# Elucidating the Mechanism of Lipid Exchange by Cholesteryl Ester Transfer Protein (CETP)

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

**DOCTOR OF PHILOSOPHY**

BY

SUKRITI SACHER (PHD20201)

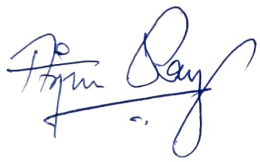
COMPUTATIONAL BIOLOGY

INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY DELHI

NEW DELHI- 110020

# THESIS CERTIFICATE

This is to certify that the thesis titled **"Elucidating the Mechanism of Lipid Exchange by Cholesteryl Ester Transfer Protein (CETP)"**, submitted by **Sukriti Sacher**, to the Indraprastha Institute of Information Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is an original research work carried out by her under my supervision. In my opinion, the thesis meets the standards required by the regulations for the degree. The results contained in this thesis have not been submitted in part or in full to any other university or institute for the award of any degree/diploma.



**Dr. Arjun Ray**
Thesis Supervisor
Assistant Professor
Dept. of Computational Biology
IIIT Delhi, 110020

Place: New Delhi

Date: November 14, 2025

# ACKNOWLEDGEMENTS

**"If a thing is worth doing, it is worth doing well. If it is worth having, it is worth waiting for. If it is worth attaining, it is worth fighting for. If it is worth experiencing, it is worth putting aside time for."~ Oscar Wilde**

This doctoral journey has been the truest definition of Wilde's sentiment. It was an undertaking worth doing well, demanding the best of my intellectual energy and focus. It was certainly worth the wait, as it tested the boundaries of my patience and perseverance over many years. And, ultimately, it was a goal worth fighting for, one that required a resilience I didn't know I possessed until the process demanded it. The experience itself has proven to be the most invaluable reward. Beyond the research and the final findings, pursuing this degree has taught me countless lessons in discipline, humility, and the quiet power of simply showing up every day.

Now, as I look back at all the moments—the late nights, the unexpected breakthroughs, and the critical feedback—that have culminated in this final moment, I am filled with profound gratitude. This achievement was not forged in isolation. It stands as a testament to the belief, guidance, and endless support of the many remarkable people who showed up, guided me, and helped illuminate the path along the way.

I owe my deepest gratitude to my supervisor, **Dr. Arjun Ray**. Your belief in my potential, especially during moments when I didn't believe in myself, was the most essential catalyst for this work. You didn't just provide direction; you patiently taught me everything from scratch, always guiding me toward independence so that I could ultimately master the process on my own. The strategic challenges and immense opportunities you provided throughout these years have helped shape my skills and confidence immeasurably. Above all, I remain continually inspired by your intellectual spark—it is the innovative way you look at proteins and draw connections that kept the fire of my own curiosity burning brightly. I sincerely hope to carry forward the high standards and dedication you've instilled in me.

Beyond my primary mentor, this thesis owes its final form to the expertise and generosity of several key individuals. I would like to thank my thesis committee members, **Dr. Ganesh Bagler**, **Dr. Jaspreet Kaur Dhanjal**, for their constant encouragement and insightful suggestions, which were instrumental in shaping this work to its final iteration. I would also like to thank **Dr. Gaurav Ahuja**, **Dr. Debarka Sengupta**, **Dr. N. Arul Murugan**, and **Dr. Tarini Shankar Ghosh** whose valuable suggestions and advice shaped this work. My research was further enriched by invaluable collaborative relationships. To **Dr. Kausik Chakraborty**, who also served on my SRF committee: thank you for fundamentally shifting my approach. You taught me how to truly think and design experiments to address core biophysics questions, and your suggestions led to a pivotal change in direction that made this thesis exponentially stronger. I am deeply indebted to **Dr. Shantanu Sengupta**, who taught me the twin virtues of resilience and perfection. Your meticulous standards and critical insights shaped the entire final body of work. A special note of thanks goes to the team at Lp2M/Université Côte d'Azur—**Prof. Laurent Counillon's** team, and especially **Dr. Mallorie Poet**, who not only mentored me through eight months of complex experimental work but also took care of me in a foreign country, providing the necessary encouragement to keep going, despite numerous failed Western blots. Finally, to **Dr. Abhishek Mukherjee**, who has long been a senior, mentor, and friend before becoming a collaborator: thank you for devoting the time to help me troubleshoot my experiments and, most importantly, for teaching me the crucial art of scientific writing.

The daily journey within the lab was not only productive but genuinely joyful, thanks to my colleagues. I must first thank **Gayatri**, who, for a significant time, was the only other person in the lab. Our enduring friendship was forged over everything from learning how to log in remotely via terminal to tackling complex MD simulation and machine learning concepts. Whether it was the delicious food and even more intriguing discussions shared at lunch, or the lively house parties at my place, your contribution to my sanity and success was monumental. I also extend my gratitude to **Parth**, **Prutyay**, **Mrinal**, and **Atul**, who interned with us during my first year. You guys effectively taught me everything I now know about programming and bash. Finally, to my current lab mates **Arvinpreet**, **Sadiyah**, **Riddhi**, **Bhumika**, **Diksha**, **Khabeer**, **Deeksha**, **Bharat**: thank you for the countless parties we've shared, both official and sneakily planned. I appreciate all the teasing, the running jokes, the important post-8 PM discussions, and

the constant readiness to help book my Uber rides from time to time. That last Uber black ride may not be so far away now.

The emotional backbone of this entire endeavor was provided by my incredible circle of batchmates and friends. To my fellow PhD batchmates, **Vishakha**, **Sanjay**, **Mansi**, and **Pradeep**: thank you for making the enduring and often difficult nature of the process more manageable. Sharing this intense experience with you created a bond of understanding and mutual support that I relied on constantly. To **Ridam** and **Vishakha**, my house parties are incomplete without you two. A heartfelt thank you to **Sakshi**, **Shubhadeep**, **Arushi**, **Sonam**, **Saveena**, **Shiva**, **Suvendu**, who were all my adopted juniors. A shoutout to friends I made abroad, **Baibhavi**, **Jeanne**, **Hatice**, **Marta**, **Victor**, **Alexis**, **Gisele** and **Gaelle**. I couldn't have survived Nice without you. To **Ankita**, **Sanya** and **Uma**, who patiently listened to me whine and cry through the inevitable tough stretches—your empathy and unwavering presence served as my most critical emotional outlet. My deepest gratitude also extends to the wider group of friends, **Indrani**, **Ratul**, **Sunil**, **Uzma**, and **Ansh**, with whom I have shared many beautiful memories. These moments of normalcy and pure joy provided the necessary breaks that kept my spirit whole. And finally, to the most reliable stress reliever of all, **Muffin**. Your unconditional love, boundless energy, and simple, perfect ability to help me forget the bad days were truly essential. Every wag of your tail was a reminder of life's simple joys.

I would also like to acknowledge the administrative staff at IIIT-D, specifically **Mrs. Priti Patel**, **Mrs. Shipra Jain**, **Mrs. Anshu Dureja**, and **Mr. Raju Biswas**, for always being available to address my queries and resolve my academic issues promptly. My sincere gratitude also goes to **Mr. Imran Khan**, **Mrs. Sarika**, **Mr. Mohit**, and **Mr. Kapil Dev Garg**. I would also like to thank the IIITD IT staff, especially **Mr. Adarsh**, for always being there when I needed his support or technical expertise. I am deeply grateful to all the people I have met during this journey who have shaped my path and inspired me in countless ways.

Finally, and most importantly, I would like to thank my family. To my **Mom**, thank you for always giving me the best advice; sometimes I didn't even realize I was wrong until I talked to you, a testament to your quiet wisdom. To my **Dad**, thank you for always keeping my spirits up, for motivating and inspiring me, and for making sure I would never back down from a fight. To my **Sister**, for her wisdom and comforting words that,

arriving at just the right time, helped me see the light at the end of the tunnel. And to my Brother , for all his care and concern for my well-being, for listening to me patiently, always welcoming me back home with a warm hug, and for just being there. This degree belongs to us all.

Sukriti Sacher

# ABSTRACT

Cholesteryl Ester Transfer Protein (CETP) exchanges cholesteryl esters (CEs) and triglycerides (TGs) between lipoproteins, modulating their composition. This process regulates plasma CE and TG content and therefore strongly correlates with cardiovascular disease (CVD) risk factors. While CETP is a significant therapeutic target, its mechanism of lipid transfer, particularly the role of its structural components, such as its cavity and associated phospholipids, remains incompletely understood. Identification of molecular determinants that can influence CETP activity is paramount for designing and evaluating next-generation therapeutics. Therefore, in this work, we have employed advanced molecular dynamics simulations to track the movement of lipids through CETP, highlighting the regulatory interactions that influence their lipid transfer efficiency and specificity.

CETP encloses a tunnel that serves as a medium for lipid exchange; the movement of lipids can only be understood post identification and characterization of this tunnel. Although several methods exist for protein cavity identification, they are generally limited by the type of cavity they can identify, as well as the automation and resolution of their cavity detection. We addressed these limitations of existing methods by developing CICLOP (Characterization of Inner Cavity Lining Of Proteins), a new method utilizing a hybrid grid- and tessellation-based approach to identify internal cavities (tunnels, channels, pores, and voids). CICLOP offers superior performance and automation, accurately identifying the protein cavity and its functional characterization (diameter, volume, hydrophobicity, charge, and conservation). Applying CICLOP, we identified a single, continuous hydrophobic tunnel within CETP that contradicted earlier models that hypothesized diffused, smaller cavities within CETP.

Next, we determined the lipid transfer mechanism using steered MD simulations. The CETP cavity, in addition to its terminal openings, has two additional openings, which are plugged by two phospholipids (PLs); however, the role of these openings in

CETP structure and function has remained elusive to date. Our structural and functional analyses revealed that lipid traversal through CETP's central tunnel is facilitated through hydrophobic interaction-mediated diffusion. Moreover, the tunnel's function is dictated by its dynamic plasticity that allows lipid movement through a peristaltic wave-like motion. Further, the PLs are indispensable in establishing the optimal architecture of the CETP tunnel and accelerating lipid traversal through a novel "gliding" mechanism. Using free energy calculations and in vitro mutagenesis, an in-depth understanding of the mechanism of lipid exchange by CETP, guided and accentuated by its interaction with PLs, was obtained.

To further understand the complete lipid transfer mechanism, we investigated both the global process and lipid-specific factors governing movement. Our free energy calculations demonstrated the non-specificity of the CETP termini for lipid entry, indicating that this process is primarily governed by the lipid's surface availability, rather than any inherent protein bias for CE or TG. Furthermore, both lipid types follow a single, conserved physical path once inside the tunnel. Importantly, the CETP tunnel can accommodate two lipids simultaneously moving in opposite directions, providing compelling. The dynamics of lipid traversal are significantly influenced by lipid-specific factors such as acyl chain length and conformation. Long-chain TGs (LCTs) in specific conformations like 'Fork' and 'T' exhibit the longest residence times because they form a greater number of stable hydrophobic contacts with the tunnel residues. This finding is critical in the context of cardiovascular disease (CVD) where the slower transfer kinetics of LCTs—which are prevalent in the plasma of CVD patients— may lead to adverse, pro-atherogenic lipoprotein remodeling outcomes.

Altogether, this study advances our understanding of CETP by revealing that its function relies on a precisely orchestrated interplay of tunnel plasticity and optimal hydrophobicity, which allows lipid entry and diffusion through the tunnel. The study also identified PL-plugs as essential co-factors that strongly influenced CETP function, with their ability to modulate both the tunnel hydrophobicity and lipid transfer dynamics. Lipid-specific factors, such as acyl chain length, which are directly influenced by diet, also influence lipid transfer dynamics. The study provides compelling evidence in favor of a ternary complex, wherein CETP bridges two lipoproteins, simultaneously facilitating the concurrent exchange of CE for TG. These insights open new avenues for designing and evaluating next-generation CETP-targeted drugs by focusing on the

solvent-accessible PL-binding pockets or the integrity of the tunnel, providing a refined approach for modulating CETP function in patients with atherosclerosis.

