*, 48(2):397–408, February 2008.

-
- ²⁰⁸ C. M. Labbé, T. Pencheva, D. Jereva, D. Desvillechabrol, J. Becot, B. O. Villoutreix, I. Pajeva, and M. A. Miteva. AMMOS2: a web server for protein-ligand-water complexes refinement via molecular mechanics. *Nucleic Acids Res.*, 45(W1):W350–W355, July 2017.
- ²⁰⁹ D. Bucher, P. Stouten, and N. Triballeau. Shedding Light on Important Waters for Drug Design: Simulations versus Grid-Based Methods. *J Chem Inf Model*, 58(3):692–699, 2018.
- ²¹⁰ D. Chen, G. Menche, T. D. Power, L. Sower, J. W. Peterson, and C. H. Schein. Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. *Proteins*, 67(3):593–605, May 2007.
- ²¹¹ P. Chaskar, V. Zoete, and U. F. Röhrig. On-the-Fly QM/MM Docking with Attracting Cavities. *J Chem Inf Model*, 57(1):73–84, 2017.
- ²¹² M. A. Lill. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry*, 50(28):6157–6169, July 2011.
- ²¹³ R. P. Venkitakrishnan, E. Zaborowski, D. McElheny, S. J. Benkovic, H. J. Dyson, and P. E. Wright. Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle. *Biochemistry*, 43(51):16046–16055, December 2004.
- ²¹⁴ V. Oleinikovas, G. Saladino, B. P. Cossins, and F. L. Gervasio. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.*, 138(43):14257–14263, 2016.
- ²¹⁵ G. Bottegoni, W. Rocchia, M. Rueda, R. Abagyan, and A. Cavalli. Systematic exploitation of multiple receptor conformations for virtual ligand screening. *PLoS ONE*, 6(5):e18845, 2011.

Deniz Aydin

✉ denizaydin1990@gmail.com in www.linkedin.com/in/deniz-aydin-biotech
📞 0000-0002-2758-2480 🌐 http://bit.ly/ScholarDA



Education

- 11/2014 - 06/2019 **Ph.D. in Biotechnology and Bioengineering**
Swiss Federal Institute of Technology (EPFL)
Advisor: Matteo Dal Peraro
Thesis: Characterization of coenzyme Q biosynthesis proteins through integrative modeling
Committee: Siewert-Jan Marrink, Stefano Vanni, Giovanni D'Angelo
- 09/2012 - 09/2014 **M.Sc. in Chemical and Biological Engineering**
Koç University, Istanbul
Advisors: Burak Erman, Mehmet Sayar
Thesis: Molecular dynamics studies to understand pathogenicity in Behçet's Disease
GPA: 3.79 / 4.00
- 09/2008 - 06/2012 **B.Sc. in Chemistry**
Istanbul Technical University
Advisor: Umit Tunca
Thesis: Main chain functionalization of polyoxanorbornenes via thiol-ene click reactions
GPA: 3.71 / 4.00
First-Class Honors
Ranked 1st in graduating class
- 01/2011 - 07/2011 **Erasmus European Exchange Programme Scholar**
University of Nantes, France
Advisor: Valérie Fargeas
Research topic: Synthetic and structural organic chemistry