KEYWORDS: CETP ; Lipoproteins ; TG ; CE ; Lipid Exchange ; MD simulation ; Umbrella Sampling

# LIST OF PUBLICATIONS

## 0.1 Publications From Thesis

- **Sacher, S.**, Singh, P., Mukherjee, A., Bhaskar, A.K., Chakraborty, K., Counillon, L., Sengupta, S., Poet, M., and Ray, A. Phospholipids that plug the pores of cholesteryl ester transfer protein control its lipid transfer dynamics. *Biorixiv*, https://doi.org/10.1101/2025.09.04.674379 **(First author), Under Review**
- **Sacher, S.** and Ray, A. Solvation Dynamics of a Single Triglyceride as a Function of its Chain Length. *Biorixiv*, https://doi.org/10.1101/2022.03.28.486057 **(First author), Under Review**
- **Sacher, S.**, Mukherjee, A., and Ray, A. Deciphering structural aspects of reverse cholesterol transport: mapping the knowns and unknowns. *Biological Reviews Cambridge Philosophical Society*, 98(4), 1160–1183, (2023) **(First author)**
- Garg, P., **Sacher, S.**, Mrinal, Atul, Gautaum, P. and Ray, A. CICLOP: a robust and accurate computational framework for protein inner cavity detection. *Bioinformatics*, 38(8), 2153–2161, (2022) **(Co-first author)**

## 0.2 Others

- Sain, S., **Sacher, S.**, Solanki, B., and Kumar, N. Binding Dynamics of Helicobacter pylori CagF-CagV: Biochemical, Computational, and Structural Insights into the Inner Membrane Core Complex of Cag-Type 4 Secretion System. *Journal of Pure Applied Microbiology*, 19(3), 2042–2053, (2025) **(Co-author)**
- Francis, A.X., Chilkoti, M., Atul, Mrinal, **Sacher, S.**, and Ray, A. An Algorithm-Based Investigation to Reveal the Differential Dynamics of Water inside Protein Cavities as a Function of Distance from Their Wall. *Journal of Chemical Information and Modelling*, 65(10), (2025) **(Co-corresponding author)**
- Sunidhi, S., **Sacher, S.**, Atul, Garg, P., and Ray, A. Elucidating the Structural Features of ABCA1 in its Heterogeneous Membrane Environment. *Frontiers in Molecular Biosciences*, 8, (2021) **(Co-first author)**

## 0.3 Book Chapters

- **Sacher, S.** and Ray, A. In Silico strategies for characterizing inner cavities of lipid-binding proteins. In: Drin, G. (eds) Intracellular Lipid Transport. Methods

in Molecular Biology, vol 2888, 978—1—0716—4318—1, Springer (2025)

- **Sacher, S.**, Afroz, S. Kaur, A. and Ray, A. Probing lipid-protein interactions through molecular dynamics simulation. In: Bhattacharya, S.K. (eds) Lipid Analysis in Biology, Health, and Disease, 978-0-443-19116-9, Elsevier (2025)

## 0.4 Presentations in Conferences

- Presentation titled *Phospholipid bound to central pores of cholesteryl ester transfer protein (CETP) aids in the transfer of lipid through CETP tunnel* at the **European Atherosclerosis Society Meeting 2024, Lyon, France**, May 26-29, 2024
- Presentation titled *Decoding the tripartite relation between CETP and its bound phospholipids with triglycerides* at the **Cardiovascular Research Convergence 2023, New Delhi, India**, Oct 28, 2023
- Poster titled *Solvation dynamics of a single triglyceride as a function of its chain length* at the **Advances in Cardiovascular Medicine and Research 2023, Chandigarh, India**, Feb 16-18, 2023

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABBREVIATIONS

| | |
|---|---|
| **ACF** | Auto-Correlation Function |
| **ApoA1** | Apolipoprotein A1 |
| **ApoC1** | Apolipoprotein C1 |
| **ApoE** | Apolipoprotein E |
| **ASCVD** | Atherosclerotic Cardiovascular Disease |
| **BPI** | Bactericidal Permeability Increasing Protein |
| **CE** | Cholesteryl ester |
| **CETP** | Cholesteryl Ester Transfer Protein |
| **CHARMM** | Chemistry at HARvard Macromolecular Mechanics |
| **CI** | Confidence Interval |
| **CICLOP** | Characterization of Internal Cavities of Proteins |
| **COM** | Center of Mass |
| **CONAN** | CONtact ANalysis |
| **CVD** | Cardiovascular Disease |
| **DOPC** | 2-dioleoyl-sn-glycero-3-phosphocholine |
| **EM** | Electron Microscopy |
| **FFA** | Free Fatty Acids |
| **FN** | False Negatives |
| **FP** | False Positives |
| **FPR** | False Positive Rate |
| **GROMACS** | GROningen MAchine for Chemical Simulations |
| **HDL** | High-density lipoprotein |
| **HL** | Hepatic Lipase |
| **HPeV** | Human parechovirus Envelope |
| **IDL** | Intermediate-density lipoprotein |
| **LBP** | Lipopolysaccharide Binding Protein |
| **LCTs** | Long Chain TGs |
| **LD** | Lipid Droplets |

| | |
|---|---|
| **LDL** | Low-density lipoprotein |
| **LINCS** | LINear Constraint Solver |
| **LPL** | Lipoprotein Lipase |
| **LTP** | Lipid Transfer Proteins |
| **MD** | Molecular Dynamics |
| **MCTs** | Medium Chain TGs |
| **NEFA** | Non-Esterified Fatty Acids |
| **OPLS-AA** | Optimized Potentials for Liquid Simulations—All-Atom |
| **PBC** | Periodic Boundary Conditions |
| **PCA** | Principal Component Analysis |
| **PDB** | Protein Data Bank |
| **PL** | Phospholipids |
| **PLTP** | Phospholipid Transfer Protein |
| **PLUNC** | Palate Lung and Nasal epithelium Clone |
| **PMF** | Potential of Mean Force |
| **PME** | Particle Mesh Ewald |
| **POPC** | 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine |
| **PSI-Blast** | Position-Specific Iterated BLAST |
| **RCT** | Reverse Cholesterol Transport |
| **rHDL** | Recombinant High Density Lipoprotein |
| **RMSD** | Root Mean Square Deviation |
| **SARS-CoV-2** | Severe Acute Respiratory Syndrome Coronavirus 2 |
| **SCARB1** | Scavenger Receptor Class B Type 1 |
| **SCTs** | Short Chain TGs |
| **SD** | Standardad Deviation |
| **SMD** | Steered Molecular Dynamics |
| **SPC** | Simple Point Charge |
| **T2D** | Type 2 Diabetes |
| **TG** | Triglyceride |
| **THQ** | tetrahydroquinoline |
| **TIP3P** | Transferable Intermolecular Potential with 3 points |
| **TN** | True Negatives |
| **TP** | True Positives |

| | |
|---|---|
| **TRL** | Triglyceride-Rich Lipoproteins |
| **VLDL** | very low-density lipoprotein |
| **WHAM** | Weighted Histogram Analysis Method |

# CHAPTER 1

# INTRODUCTION

Cardiovascular diseases (CVDs) encompass a broad range of heart and blood vessel conditions, including coronary heart disease, stroke, and other related issues. These diseases are a significant health concern, contributing significantly to mortality and morbidity worldwide (Cesare Di *et al.*, 2024). Influenced by a combination of behavioral, environmental, and socio-economic factors, along with an individual's genetic makeup, CVDs are often considered complex in their treatment and prognosis (Bhatnagar, 2017). At the molecular level, dysregulation of cholesterol and triglycerides (TG), either due to their imbalanced cellular synthesis, degradation, oxidation, or due to their abnormal plasma concentrations within lipoproteins, is a primary factor responsible for CVDs (Wazir *et al.*, 2023). The plasma levels of cholesterol-carrying lipoproteins (high-density, low-density, and very low-density lipoproteins, HDL, LDL, and VLDL) regulate the activity of cholesteryl ester transfer protein (CETP). CETP exchanges cholesteryl esters and TGs between lipoproteins, changing their composition and affecting their half-life. It is secreted in response to elevated plasma TG and cholesterol.

## 1.1 Triglyceride and Cholesterol Metabolism

The plasma life cycle of these lipids begins with the dietary uptake of TG that occurs in the intestine, where it is packed into chylomicrons. The TG in the core of chylomicrons is hydrolysed into free fatty acids and glycerol by lipoprotein lipase, forming chylomicron remnants, which are delivered to the liver through an LDL-like receptor protein (Fig. 1.1). A delayed removal of chylomicron remnants is associated with atherogenesis (Kwiterovich, 2000). Subsequently, TG-rich VLDLs are synthesized and secreted from the liver to deliver TGs and free fatty acids to peripheral tissues. While in plasma, VLDL is similarly acted upon by lipoprotein lipase to produce intermediate density lipoprotein (IDL), which is further hydrolysed into LDL by hepatic lipase. The contents of LDL are similarly adsorbed through the LDL receptor, which is ubiquitously expressed (Fig. 1.1).

An elevated VLDL leads to hypertriglyceridemia, and overproduction of its downstream remnants IDL and LDL, all of which are hallmarks of CVDs (Packard *et al.*, 2020). Furthermore, oxidation of LDL initiates the inflammatory process at the heart of atherogenesis (Khatana *et al.*, 2020). Several dyslipoproteinemias have been described as the result of defects in the endogenous VLDL-IDL-LDL metabolic pathway, referred to as LDL metabolic syndrome (Fan *et al.*, 2019). These are often accompanied by hyperinsulinism, glucose intolerance, hypertension, and low HDL-cholesterol, further linking them to other lipid metabolism-related disorders (Kwiterovich, 2000; Fan *et al.*, 2019).

In the case of cholesterol, de novo synthesis accounts for most of the plasma pools of cholesterol in lipoproteins, unlike TGs, which are primarily absorbed from the diet. (Jones *et al.*, 1996). This cholesterol is effluxed out of the cell through ATP-binding cassette transporter A1 and packaged into HDL. Subsequently, cholesterol in HDL is acted upon by lecithin-cholesterol acyltransferase to produce cholesteryl esters (CEs), forming a spherical HDL that readily circulates in the plasma (Sacher *et al.*, 2023). HDL is involved in reducing atherosclerotic risk by removing cholesterol from macrophages in the artery wall (Rye *et al.*, 2009). It also inhibits LDL oxidation (Negre-Salvayre *et al.*, 2006), improves endothelial function (Mineo *et al.*, 2006) and has anti-thrombotic and anti-inflammatory properties (Murphy *et al.*, 2008).

During dyslipidemia, an increased level of free fatty acids in plasma leads to an increased flux of free fatty acids to the liver. Similarly, the inability of adipose tissue to process and store excess fats as TGs also has the same fate. This leads to an elevated production and secretion of VLDL (Kwiterovich, 2000). Similarly, the excess of cholesterol that is fluxed back to the liver is metabolized and excreted into bile (Sacher *et al.*, 2023). Elevated levels of plasma TG and cholesterol further result in the release of CETP and hepatic lipase (Kwiterovich, 2000; Salekeen *et al.*, 2022). CETP activity is pro-atherogenic in nature.