Employment History

- 11/2014 - 07/2019 **Swiss Federal Institute of Technology (EPFL)**
Research Assistant, Laboratory for Biomolecular Modeling
Teaching Assistant, School of Life Sciences
- 02/2017 - 08/2017 **University of California San Diego (UCSD)**
Visiting Researcher, SNSF Doctoral Mobility Fellow
Skaggs School of Pharmacy and Pharmaceutical Sciences
Advisor: Prof. Ruben Abagyan
Research topic: Structure-based drug design, molecular docking
- 09/2012 - 09/2014 **Koç University, Istanbul**
Research Assistant, Computational Drug Design & Quantitative Biology Lab
Teaching Assistant, Chemical and Biological Engineering Department
- 09/2011 - 02/2012 **Istanbul Technical University**
Undergraduate Researcher in Complex Macromolecular Structure Center
- 08/2011 - 09/2011 **Diversey, Inc., Istanbul**
Intern in Chemical Manufacturing Department
- 05/2011 - 06/2011 **University of Nantes, France**
Intern in Organic Synthesis Laboratory
- 06/2010 - 07/2010 **Istanbul Technical University**
Intern in Complex Macromolecular Structure Center
- 08/2009 - 09/2009 **The Scientific and Technological Research Council of Turkey (TUBITAK)**
Intern in Polymer and Organic Chemistry Technologies Laboratory

Awards and Fellowships

- 06/2019 **European Molecular Biology Organization Long-Term Fellowship**, *EMBO*
- 05/2019 **Swiss National Science Foundation Early Postdoctoral Mobility Fellowship**, *Switzerland*
- 01/2019 **The American Chemical Society Division of Biological Chemistry Travel Award**, *USA*
- 02/2018 **The Biophysical Society Student Research Achievement Award**, *USA*
- 02/2018 **The Biophysical Society Best Poster Presentation in Membrane Structure & Assembly**, *USA*
- 02/2018 **The Biophysical Society International Travel Award**, *USA*
- 02/2017 - 08/2017 **Swiss National Science Foundation Doctoral Mobility Fellowship**, *Switzerland*
- 09/2012 - 09/2014 **Koç University Academic Excellence Scholarship and stipend for M.Sc. studies**, *Turkey*
- 06/2012 **Ranked first in Department of Chemistry graduating class**, *Turkey*
- 06/2012 **Bachelor of Science with First-Class Honors**, *Turkey*
- 01/2011 - 07/2011 **Erasmus European Exchange Programme Scholarship**, *Turkey*

Publications in Peer-Reviewed Journals

- 2019 Lohman DC*, **Aydin D***, Vonbank H, Smith R, Linke V, Weisenhorn E, McDevitt MT, Hutchins P, Wilkerson E, Russell J, Stefely MS, Beebe ET, Jochem A, Bingman CA, Coon JJ, Dal Peraro M, Pagliarini DJ.
“An Isoprene Lipid-Binding Protein Promotes Eukaryotic Coenzyme Q Biosynthesis.”
Molecular Cell 73(4): 763-774. (*Co-First Authorship). URL: 10.1016/j.molcel.2018.11.033
Media coverage: [EPFL SV News](#), [EurekAlert-AAAS](#), [AlphaGalileo](#). Highlighted with a [Preview](#).
- 2018 Reidenbach AG, Kemmerer ZA, **Aydin D**, Jochem A, McDevitt MT, Hutchins PD, Wilkerson EM, Stark JL, Stefely JA, Johnson IE, Bingman CA, Markley JL, Coon JJ, Dal Peraro M, Pagliarini DJ.
“Conserved Lipid and Small-Molecule Modulation of COQ8 Reveals Regulation of the Ancient Kinase-Like UbiB Family.”
Cell Chemical Biology 25(2): 154-165. ([Cover Article](#), highlighted with a [Preview](#)).
URL: 10.1016/j.chembiol.2017.11.001
- 2016 Stefely JA, Licitra F, Laredj L, Reidenbach AG, Kemmerer ZA, Grangeray A, Jaeg T, Minogue CE, Ulbrich A, Hutchins PD, Wilkerson EM, Ruan Z, **Aydin D**, Hebert AS, Guo X, Freiburger EC, Reutenauer L, Jochem A, Chergova M, Johnson IE, Lohman DC, Rush MJP, Kwiecien NW, Singh PK, Floyd BJ, Forsman U, Sindelar PJ, Westphall MS, Pierrel F, Dal Peraro M, Kannan N, Bingman CA, Coon JJ, Isope P, Puccio H, Pagliarini DJ.
“Cerebellar Ataxia and Coenzyme Q Deficiency Through Loss of Unorthodox Kinase Activity.”
Molecular Cell 63(4): 608-620. ([Cover Article](#)). URL: 10.1016/j.molcel.2016.06.030