## 1.2 CETP: An Important Mediator of Lipid Metabolism

As the TG-enriched VLDL is secreted by the liver at an accelerated rate, the TGs in VLDL are exchanged for CE in the core of LDL, producing CE-depleted, but TG-enriched LDL particles through CETP-mediated action (Fig. 1.1). These particles are

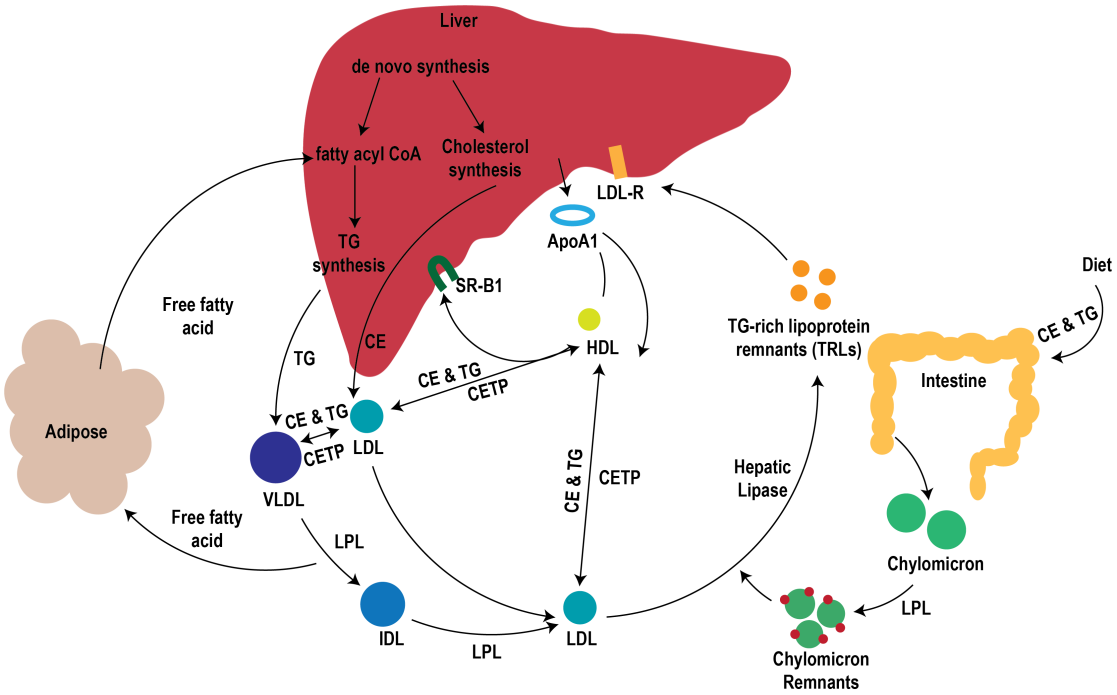



**Figure 1.1. Metabolism of cholesterol and triglycerides.** Fatty acids (FFA) stored in adipose tissue, or those obtained from diet enter the liver and together with FFA derived from de novo lipogenesis are used for the synthesis of TG and other complex lipids (not shown). The TGs are stored in part either within lipid droplets or used for the assembly of VLDL, which are then released into the circulation. Here, VLDL, together with intestinal-derived chylomicrons (CM), are processed in the capillaries of adipose tissues, the heart, and skeletal muscle by lipoprotein lipase (LPL). LPL hydrolyzes TGs to generate FFA that are taken up by the cells of the respective organ for storage or energy generation. Through LPL-dependent triglyceride hydrolysis, TRLs become smaller and are thus transformed to cholesterol-enriched TRL remnants that are eventually taken up by hepatocytes through endocytosis and digested in the endo-lysosomal compartment. Part of the VLDL remnants are further processed by other intravascular enzymes including hepatic lipase (HL) and phospholipid transfer protein (PLTP) to become LDL. Similarly, de novo cholesterol biosynthesis results in the release of cholesterol-rich LDL into the circulation from the liver. Furthermore, HDL particles are generated from lipid-poor apolipoprotein A-I (APOA1) secreted by the liver. Both LDL and HDL exchange cholesterol ester for triglyceride with TRL/VLDL in a reaction catalyzed by CETP.

then rapidly hydrolyzed by hepatic lipase to produce small dense LDL particles (sdLDL) (Kwiterovich, 2000). These sdLDL have low affinity for LDL receptors and are slowly cleared from the plasma, increasing their likelihood of being deposited into the arterial wall through oxidation (Martin *et al.*, 1993). Therefore, CETP-mediated TG exchange between VLDL and LDL has a strong pro-atherogenic potential. Alternatively, due to the non-specific nature of CETP, the CEs in the core of HDL are also exchanged for TG in VLDL, producing a TG-enriched but CE-depleted HDL (Fig. 1.1). These TG-rich HDL are catabolized more rapidly by the kidney, reducing plasma HDL (Ivanova *et al.*, 2017). Ultimately, the atherogenic potential of CETP activity may indeed depend on the metabolic context (Barter, 2000; Harder *et al.*, 2007). When CETP deficiency does not result in an increase in HDL levels, the outcome tends to be pro-atherogenic. Conversely, if CETP deficiency results in a significant rise in HDL levels, the outcomes appear to be anti-atherogenic (Ivanova *et al.*, 2017; Barter, 2000).

CETP belongs to the bacterial permeability increasing protein/lipopolysaccharide binding protein family (BPI/LBP), and it exhibits anti-inflammatory functions similar to other family members, regardless of its influence on lipoproteins. It has been demonstrated that CETP in macrophages and the liver inhibits the interaction of lipopolysaccharide with Toll-like receptor 4, which in turn reduces the inflammatory response (Parini and Rudel, 2003). Further, CETP expression lowers the production of mitochondrial superoxide anions and the release of $H_2O_2$, reducing the expression of pro-inflammatory genes and phagocytic activity (Venancio *et al.*, 2016). While an exhaustive analysis of CETP expression or inhibition and its physiological effects is beyond the scope of this thesis, it is evident that more work is needed to understand how CETP influences the concentration, composition, metabolism, and function of lipoproteins.

Since CETP represents a focal point in regulating lipoprotein composition and their plasma turnover, CETP has been actively targeted for CVD relief. Notably, its pro-atherogenic effect due to increased HDL clearance and increased sLDL production in conjunction with hepatic lipase activity has favored its inhibition, showing promising CVD outcomes. Results from clinical trials of CETP inhibitors have demonstrated that, compared to placebo, torcetrapib, evacetrapib, anacetrapib, and obicetrapib reduced LDL and apoB levels while increasing HDL levels. The randomized assessment of anacetrapib in a trial focused on lipid modification and cardiovascular outcomes, where it was administered alongside a high-intensity statin, indicated a 17% decrease in LDL

levels and a 9% reduction in major adverse cardiovascular events [rate ratio, 0.91; 95% confidence interval (CI), 0.85, 0.97; P = 0.004] (Bowman *et al.*, 2017). Despite this, the manufacturer chose not to seek regulatory approval because anacetrapib had a prolonged half-life and tended to accumulate in adipose tissue. Similarly, previous CETP inhibitors did not secure market approval due to various compound-specific off-target effects. Recently, obicetrapib, a next-generation polar CETP inhibitor, has demonstrated potential regarding its effectiveness and is presently in Phase III trials (Dorighello *et al.*, 2022; Davidson *et al.*, 2024). This highlights the potential for advancements in the design and effectiveness of CETP inhibitors.

## 1.3 Research Gaps

Considering that CETP lies at the crossroads of lipoprotein metabolism, which may lead to atherogenesis, it is crucial to understand its mechanism of action and how it contributes to metabolic balance in the context of CVDs. To achieve this, a comprehensive biophysical characterization of the protein, including its structural components, is essential. However, studying lipid transfer through the CETP tunnel remains challenging, as it cannot be directly visualized through experimental biophysical approaches. CETP possesses a central tunnel flanked by two $\beta$-barrel domains, N- & C- terminal barrels connected by a neck region (Qiu *et al.*, 2007; Liu *et al.*, 2012). The protein harbors a tunnel 60-Å long, which is the longest reported for any lipid transfer protein. The tunnel is highly hydrophobic and contains only two bound water molecules. The size and hydrophobicity of the tunnel suggest its inability to discriminate between different neutral lipids, implying that different combinations of neutral lipids may be accommodated inside the tunnel (Qiu *et al.*, 2007). Further, the protein contains a series of small, isolated hydrophobic cavities (diameter >5 Å) along the central axis of both $\beta$-barrels that are hypothesized to connect due to conformational changes, forming a single continuous hydrophobic tunnel  105 Å in length (Zhang *et al.*, 2012). The exact mechanism by which this tunnel facilitates lipid transfer is unknown, specifically whether it functions as a passive pore or actively directs and regulates lipid entry and exit.

Further, the central tunnel is occluded at the neck region by two phospholipid (PL) molecules—typically 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) embedded such

that their hydrophobic tails are sequestered within the tunnel and their polar head groups remain solvent-exposed (Qiu *et al.*, 2007). While early hypotheses posited that these PLs participate in dynamic exchange during lipid transport (Liu *et al.*, 2012), mounting evidence suggests that they serve as structural elements, maintaining tunnel integrity by shielding it from aqueous exposure (Koivuniemi *et al.*, 2012). Additionally, bacterial permeability-increasing protein (BPI), a structurally homologous protein to CETP, also harbors PLs within its concave tunnel with a similar orientation (Beamer *et al.*, 1997), implying a conserved architectural role. These bound PLs influence the conformational dynamics of CETP; however, their functional relevance to lipid transfer remains unresolved.

Because of the protein's characteristic boomerang form, it was previously believed that CETP adjusts its curvature to bind to lipoproteins, with the bent form interacting with the smaller HDL and a straight form interacting with the much bigger VLDL (Qiu *et al.*, 2007). Lipids entered CETP through the protein's central pores at the floor, which were closed by PLs as the protein detached (Qiu *et al.*, 2007) (Fig. 1.3 A). CETP mutants that changed the shape of these lipid-binding pockets were found to affect CETP activity (Qiu *et al.*, 2007). As a result, CETP acted as a lipoprotein shuttle, exchanging neutral lipids along the way. However, negative staining electron microscopy of CETP-lipoprotein complexes revealed barrels breaching the lipoprotein bilayer (Zhang *et al.*, 2012, 2015) (Fig. 1.2). Despite the low resolution of these images, the N-terminal barrel was shown to be largely connected with HDL, whilst the C-terminal barrel interacted with VLDL (Zhang *et al.*, 2012). Furthermore, ternary complexes including two lipoproteins connected by CETP have been seen, implying a tunnel mechanism wherein the lipid entry occurred through the open termini and was governed by the concentration gradients across lipoprotein surfaces (Zhang *et al.*, 2015) (Fig. 1.2).

Despite the importance of the initial lipoprotein-sensing phase for drug development, irrespective of the shuttle or tunnel mode of transfer, the full regulatory process involved in CETP-mediated lipoprotein remodelling remains largely unknown. While its immobilization on HDL has been linked to POPC, ApoA-I, and E2 fatty acyl esters (Cilpa-Karhu *et al.*, 2015; Morton and Greene, 2003*a*), the crucial CETP-LDL/VLDL interaction is unexplored. Similarly, the factors controlling CETP detachment from lipoproteins remain unresolved. Lipid entry into CETP tunnel has been studied through molecular modelling of CETP-HDL complexes, which show that cholesteryl ester can diffuse from

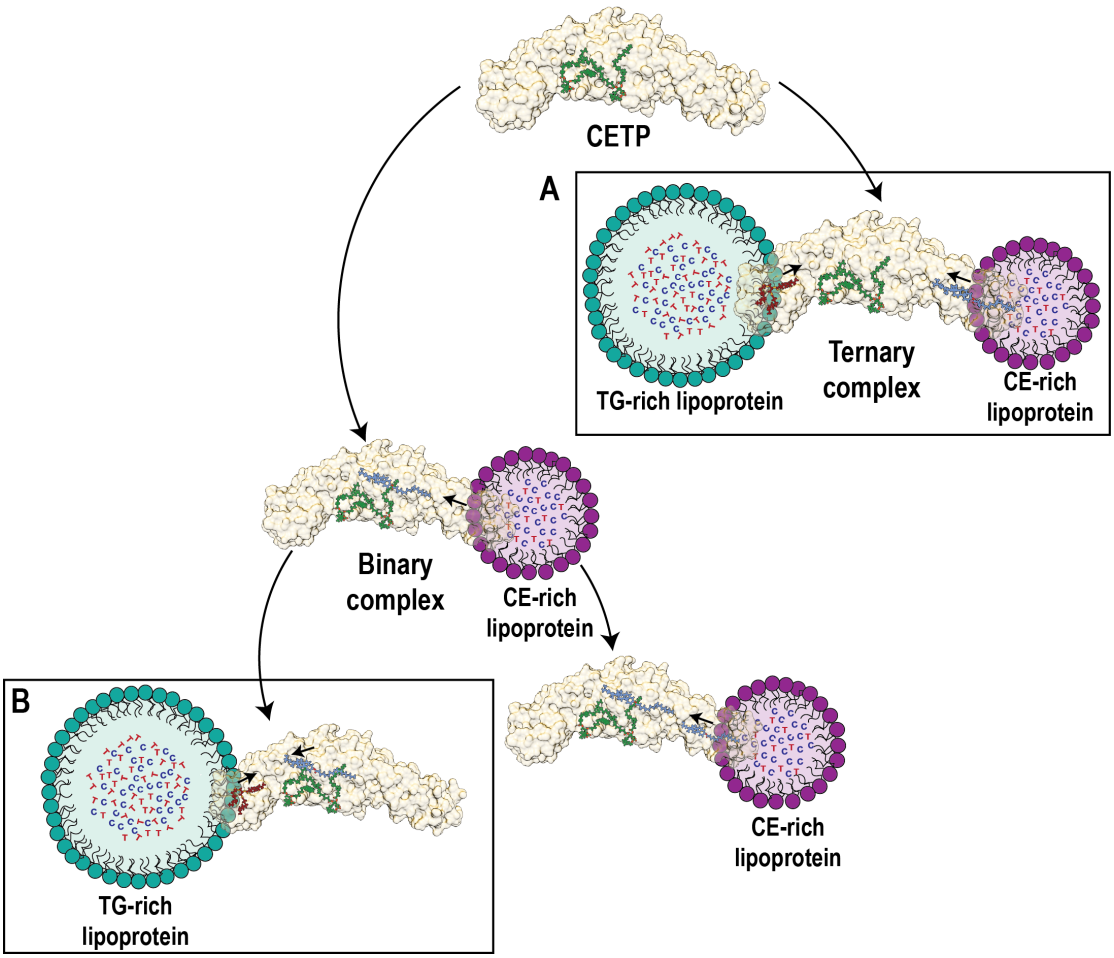



**Figure 1.2. Schematic representation of neutral lipid exchange by CETP. A.** Tunnel mechanism of lipid transfer showing CETP forming a bridge between two lipoproteins. **B.** Shuttle mechanism of lipid transfer showing CETP shuttling between a TG-rich and CE-rich lipoprotein

the core of HDL into the CETP tunnel (Cilpa-Karhu *et al.*, 2015) (Fig. 1.3 B). Similarly, a subsequent steered MD (SMD) study employing a minimal ternary CETP complex embedded in lipid monolayers demonstrated full CE translocation while identifying energy barriers (Fig. 1.3 C). However, the movement of triglycerides through the tunnel remains largely unexplored. Furthermore, molecular determinants such as residues that facilitate or inhibit lipid entry through the terminal openings, or the stoichiometry of exchange—that is, whether lipids are exchanged one at a time or concurrently—have never been explored.

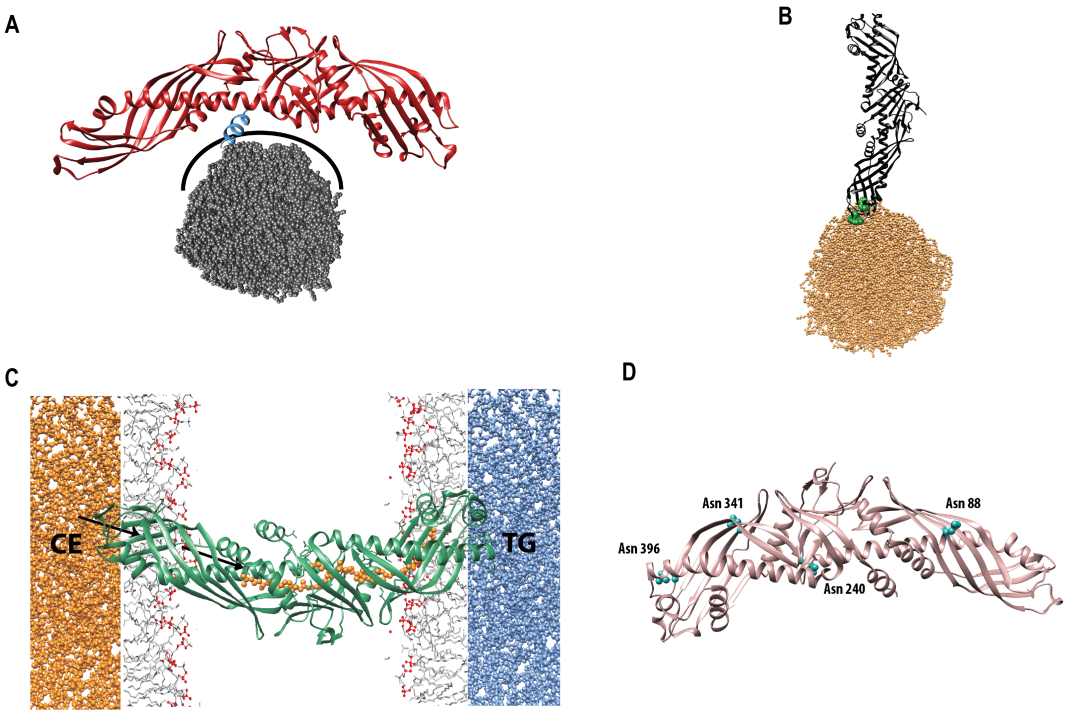



**Figure 1.3. MD simulation systems used to study lipoprotein re-modelling through CETP. A**
CETP was postulated to bind HDL through its concave surface. CETP (red) was simulated such that its concave surface faced the HDL droplet (grey). Helix X (blue) was found to act as a lid, regulating accessibility to the C-terminal tunnel opening of CETP. **B** Simulation system showing the Trp residues (green) in the N-terminal barrel domain of CETP (black) as the sites facilitating HDL interaction. **C** simulated system for the ternary complex formed by HDL-CETP-LDL association showcasing the tunnel mechanism of CE transfer from HDL through CETP via steered molecular dynamics. **D** Glycation sites (blue) of CETP (pink) speculated to impact its structure and dynamics of CE transfer.

## 1.4 Thesis Scope: Understanding the Molecular Determinants for Lipid Transfer Through CETP

### 1.4.1 Objective of the Thesis

This thesis aims to advance our understanding of CETP-mediated lipid exchange, a key player in lipoprotein remodeling. Specifically, it uses extensive classical and steered molecular dynamics to identify important structural and molecular components that influence CETP activity. It also uses umbrella sampling for free energy calculations to highlight the thermodynamic costs of lipid traversal through CETP. Furthermore, it establishes a proof-of-concept for the simultaneous heteroexchange of lipids (TG and CE) to address the transfer stoichiometry. The research is structured around four key objectives:

***Objective 1:*** **To characterize the tunnel within CETP and elucidate its role in lipid transfer.**

***Objective 2:*** **To elucidate the mechanism of triglyceride transfer through CETP and determine the role of structural PL plugs in this process.**

***Objective 3.1:*** **To elucidate additional factors that influence the movement of lipid within the CETP tunnel.**

***Objective 3.2:*** **To investigate the movement of triglyceride within CETP as a function of acyl chain length.**

The investigation is strictly confined to the molecular features that govern the movement of lipid through CETP, specifically addressing the following structural and mechanistic hypotheses:

***Hypothesis 1***: Whether the cavity within CETP is continuous or formed by connecting the several smaller cavities, and whether it functions as a passive pore or actively directs the movement of lipid.

***Hypothesis 2***: Whether the PL plugs are static structural components or serve a dynamic functional role in lipid transfer.

***Hypothesis 3***: Whether lipids can move through the tunnel one at a time or concur-

rently.

## 1.5 Organization of the Thesis

The organization of this thesis is structured to systematically address the outlined research objectives and provide a cumulative mechanistic understanding of CETP-mediated lipid exchange. The subsequent eight chapters move logically from establishing the necessary foundational context and advanced computational methodologies to presenting the novel findings regarding the CETP tunnel characteristics, the functional role of phospholipid, and the specific dynamics of neutral lipid transfer.

*Chapter 1* lays the groundwork, introducing CETP's importance in biology and the functional gaps in our understanding that this thesis aims to address.

*Chapter 2* presents a detailed review of the literature describing the different biochemical, biophysical, and computational approaches that have been used to dissect the mechanism of CETP and to design therapeutic strategies that target CETP.

*Chapter 3* describes the methodology employed in this research, specifically the physical foundation of molecular dynamics simulations, steered molecular dynamics, and umbrella sampling simulations for free energy calculations in detail.

*Chapter 4* discusses the problems with cavity identification algorithms and presents a generalized and automated workflow, CICLOP, for elucidating and characterizing the cavity lining of proteins and, in particular, CETP, addressing Objective 1. The chapter describes the approach used within CICLOP and its extensive benchmarking against existing cavity detection algorithms, demonstrating its effectiveness in identifying and characterizing the physicochemical properties of protein cavities.

*Chapter 5* describes the movement of a single TG inside the CETP tunnel simulated using steered molecular dynamics to address Objective 2. It identifies and highlights several structural features of CETP that facilitate the movement of lipids inside the tunnel. It also highlights the synergy between the PLs plugging the central pores of the CETP tunnel and their impact on the lipid transfer activity of CETP.

*Chapter 6* explores additional factors that influence lipid transfer through the CETP

tunnel, addressing Objective 3.1. Molecular determinants regulating lipid entry through the terminal barrels and concurrent movement of two lipids through CETP tunnel are described.

*Chapter 7* investigates the role of acyl chain length and conformation of TG in governing its ability to enter and traverse the CETP tunnel, addressing objective 3.2.

*Chapter 8* concludes the thesis by summarizing the findings and discussing their implications. It also discusses the limitations of the approaches and proposes future research directions.

# CHAPTER 2

# Background and Related Works

CETP facilitates the transfer of TG from VLDL to HDL and/or LDL in exchange for CE, leading to the formation of larger, relatively TG-rich HDL and LDL particles (Wazir *et al.*, 2023; Kwiterovich, 2000). It is a hydrophobic glycoprotein expressed in several tissues of humans, primates, rabbits, hamsters, reptiles, and fish, while mice, rats, and dogs do not express this protein (Raposo *et al.*, 2021). In humans, CETP is highly expressed in adipose tissue, with higher expression observed in small adipocytes that have a lower lipid content. It is also highly expressed in the liver, where its expression is differentially regulated based on the fasting state (MacLean *et al.*, 2000).

## 2.1 Physiological Importance of CETP and a Disease Perspective

The primary role of CETP is to exchange CE from HDL or LDL with TG from VLDL. As this exchange proceeds, CE-depleted but TG-enriched HDL or LDL particles are generated along with TG-depleted and CE-rich VLDL. CE-rich VLDL is acted upon by hepatic lipase to generate IDL and LDL. While TG-rich LDL particles are rapidly acted upon by hepatic lipase to generate small LDL (sLDL) (Kwiterovich, 2000). sLDL has a low affinity for LDL receptors, which decreases their clearance, significantly increasing their circulation time. This further increases their likelihood of being oxidized, generating an atherogenic phenotype (Martin *et al.*, 1993). Alternatively, a TG-enriched but CE-depleted HDL due to CETP exchange is acted upon by PLTP, leading to the formation of small, dense HDL particles (HDL3). These TG-rich HDL are catabolized more rapidly by the kidney, decreasing the total plasma HDL-C level. This results in an increased flux of cholesterol into RCT; faster removal of HDL leads to generation of HDL through increased efflux of cholesterol out of peripheral tissues (Sacher *et al.*, 2023).