Unpublished Manuscripts in Preparation

1. **Aydin D***, Prunotto A*, Abriata LA, Dal Peraro M.
“Integrative Modeling at the Protein-Membrane Interface.”
(Invited review article for *Accounts of Chemical Research*, *Co-First Authorship)
2. Tamò G*, **Aydin D***, Träger S, Dal Peraro M.
“Using Energy Grids to Constrain the Search Space of Small Molecules During Docking.”
*(Co-First Authorship)
3. Träger S, Tamò G, **Aydin D**, Fonti G, Dal Peraro M.
“CLONe: Clustering Based on Local Density Neighborhoods Applied to Integrative Dynamic Modeling.”

Oral Presentations in International Conferences

- 2019 **Aydin D.** “Overcoming the Hydrophobic Barrier of the Membrane: A Lipid-Binding Protein Promotes Coenzyme Q Biosynthesis.”
The American Chemical Society (ACS) National Meeting, Orlando, FL, USA, March 31 - April 4.
Division of Biological Chemistry Travel Award Winning Abstract.
- 2019 **Aydin D.** “Overcoming the Hydrophobic Barrier of the Membrane: A Lipid-Binding Protein Promotes Coenzyme Q Biosynthesis.”
CECAM Meeting on Multiscale Modeling, EPFL, Switzerland, February 4 - 6.
- 2018 **Aydin D.** “Isoprene and Membrane Binding Capabilities of COQ9 in Coenzyme Q Biosynthesis.”
The Young Scientist Symposium, Bordeaux, France, May 4 - 5.
- 2017 **Aydin D.** “An Integrative Approach to Study Membrane Proteins, and New Optimization Strategies for Computational Drug Design.”
Invited talk at Stanford University, CA, USA, July 3.
- 2017 **Aydin D.** “An Integrated Computational and Experimental Study to Investigate the Role of COQ8 in Promoting Coenzyme Q Biosynthesis.”
Gordon Research Seminar: Proteins, Holderness, NH, USA, June 17-18.
Participated as part of the SNSF Doctoral Mobility Fellowship.
- 2017 **Aydin D.** “An Integrated Computational and Experimental Study to Investigate the Role of COQ8 in Promoting Coenzyme Q Biosynthesis.”
Lightning talk in *Gordon Research Conference: Proteins*, Holderness, NH, USA, June 18-23.
Participated as part of the SNSF Doctoral Mobility Fellowship.

Poster Presentations in International Conferences

- 2018 **Aydin D,** Lohman DC, Pagliarini DJ, Dal Peraro M. “An Integrated Computational and Experimental Study to Investigate the Role of COQ9 in Promoting CoQ Biosynthesis.”
The Biophysical Society (BPS) Annual Meeting, San Francisco, CA, USA, February 17-21.
Best Poster Presentation and Student Research Achievement Award in Membrane Structure & Assembly.
International Travel Award Winning Abstract.
- 2017 **Aydin D,** Reidenbach AG, Pagliarini DJ, Dal Peraro M. “An Integrated Computational and Experimental Study to Investigate the Role of COQ8 in Promoting CoQ Biosynthesis.”
Computational Advances in Drug Discovery, EPFL, Switzerland, September 5-8.
- 2017 **Aydin D,** Reidenbach AG, Pagliarini DJ, Dal Peraro M. “An Integrated Computational and Experimental Study to Investigate the Role of COQ8 in Promoting CoQ Biosynthesis.”
Gordon Research Conference: Proteins, Holderness, NH, USA, June 18-23.
Participated as part of the SNSF Doctoral Mobility Fellowship.
- 2016 **Aydin D,** Dal Peraro M. “Unraveling the Interaction Surface of COQ8A Kinase with Inner Mitochondrial Membrane.”
The Biophysical Society (BPS) Thematic Meeting on Liposomes, Exosomes, and Virosomes, Ascona, Switzerland, September 11-16.
- 2016 **Aydin D,** Reidenbach AG, Pagliarini DJ, Dal Peraro M. “Unraveling the Interaction Surface of COQ8A Kinase with Inner Mitochondrial Membrane.”
Swiss Institute of Bioinformatics (SIB) Days, Biel/Bienne, Switzerland, June 7-8.
- 2015 **Aydin D,** Marcaida MJ, Dal Peraro M. “Atomistic Molecular Dynamics Study of 4-Thiazolidinone Compounds Bound to Aurora-A Kinase.”
Computational Advances in Drug Discovery, EPFL, Switzerland, September 22-25.