Mutations in CETP are rare in Western populations but have been reported in Japanese populations. Two common mutations of the gene include an intron 14 splicing defect and

D442G, both of which cause marked abnormalities in the concentration, composition, and functions of both HDL and LDL, leading to increased susceptibility to coronary atherosclerosis (Yamashita *et al.*, 2000). Another Taq1 B polymorphism, B2 allele, is associated with low CETP mass, increased HDL-cholesterol, and a decreased risk of CVD (Nagano *et al.*, 2004). Complete or partial CETP deficiency is rare; however, variations in coding or non-coding regions of the CETP gene are pretty common (Weber *et al.*, 2010).

In addition to CETP's direct role in modulating plasma lipoprotein composition, CETP has been shown to influence lipoproteins in the cerebrospinal fluid, majorly affecting ApoE composition of HDL and the brain cholesterol concentration, which is linked with Alzheimer's disease risk (Oestereich *et al.*, 2022; Zak *et al.*, 2002). While the effect of CETP on HDL and the association of ApoE have not been studied directly, CETP inhibition has been shown to decrease the association of ApoE with HDL, thereby improving AD prognosis (Mehta *et al.*, 2023).

Additionally, CETP inhibition ameliorates the risk of Type 2 diabetes (T2D) by modulating reverse cholesterol transport (Metzinger *et al.*, 2020). HDL-mediated cholesterol efflux enhances insulin sensitivity, thereby increasing glucose uptake and reducing plasma glucose levels in type 2 diabetes (T2D) patients (Mehta *et al.*, 2023). Furthermore, CETP has been shown to stimulate lipolysis and thermogenesis, thereby increasing the body's energy expenditure and resulting in a significant reduction in body fat (Raposo *et al.*, 2021).

Despite the debate regarding the biochemical effects of CETP activity, gaining a better understanding of CETP's molecular interactions may provide clarity in defining its function.

## 2.2 Structure of CETP

CETP is a boomerang-shaped protein measuring 135 Å × 30 Å × 35 Å and has a fold similar to BPI (Qiu *et al.*, 2007; Beamer *et al.*, 1997). This fold comprises two $\beta$-barrels joined by a linker consisting of $\beta$-sheets. Each barrel is composed of a highly twisted $\beta$-sheet along with two helices. Upon superimposition on Bactericidal Permeability Increasing Protein (BPI) that belongs to the same family as CETP, it was observed that

the two barrels differ by 25 Å at the tips. This showed that the barrels are capable of movement. The bottom surface in BPI is almost flat, compared to that of CETP, which is concave (Qiu *et al.*, 2007).

CETP is glycosylated at N88 (N-barrel), N240, N341 (neck), and N396 (C-barrel). These glycans were observed to form stabilizing interactions with nearby residues and modulate CETP tunnel integrity (Hao *et al.*, 2022) (Fig. 1.3 D). The structure of CETP has four openings: two at the terminal ends of the two barrels and two at the center of the protein, which are plugged by two dioleoylphosphatidylcholines. These phospholipids are oriented such that they bury their acyl chains inside the tunnel and expose their zwitterionic head groups to solvent. PLs bound to CETP were shown to maintain its bent-untwisted conformation, while their absence triggered conformational flexibility (Revanasiddappa *et al.*, 2018).

The protein encompasses a continuous tunnel, 60-Å-long. This tunnel is occupied by two cholesteryl oleate molecules. The accessible surface area and volume of the CETP tunnel are 2,100 $Å^2$ and 2,560 $Å^3$, respectively. The tunnel is highly hydrophobic and contains only two bound water molecules. The size and hydrophobicity of the tunnel suggest its inability to discriminate between different neutral lipids (Qiu *et al.*, 2007). In the structure, one CE is located between the N-barrel and the central $\beta$-sheet. This CE is surrounded by mostly hydrophobic residues (13C, I15, T138, V198, F263, F265, L457, M459) and only a few polar residues (Q199, S230, H232). H232 forms a $\pi$-stacking interaction with the ester of CE, and the oleoyl tails of CE and N-PL (located between N-barrel and central $\beta$-sheet) stack with F265 and F263, respectively (Fig. 2.1). The N-PL forms a salt bridge with R201. The other CE is surrounded by L425, V428, F429, Y375, M433, I443, while R282 forms a salt-bridge with C-PL (located between C-barrel and the central $\beta$-sheet) (Qiu *et al.*, 2007) (Fig. 2.1).

## 2.3 CETP Activity

CETP mediates equimolar exchange between lipoproteins. Recombinant CETP (rCETP) expressed in HEK293 cells mediated a net flux of TG from VLDL to HDL, and CE from HDL to VLDL, where the total gain of TG in HDL and loss of CE from HDL were equal. The assay, however, measured the endpoint accumulation of radioactively labelled TG

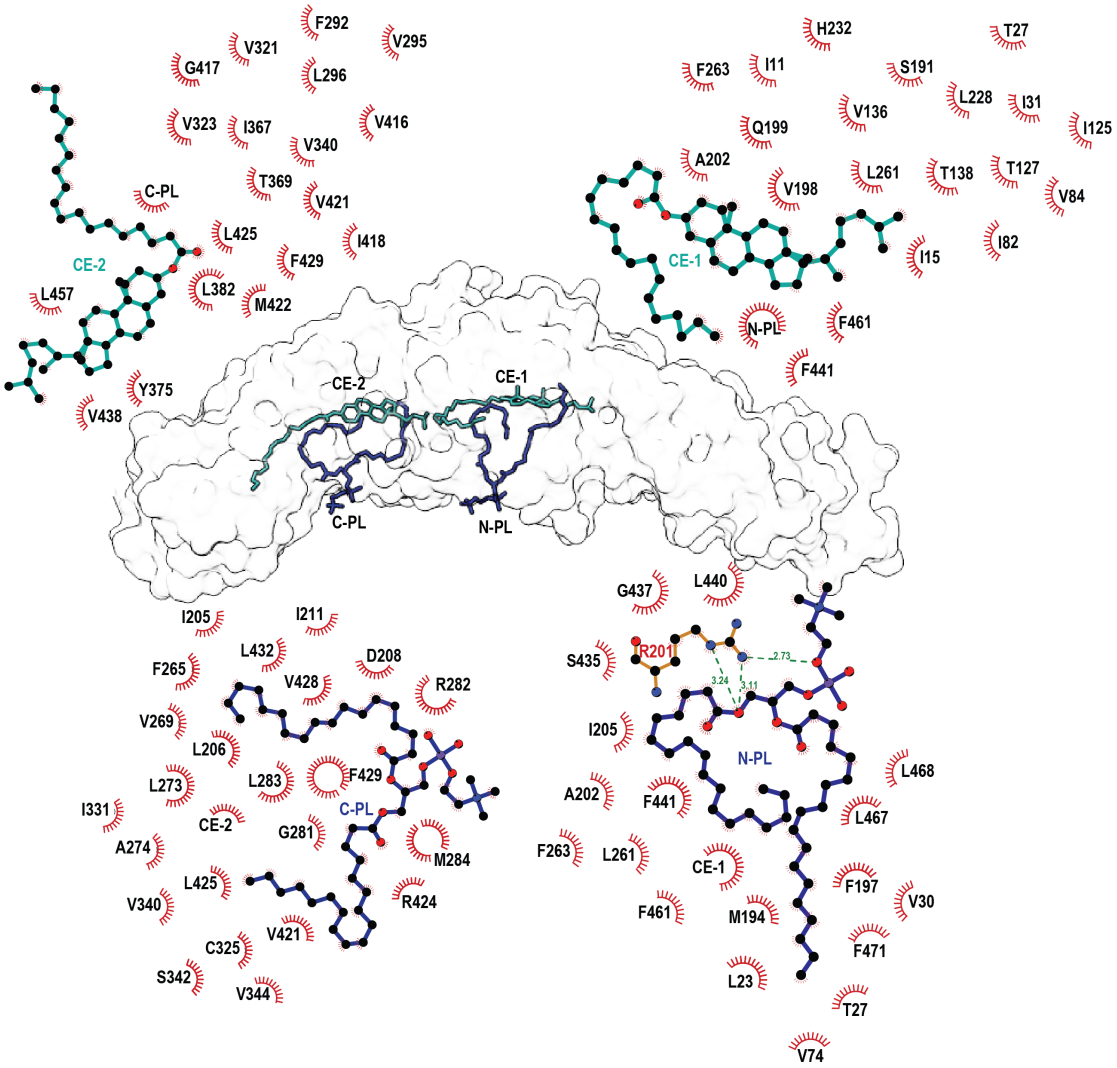



**Figure 2.1. Residues interacting with two cholesteryl oleate and DOPC within the CETP tunnel as shown in PDBID:2OBD**

or CE between two lipoprotein compartments (Morton and Izem, 2014).

CETP is physiologically inhibited by apoC1 when associated with the plasma HDL fraction. This effect is associated with its ability to modulate the electrostatic charge of the HDL surface, possibly affecting CETP-lipoprotein binding. Interestingly, in T2D patients, apoC1 becomes dysfunctional as a constitutive CETP inhibitor owing to its glycation. This has been shown to alter its physicochemical properties, leading to impairment of CETP inhibition (Dumont *et al.*, 2005; Bouillet *et al.*, 2014).

CETP has also been implicated in phospholipid exchange, along with PLTP, another homologous protein belonging to the same family as CETP. This is further supported by the fact that CETP genetic deficiency reduced plasma phospholipid exchange activity by 50 %, suggesting both CETP and PLTP may have overlapping functions. However, the introduction of CETPtg into PLTP-KO mice did not ameliorate PL exchange activity, clearly suggesting that CETP does not exchange PLs (Kawano *et al.*, 2000).

CETP activity was also dependent on the biochemical properties of neutral lipid substrate. The increase in acyl chain length led to an increase in CETP-mediated transfer, while an increase in unsaturation of acyl chains declined CETP-mediated lipid uptake (Green and Pittman, 1991). The acyl chain length and unsaturation have been shown to regulate monolayer curvature. Furthermore, the presence of medium-chain fatty acids [lauric (12:0) and myristic (14:0)] influences the lipoprotein surface mobility more than short chains [octanoic (8:0) and decanoic (10:0)] and longer chains [palmitic (16:0) and stearic (18:0)] acids. The further introduction of unsaturation also increased this mobility, possibly affecting the surface availability of neutral lipids and, therefore, influencing their uptake (Foucher *et al.*, 1996).

## 2.4 Mechanism of Lipid Transfer

The lipid transfer through CETP has been explained using two hypotheses: the shuttle model and the tunnel model. In the shuttle model, CETP binds to a single lipoprotein, forming a binary complex, followed by transfer of lipids between lipoprotein and CETP. This results in dissociation, allowing CETP to continue binding to and exchanging lipids with other lipoproteins. However, in the tunnel model, CETP forms a ternary complex by simultaneously bridging two lipoproteins. This is followed by a bidirectional transfer

between lipoproteins, and finally, the dissociation of this ternary complex (Sacher *et al.*, 2023). Based on the concave shape of the protein, it was presumed that CETP could bind a single lipoprotein at a time by modulating its curvature (Qiu *et al.*, 2007). However, electron micrographs did not show CETP adhering to the convex surface of HDL through its arch (Zhang *et al.*, 2015). It was later shown to penetrate HDL (~ 30 Å ± 10 Å), LDL (~ 25 Å ± 10 Å) and VLDL ( ~ 20 Å ± 10 Å) (Zhang *et al.*, 2012).