Supervision of Students

- 06/2018 - 07/2018 **Bachelor's internship project**, Yildiz Aydin, *EPFL*.
"Atomistic molecular dynamics simulations to probe protein-protein interactions."
Supervised the parametrization of a large dataset for the initial phase of a Swiss National Supercomputing Centre (CSCS) funded project.
- 03/2016 - 01/2017 **Master's internship project**, Alexander Lorkowski, *EPFL*.
"Comparative molecular dynamics simulations of atypical kinase COQ8 structures."
Supervised the comparative analysis of ATP- and ADP-bound, and unbound states of COQ8 to understand how ligand binding affects the conformational dynamics.
- 09/2015 - 02/2016 **Master's internship project**, Helena Vincentelli, *EPFL*.
"Structural analysis and ligand binding capabilities of COQ8 and kinase-like UbiB proteins."
Supervised the investigation of small-molecule binding capabilities and dimerization mechanism of COQ8 and homologues.

Teaching Activities

- Spring 2016 & Spring 2015 **Graduate Biomolecular Structure and Mechanics**, *EPFL*
Developed exercises and projects; corrected assignments; assisted students for the practical sessions on computational structural biology; and supervised students for their semester project and presentation.
- Spring 2016 **Undergraduate Global Issues: Health**, *EPFL*
Proposed and assigned major global health issues to five groups of students; and supervised the preparation of their semester project and final poster presentation.
- Fall 2015 **Graduate Scientific Literature Analysis in Bioengineering**, *EPFL*
Proposed and assigned controversial, retracted articles to a group of students; and supervised the critical analysis of the scientific literature and the preparation of an oral presentation.
- Spring 2015 **Undergraduate Physical Biology of the Cell II**, *EPFL*
Developed weekly exercises and assisted students for the practical sessions.
- Spring 2014 **Undergraduate General Chemistry Lab**, *Koç University*
Developed weekly quizzes; graded quizzes and lab reports; and taught lab sessions.
- Fall 2013 **Graduate Statistical Thermodynamics**, *Koç University*
Developed and graded weekly exercises; and taught practical sessions.
- Spring 2013 **Undergraduate Introduction to Science**, *Koç University*
Assumed full responsibility on interactive lectures and lab sessions on important scientific discoveries, which involved designing class lectures and assignments, teaching, and grading.
- Fall 2012 **Undergraduate Organic Chemistry Lab**, *Koç University*
Developed weekly quizzes; graded quizzes and lab reports; and taught lab sessions.

Skills

■ Languages

English (*Fluent*)
French (*Advanced*)
Turkish (*Native*)

■ Computer Skills

Molecular Dynamics Simulations: Gromacs, NAMD, AMBER, Gaussian
Molecular Docking: ICM, Schrödinger Suite (Glide)
Molecular Modeling and Visualization: PyMOL, VMD, Schrödinger Suite (Maestro)
Computing: Python, Matlab, Tcl, Shell, LaTeX
Computer Graphics: Illustrator, Photoshop