Further, co-incubation of CETP with HDL and LDL, or HDL and VLDL, EM micrographs showed 25 % of HDL-CETP-LDL complexes and 30 % of HDL-CETP-VLDL complexes, respectively (Zhang *et al.*, 2012). These observations do not totally exclude the shuttle mechanism, as the ternary complexes were found to coexist with the binary complex of HDL/LDL/VLDL-CETP (Lauer *et al.*, 2016). It is also proposed that binding to one lipoprotein initiates a conformational change that elicits binding to the second lipoprotein with increased affinity (Zhang *et al.*, 2017). In the absence of direct evidence supporting either mechanism, the debate over the two hypotheses persists. Irrespective of the shuttle or tunnel mode, the lipid transfer occurs by: *1)* Binding of CETP to one or two lipoproteins *2)* Lipid transfer from lipoproteins into CETP *3)* Detachment from lipoproteins

### 2.4.1 Binding of CETP to One or Two Lipoproteins

Although a simple lock-key interaction mediated by CETP and HDL curvature seems a likely player in lipoprotein binding, it does not agree with CETP lipoprotein penetration observed in EM micrographs. Immobilization of CETP on rHDL and liposome surfaces showed that binding of CETP to HDL is non-specific and dominated by phospholipids rather than apoA1 or apoA2 (Zhang *et al.*, 2015). However, once bound, this interaction may be stabilized by apoA1 and the presence of E2 fatty acyl esters (Cilpa-Karhu *et al.*, 2015). Furthermore, lipoproteins may associate with one or more CETPs at the same time, an ability entirely attributed to the size, differential surface properties, apolipoprotein compositions, and curvature of lipoproteins (Zhang *et al.*, 2015, 2012; Morton and Greene, 2003*b*). Increasing the electronegativity of the lipoprotein also stabilizes the CETP-lipoprotein interaction; however, the lipid transfer activity remains unaffected (Morton and Greene, 2003*a*).

The circulating molar ratios of lipoproteins in themselves are important physiological determinants for CETP-lipoprotein binding selectivity. Normogenic plasma levels of lipoproteins are in the order: HDL (~ 13 $\mu$M) > LDL (~ 2 $\mu$M) > VLDL (~ 0.1 $\mu$M), which naturally provides a kinetic argument in CETP's predominant association with HDL as observed by a high number of HDL-CETP complexes in electron micrographs compared to LDL or VLDL (Zhang *et al.*, 2015; Charles and Kane, 2012). However, despite the high plasma molar concentration of HDL, no HDL-CETP-HDL ternary complexes are formed, implying the primary motive for ternary complex is to facilitate a heteroexchange. (Zhang *et al.*, 2015, 2012)

Molecular assessment of CETP-HDL complex through MD simulation has identified a Trp cluster (W105, W106, and W162) and flexible regions (res 90-110 and res 150-170) at the N-barrel domain end of CETP to play an important role in anchoring and penetration (Cilpa-Karhu *et al.*, 2015; Koivuniemi *et al.*, 2012). Upon penetration, these regions are pulled apart to give rise to an opening through which neutral lipid enters the CETP tunnel (Cilpa-Karhu *et al.*, 2015; Lei *et al.*, 2016). The role of N-terminal flexible regions in binding and penetration is further supported by the fact that this region, including W105, is highly conserved amongst 26 CETP orthologs (Charles and Kane, 2012). Further, mutants of residues A48, G53, K98, and N165 result in declined CETP lipid transfer activity (Wang *et al.*, 1991).

### 2.4.2 Lipid Transfer into the Tunnel

Given the distinctive lengths of the two barrel domains (N-barrel is narrower and longer than C-barrel domain) and the differential depth at which these domains penetrate lipoproteins, a distinct mechanism may exist to allow neutral lipid access (Zhang *et al.*, 2012; Charles and Kane, 2012). Specifically, since HDL is smaller, CETP can penetrate the core of HDL, while it may just puncture the surface of the much larger VLDL (Koivuniemi *et al.*, 2012). It has been shown that the lipid transfer rates of CETP are dependent on the concentration of surface available CE and TG (Morton and Greene, 2003*b*). Moreover, a concentration of 0.1 % mole of neutral lipids in the surface of lipoprotein was essential to initiate a lipid transfer event, failing which the complexes may simply dissociate without being productive (Morton and Greene, 2003*b*).

Neutral lipids located within the core of lipid droplets (LD) are known to actively drive LD expansion by inserting into the surface phospholipid (PL) monolayer (Kim *et al.*, 2021). This surface reorientation of neutral lipids functions to reduce high surface tension during droplet growth and simultaneously generates surface defects. These defects are critical for protein recognition, a process often mediated solely by Tryptophan (Trp) residues due to their strong affinity for membrane interfaces (Kim *et al.*, 2021). Although lipoproteins differ from LDs due to the presence of regulatory apolipoproteins, it's plausible that similar neutral lipid-mediated surface defects might also form on lipoproteins, creating the specific sites recognized by the Trp residues lining the barrel of CETP (Cilpa-Karhu *et al.*, 2015).

Upon binding to lipoprotein through its flexible ends, CETP barrels undergo low-energy fluctuations that may gate lipid entry into the central tunnel (Chirasani *et al.*, 2016*a*). Further, a series of small, isolated hydrophobic cavities (diameter >5 Å) present along the central axis of both $\beta$-barrel domains may align linearly, on conformational changes, pushing the protein from its inactive state (disconnected cavities) to the active state (connected cavities with a continuous tunnel) that can facilitate lipid transfer (Zhang *et al.*, 2012).

The entry of lipids into the tunnel is proposed to be guided by a concentration gradient (Zhang *et al.*, 2015). In the case of a binary complex, this gradient may result from a high molar concentration in the donor particle, with no lipid within the CETP tunnel. In the case of a ternary complex, this gradient may be generated due to the differential concentration of CEs or TGs on the surfaces of donor and acceptor lipoproteins. Further, the CEs and TGs in the surface of lipoproteins may themselves compete amongst each other to gain entry into the tunnel (Cilpa-Karhu *et al.*, 2015; Morton and Greene, 2003*b*), as there are no additional proteins reported to facilitate lipid entry into the tunnel. Further, the movement of lipid inside the tunnel is proposed to be guided by simple diffusion (Morton and Greene, 2003*b*; Lei *et al.*, 2016).

Steered MD study of a minimal ternary complex showed that as CE penetrates the pore of CETP and continues to move up the tunnel, it leads to counter-clockwise rotation of the CETP $\beta$-barrel domain to generate a continuous tunnel (Lei *et al.*, 2016). During its movement through the tunnel, it forms physical contacts with F115, R158 and F167 in the N-barrel domain that attracts the CE molecule initially allowing it to enter the

tunnel; I15, L23, A202, I205, L206, F263, F265, M433 in the neck region that act as an energy barrier; and F301, M412 in the C-barrel region that allows for the rotation of the sterol ring and its eventual exit from the tunnel.

Several mutations in the neck region significantly alter the lipid transfer activity of CETP, supporting the above route of CE traversal. I15W, S230A, H232A, M433W, L457W, M459W and the double mutants I443W-V198W, I443W-L457W, L457W-M459W affect both CE and TG transfer. Moreover, Y375S, F265R, and F270R were not secreted, suggesting these mutants destabilized the CETP structure (Qiu *et al.*, 2007). Additionally, the V198W mutation led to a significant decline in TG transfer, while CE transfer remained unaffected (Qiu *et al.*, 2007). Q199A significantly affected CE transfer, with no observed effect on TG transfer through CETP (Qiu *et al.*, 2007). Considering the tunnel is extremely hydrophobic, the introduction of a bulky and rigid amino acid, such as tryptophan, or a charged Amino Acid like Arginine or a polar serine may significantly impact tunnel characteristics. These large groups may also occlude the neck, further constricting the path available for lipid movement (Qiu *et al.*, 2007).

Another structural component of CETP is the PLs that plug the two openings at the concave surface. The PL-bound orifices lie 15 Å away from the surface of HDL in the CETP-HDL complex (Zhang *et al.*, 2012). However, their origin hasn't been elaborately reflected upon, especially since the electron micrographs showed CETP penetration using its barrel ends. However, their role in CE transfer has been addressed using MD simulations. It was observed that CETP-bound PLs contribute to the translational motion of CE by converting its bent form (unfavorable) into a linear conformation (favourable) (Keshavamurthy *et al.*, 2023).

## 2.5 CETP Inhibition

The rationale for inhibiting CETP stems from the observation that individuals with CETP deficiency exhibit hyperalphalipoproteinemia, characterized by elevated plasma levels of HDL and apoA1. Since HDL is associated with several protective roles, including antioxidant, anti-thrombotic, and anti-diabetic effects, in addition to its primary role as a cholesterol efflux mediator, increasing plasma HDL mass appears to be a beneficial approach. Besides inhibiting CETP, it also leads to a reduction of small dense

LDL particles, which are known to be pro-atherogenic in nature.

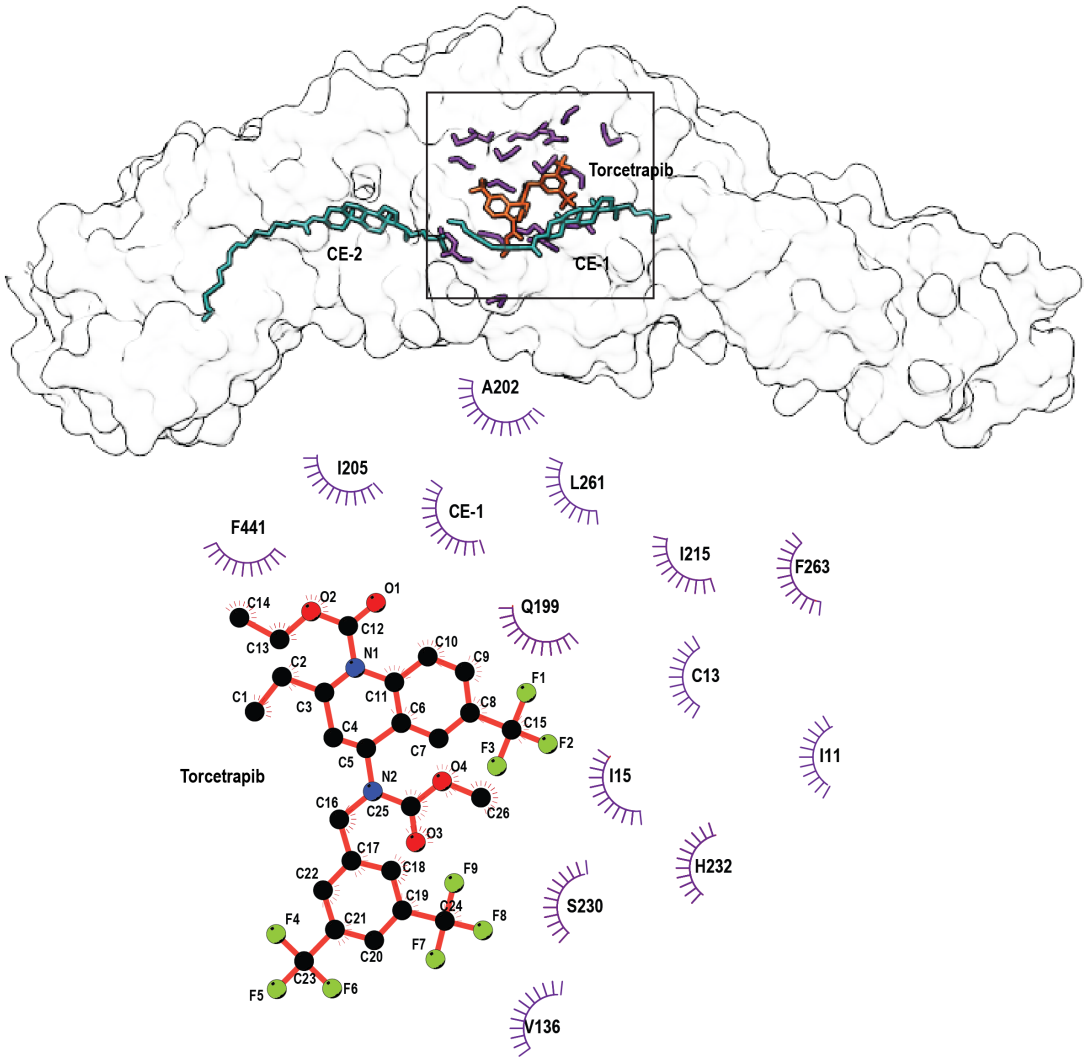



**Figure 2.2. Residues interacting with torcetrapib (a CETP inhibitor) as it binds at the interface of N-barrel and neck displacing N-PL as shown in PDBID:4EWS**

Considering the narrowest region of the CETP tunnel (neck), CETP inhibitors were designed to bind at the interface of the neck and the N-barrel domain. CETP bound by torcetrapib (a representative of the tetrahydroquinoline (THQ) series of inhibitors, $IC_{50}$ = 50nM) displaced PL bound to N-pocket (N-PL) and the CE bound near this region (Liu *et al.*, 2012) (Fig. 2.2). Further, the THQ group made hydrophobic contacts with F263, I11, C13, I215, and the aromatic faces of H232. And one trifluoromethyl group made contacts with V136, Q199, L129, V136, V198, P221 and L228 while methyl carbamate occupied a pocket formed by V198, Q199, A202 and I215 (Liu *et al.*, 2012) (Fig. 2.2). Another CETP inhibitor, with trifluoropropan-2-ol and tetrafluoroethoxyphenyl, called compound 2 and anacetrapib, was observed to bind at the same site as torcetrapib (Liu *et al.*, 2012; Chirasani *et al.*, 2016*b*). Further, due to chemical similarities between torce-

trapib, evacetrapib, and dalcetrapib, all of these inhibitors bind in the N-terminal pocket, which is also supported by experimental evidence that showed competitive inhibition for CETP binding amongst these inhibitors (Liu *et al.*, 2012; Ranalletta *et al.*, 2010). Steered molecular dynamics simulations revealed that inhibitors physically occlude the tunnel, causing CE to take a different path to exit the tunnel, even under a driving force. Movement past the neck region is bound by the inhibitor, posing a rate-limiting step with a high energy barrier (Dixit *et al.*, 2019).

### 2.5.1 Failures of CETP Inhibitors

Despite the atheroprotective nature of several CETP inhibitory strategies (small molecule inhibitors and antisense oligonucleotides) (Barter *et al.*, 2015; Shrestha *et al.*, 2017) in animal models, multiple clinical development programs for CETP inhibition have yielded disappointing results. Torcetrapib, due to its high target specificity, was the first inhibitor to reach clinical trials (ILLUMINATE), demonstrating a 60% increase in HDL cholesterol with a 20% decrease in LDL cholesterol in addition to statins (Nissen *et al.*, 2007). However, the trial was terminated prematurely due to a 58% increase in mortality of torcetrapib-treated patients compared to the placebo group. These effects were attributed to its off-target effects, which included excessive production of aldosterone, cortisol, and endothelin in the arterial wall, thereby raising blood pressure, as well as sepsis-related mortality (Forrest *et al.*, 2008). The development wasn't deterred by these results, and it was suggested that a CETP inhibitor without such toxicity may still prove beneficial in cardiovascular relief.

Dalcetrapib was the next inhibitor in development, which raised HDL cholesterol by 30% but had no effect on LDL cholesterol levels (Rhainds *et al.*, 2012). Clinical trials (dal-OUTCOMES) did not show vascular inflammation like torcetrapib, but there was also no apparent cardiovascular relief. Therefore, the trial was stopped early due to futility (Tall and Rader, 2018). Evacetrapib, the next inhibitor, proved to be more potent as it raised HDL cholesterol levels > 125% while decreasing LDL-cholesterol > 30% (Cao *et al.*, 2011). However, the trial (ACCELERATE) was similarly stopped early due to a lack of efficacy in reducing CVD events (Tall and Rader, 2018; Nicholls and Bubb, 2020). Then, Anacetrapib raised hopes by similarly increasing HDL cholesterol by more than 130% and lowering LDL cholesterol by more than 30% in patients with a high risk

of CVD or familial hypercholesterolemia. In addition to not exhibiting torcetrapib-like vascular toxicity, a 9% reduction in CVD events was reported in the REVEAL trial (Kastelein *et al.*, 2015). This trial demonstrated a direct association between non-HDL cholesterol levels and CVD events. However, due to the high lipophilicity of the compound, it had a long half-life and significant accumulation in adipose tissue (Johns *et al.*, 2019), which ultimately led the manufacturer to withdraw its regulatory approval.

| TRIAL (drug) | Patients | Lipoprotein Changes | Duration | Outcome | Comments | Citations |
|---|---|---|---|---|---|---|
| ILLUMINATE (Torcetrapib) | 15,067 High CV Risk | 72% increase in HDL-C , LDL-C decreased | 24 months | Increased CV events, Increased death, Increased systolic blood pressure (SBP) (5mm) | Electrolyte disturbances, hyperaldosteronism identified as off-target effects | DOI: 10.1056/NEJMoa0706628 |
| dal-OUTCOMES (Dalcetrapib) | 15,871 Post acute coronary syndrome | ~30 % increase in HDL-C,LDL-C unchanged | 31 months | Frequency of CV events unchanged, Increased SBP (0.6 mm) | Trial stopped early for futility. Possible benefit in a genetic subgroup. | DOI: 10.1056/NEJMoa1206797 |
| ACCELERATE (Evacetrapib) | 12,092 High risk of vascular disease | 133% increase in HDL-C, LDL-C decreased | 26 months | Frequency of CV events unchanged, Increased SBP (1.2 mm) | Trial stopped early for futility | DOI: 10.1056/NEJMoa1609581 |
| REVEAL (Anacetrapib) | 30,449 High risk of vascular disease | 104% increase in HDL-C, 17% decrease in LDL-C | 4.1 years | Decreased CV events, Increased SBP (0.7 mm) | Trial completed decreased new onset diabetes | DOI: 10.1056/NEJMoa1706444 |
| BROADWAY (Obicetrapib) | 354 (familial hypercholesterolemia) 2,530 (ASCVD) | 140% increase in HDL-C, 29.9% decrease in LDL-C | 12 months | 3.2% reduction in CV events | | DOI: 10.1056/NEJMoa2415820 |

**Table 2.1.** Table summarizing the outcomes of CETP inhibitors on cardiovascular relief in phase III clinical trials.

The modest effect on CVD events or off-target effects for each of the inhibitors is attributed to the compound itself, rather than the drug class. Therefore, the quest for improvement in inhibitor design is ongoing. While the mechanism of action of all inhibitors in this class involves occlusion of the tunnel (Liu *et al.*, 2012; Dixit *et al.*, 2019; Koivuniemi *et al.*, 2012; Yang *et al.*, 2018), the binding site and predominantly hydrophobic mode of action remain consistent across the drug class. Accordingly, there are three polar residues in the inhibitor-binding site, Q199, S230, and H232, which have been utilized in the next inhibitor, obicetrapib, to improve its logP (Kastelein *et al.*, 2023). Indeed, the introduction of hydrophilic groups that now interact with these residues has resulted in the most polar inhibitor of the class (logP = 4.9) compared to anacetrapib (logP = 9.2) and evacetrapib (logP = 7.9) (Nicholls *et al.*, 2022). It also exhibits better bioavailability and greater efficacy at lower doses, thereby showing promise. Preliminary clinical results show a significantly reduced LDL-cholesterol, apoB, non-HDL cholesterol, Lp(a), and a significantly increased HDL-cholesterol and apoA1 in CVD pa-

tients after 4 weeks of treatment with obicetrapib in addition to statin therapy (Nicholls *et al.*, 2022; Kastelein *et al.*, 2023) Table 2.1 summarizes the major outcomes of the CETP inhibitors in phase III clinical trials.

The major learnings from the trials challenge the HDL-cholesterol perspective, suggesting that the reduction in LDL-cholesterol, rather than the induction of HDL-cholesterol, contributes to a reduced CVD risk (Farrer, 2018). Moreover, the composition and particle number of HDL rather than net plasma HDL-cholesterol may be more relevant (Kontush, 2015). Another observation from these trials highlights the impact of CETP inhibition on improving metabolic control with a lower rate of diagnosis of T2D, and a slower rate of progression from dysglycemia to diabetes (Dangas *et al.*, 2022).

# CHAPTER 3

# Methodology

The primary workflow used in this study employs classical MD to equilibrate the protein structure, followed by steered MD to pull the lipid through it, followed by free energy calculations through umbrella sampling simulations. In this chapter, these methodologies will be discussed in detail.

## 3.1 Principles of Molecular Mechanics

Molecular mechanics ignores the electronic motions and calculates the energy of the system as a function of only the atomic nuclear positions. Therefore, this is the method of choice for systems that contain a significantly large number of atoms, such as macromolecular complexes studying protein-protein interactions, protein-lipid interactions, protein-DNA interactions, and so forth.

The functioning of molecular mechanics is based on the Born-Oppenheimer approximation of the Schrodinger wave equation. Schrodinger's wave equation describes the motions of the electrons and nuclei in a molecular system from first principles, describing the spatial probability distributions corresponding to the energy states in a stationary quantum system. Born-Oppenheimer approximation assumes that the nuclei are much heavier than electrons, moving much slower, allowing for the separation of Schrodinger's equation into electronic and nuclear parts.

In theory, the overall effectiveness of molecular mechanics depends on the validity of its three underlying principles: (1) basic thermodynamic assumption, which assumes that macromolecules are driven to their native minimum energy state without the advent of external catalysts (2) additivity of the effective energy potentials that assumes that the effective molecular energy can be expressed as a sum of potentials derived from simple physical forces: van der Waals, electrostatic, mechanical-like strains arising from "ideal" bond length and angle deviations, and internal torsion flexibility (rotation of two

chemical groups about the bond joining them). The forces can be separated into local (bonded) and non-bonded (non-local) terms. (3) Transferability of these potentials assumes that potentials can be developed to incorporate all experimental data for representative structures and then be applied successfully to the prediction of large biological molecules composed of the same chemical subgroups. This is a reasonable assumption, as bond lengths and bond angles tend to adopt similar values in different molecular species under normal conditions.

The following terms are used to describe the bonded components of the system:

**Positions and distances**: For a molecular system of N atoms (possibly including atom groups) in Cartesian coordinate space, let

$$x_i = (x_{i1}, x_{i2}, x_{i3}), where i = 1, ...., N, \tag{3.1}$$

denote the position vector of atom *i*, and

$$r_{ij} \equiv x_j - x_i \tag{3.2}$$

denote the distance vector from atom i to j. The potential energy function E then depends on all Cartesian variables of the atoms:

$$E = E(X) \equiv E(x_1, x_2, ...., x_N) \tag{3.3}$$

**Bond angles**: A bond angle $\Theta_{ijk}$ formed by a bonded triplet of atoms *i–j–k* is expressed as an inner (or dot) product:

$$cos\Theta_{ijk} = \frac{(x_k - x_j) \bullet (x_i - x_j)}{r_{ij} r_{ik}} \tag{3.4}$$

**Dihedral angles** : A dihedral angle $\tau_{ijk}$ , defining the rotation of bond *i–j* around *j–k* with respect to *k–l*, is expressed as:

$$cos(\tau_{ijkl}) = n_{ab} \bullet n_{bc} \tag{3.5}$$

Here vectors $n_{ab}$ and $n_{bc}$ denote unit normal planes spanned by vectors $\{a, b\}$ and $\{b, c\}$, respectively, where $a = r_{ij}$, $b = r_{jk}$, and $c = r_{kl}$. Denoting $\Theta_{ab}$ and $\Theta_{bc}$ as angles $\Theta_{ijk}$ and $\Theta_{jkl}$, respectively:

$$cos(\tau_{ijkl}) = \frac{aXb}{\|a\| \, \|b\| \, sin\Theta_{ab}} \bullet \frac{bXc}{\|b\| \, \|c\| \, sin\Theta_{bc}} \tag{3.6}$$

Bond length potentials can be considered as "strain" terms that model small-scale deviations from reference values. The reference values for different chemical bonds can be obtained from solved X-ray crystal structures as well as from quantum mechanical solutions to equilibrium structures of small molecules.

**Improper dihedrals**: These are the angles between two planes where the two central atoms are not connected by a bond. Instead, one atom is connected to three other atoms.

During "dynamics," the bond lengths, angles, dihedrals, and improper dihedrals deviate due to molecular forces. These deviations are further described as follows:

**Bond stretching**: Bond stretching is modelled as a spring, typically using a harmonic potential (Hooke's Law). This potential describes the energy associated with the deviation of a bond length from its equilibrium value. The harmonic potential is defined by a force constant (k) and the equilibrium bond length ($r_0$).

The potential energy (V) due to bond stretching is calculated using the following equation:

$$V_{stretching} = \frac{k_s}{2}(r - r_0)^2 \tag{3.7}$$

where $k_s$ is the stretching force constant, a measure of bond stiffness determined empirically, and r is the bond length.

**Angle bending**: Angle bending refers to the energy associated with deviations of bond angles from their equilibrium values. Like bond stretching, it is treated using a harmonic potential.

$$V_{bending} = \frac{k_b}{2}(\Theta - \Theta_0)^2 \tag{3.8}$$

where $k_b$ is the bending force constant, which is determined empirically, and $\theta$ is the bond angle.

**Torsional terms**: Torsional terms describe the energy change associated with the rotation around a chemical bond of the vicinally bonded atoms. The torsional energy is typically represented using a Fourier series, enabling the modeling of various potential energy profiles with multiple minima and maxima.

$$V_{torsion} = \frac{V_0}{2}(1 + cosn\omega) \tag{3.9}$$

**Improper torsion and out-of-plane bending**: This term describes the planarity of groups of atoms, such as those around sp2 hybridized carbons in carbonyl groups or aromatic rings. These terms are not directly part of the typical bonded interactions but are added as a correction to ensure that these groups remain flat, which is important for accurately representing molecular structures and properties.

Similarly, non-bonded interactions account for atoms (greater than two bonds apart) interacting through van der Waals attraction, steric repulsion, and electrostatic attraction/repulsion depend on their distance from each other. For two approaching non-bonded atoms, the interaction is attractive (London dispersion force) until the atoms get too close and start to repel each other (van der Waals repulsion/steric strain).

The empirically derived values are specific to the force field used. Furthermore, a potential energy function is constructed, which can be minimized to compute favorable regions in multidimensional configuration space. Molecular dynamics simulations can derive thermally accessible states of the system by integrating this function. The geometric quantities constituting the potential energy function are a simple sum of the bonded and non-bonded interactions:

$$U(r) = U_{bonded}(r) + U_{nonbonded}(r) \tag{3.10}$$

Of this, the bonded term is further made up of the bonded interactions described above to give:

$$U_{bonded}(r) = \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\substack{bond \\ angles}} \frac{1}{2} K_\Theta (\Theta - \Theta_0)^2 + \sum_{\substack{improper \\ dihedral \\ angles}} \frac{1}{2} K_\xi (\xi - \xi_0)^2 + \sum_{\substack{dihedral \\ angles}} K_\varphi [1 + cos(m\varphi - \delta)] \tag{3.11}$$

Here b,$\Theta$, $\xi$, and $\varphi$ represent the bond lengths, bond angles, improper dihedral angles, and torsional dihedral angles, respectively. The variables with a subscript of 0 are the reference or ideal values and are parameters of the force field. Generally, all interactions have a harmonic functional form, except for the torsional dihedral-angle term, which has a trigonometric form. The labels K indicate force constants.

The non-bonded energies are calculated using two terms: the first is the Lennard-Jones term, which describes the van der Waals interaction, and the second is the Coulomb term, which deals with the electrostatic interactions between the partial charges of the atoms. The former term describes atomic repulsion due to atom-atom overlap, and the attraction due to London dispersion interactions

$$U_{non-bonded}(r) = \sum_{\substack{atom \\ pairs}} \left(\frac{C_{12}}{r_{12}} - \frac{C_6}{r^6}\right) + \sum_{\substack{atom \\ pairs}} \frac{q_1 q_2}{4\pi\varepsilon_0 r} \tag{3.12}$$

where $r$ is the distance between the two atoms, for which non-bonded energy is calculated. $C_{12}$, $C_6$, $q_i$, and $q_j$ are force-field parameters.

## 3.2 Force-field : CHARMM

Additive force fields calculate the potential energy function by simply adding up all the individual atom terms. This representation does not account for electronic polarization and is therefore a significant approximation for biomolecular modelling. To compensate for the lack of a polarizability term, the previous version of the additive CHARMM force fields deliberately overestimated the molecular dipoles of compounds in the gas phase to estimate electrostatic interactions that occur in condensed-phase environments

(Zhu *et al.*, 2011). Later, the energy functions for biological macromolecules were significantly revised to include polarization effects using induced dipoles, fluctuating point charges, and classical Drude oscillators (Rick and Stuart, 2002).

This study utilizes the CHARMM36 force field for the parametrization of proteins and lipids. These parameters have been optimized to reproduce NMR observables, such as scalar couplings across hydrogen bonds, residual dipolar couplings, and relaxation order parameters for backbone, side chains, and amino- and methyl-containing groups (Best *et al.*, 2012).

Similarly, for lipids, modified torisonals, Lennard-Jones, and partial atomic charges obtained from quantum mechanical and experimental thermodynamic data have been used for parameter optimization (Klauda *et al.*, 2010). Since these calculations have been explicitly performed for bilayers, and neutral lipids do not form bilayers, modified parameters for TG that are compatible with cutoff-based and particle mesh Ewald schemes for treatment of Lennard-Jones interactions were used, which have been shown to reproduce interfacial properties of neutral lipids in complex assemblies (Campomanes *et al.*, 2021).

## 3.3 Water Models

While studying molecular interactions, a system is generally assembled with all its physiological components. For modeling inter-atomic interactions within a protein, this system consists of a protein immersed in water. Water models describe the complete set of parameters necessary to perform MD simulations of water. They are designed with the physical properties of liquid water in mind and, as such, aim to reproduce those behaviors in computer simulations. There are several models available, each developed to fit a specific physical parameter of liquid water. The following discussion provides a brief overview of the most common models and the model chosen for this study.

**Three-point**: Three-point water models represent water molecules using three interaction sites, typically with charges on each atom and Lennard-Jones parameters for the oxygen atom. These models are widely used in molecular dynamics simulations due to their computational efficiency. The widely used simple point charge (SPC) model is modelled as a rigid isosceles triangle and accurately reproduces the density and dielec-

tric constant of bulk liquid water (Jorgensen *et al.*, 1983). Similarly, TIP3P is computationally less expensive and more suitable for larger systems and longer simulation times (Jorgensen *et al.*, 1983; Mark and Nilsson, 2001*a*).

**Four-point**: This model describes the interaction between molecules using a set of point charges on four interaction sites: one for oxygen and three for hydrogen or a massless "dummy" site (M) near the oxygen along the bisector of the H-O-H angle (Jorgensen *et al.*, 1983). This model is a more accurate representation of the structure of water and its dynamical properties compared to their three-point counterparts. The addition of the dummy site captures the electrostatic interactions of water better and therefore can accurately reproduce density, heat of vaporization, and diffusion coefficients of bulk liquid water (Pathirannahalage *et al.*, 2021).

**Five-point**: This model includes two additional interaction sites to better represent the lone pairs of the oxygen atom, leading to a more accurate tetrahedral structure and improved reproduction of experimental water properties (Mahoney, 2000). These models can accurately reproduce properties like radial distribution function, temperature of maximum density, in addition to density, diffusion coefficients, and dielectric constants (Pathirannahalage *et al.*, 2021). However, this model has a high computational cost.

The current study utilizes the CHARMM modified TIP3P. In comparison to the original TIP3P, this model has Lennard-Jones parameters on both the oxygen atom and hydrogen atoms (Price and Brooks, 2004). This modification prevents singularities in calculations by keeping the hydrogen atoms from getting too close to the charged atoms. Further, it reproduces the tetrahedral H-O-H angle, 104° and has slightly better structural and thermodynamic properties than SPC or SPC/E models, and a significantly lower computational cost than TIP4P and TIP5P water models. Moreover, CHARMM proteins, nucleic acids, lipids, and carbohydrates have all been developed with respect to the TIP3P water model.

## 3.4 Molecular Dynamics Simulation

MD simulation simulates the motion of a system according to classical Newton's equations of motion under the influence of a force field, yielding trajectories of particles within the system. It uses numerical integration, starting from the evaluation of the po-

tential energy of all atoms constituting the system, followed by the evaluation of atomic forces at each time step. The force evaluation is computationally dominated by the large number of non-bonded interactions, particularly by the long-range electrostatic interactions.

$$F_i = m_i a_i \tag{3.13}$$

where $F_i$ is the force exerted on particle $i$, $m_i$ is the mass of the particle and $a_i$ is the acceleration of the particle. This force can also be expressed as a gradient of the potential energy,

$$F_i = -\nabla_i V \tag{3.14}$$

Combining the above two equations gives,

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{3.15}$$

where V is the potential energy of the system.

Next,

$$F = m.a = m.\frac{dv}{dt} = m.\frac{d^2 x}{dt^2} \tag{3.16}$$

When acceleration is constant, this becomes,

$$a = \frac{dv}{dt} \tag{3.17}$$

which may be written as,

$$v = at + v_0 \tag{3.18}$$

and,

$$v = \frac{dx}{dt} \tag{3.19}$$

which may be integrated once to obtain,

$$x = v.t + x_0 \tag{3.20}$$

on combining this equation with the expression for velocity,

$$x = \frac{1}{2}t^2 + v_0 t + x_0 \tag{3.21}$$

which gives the value of $x$ at time $t$ as a function of acceleration, $a$, the initial velocity, $v_0$ and initial position, $x_0$

Further acceleration is the derivative of the potential energy with respect to the position $r$,

$$a = -\frac{1}{m}\frac{dU}{dr} \tag{3.22}$$

Therefore, to obtain a trajectory, only the initial positions of the atoms and the initial distribution of velocities and accelerations are needed, which can be determined by the gradient of the potential energy function.

The initial distribution of velocities is usually determined from Maxwell-Boltzmann or Gaussian distribution at a given temperature, giving the probability that an atom $i$ has a velocity $v_x$ in the direction $x$ at a temperature $T$.

### 3.4.1 Integration algorithms

Since the potential energy function of a complex system, such as a macromolecule, encompasses all the atoms in the system, it is impossible to find the exact analytical solution of Newton's equations for such a complex system. Therefore, these equations must be solved numerically by advancing the state of the system (positions and velocities) in small time steps. Assuming these positions, velocities, and accelerations can be

approximated by a Taylor series expansion,

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \ldots \tag{3.23}$$

$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + \ldots \tag{3.24}$$

$$a(t+\delta t) = a(t) + b(t) + \ldots \tag{3.25}$$

**Verlet algorithm**: The Verlet algorithm uses positions and accelerations at time $t$ and the positions from time $t-\delta t$ to calculate new positions at time $t+\delta t$. The Verlet algorithm uses no explicit velocities and is therefore simple, with no storage requirements; however, it is only moderately precise (Grubmüller *et al.*, 1991).

**Leap-frog algorithm**: In this algorithm, the velocities are first calculated at time $t+\frac{1}{2}dt$; these are used to calculate the positions, $r$, at time $t+\delta t$. In this way, the velocities leap over the positions, then the positions leap over the velocities. The advantage of this algorithm is that velocities are explicitly calculated, adding precision; however, they are not calculated simultaneously with the positions (Fincham, 1992).

### 3.4.2 Boundary condition

Treatment of boundaries is a crucial characteristic of molecular dynamics simulations. Since the molecules are arranged in a bounding box, it means that a relatively greater part of them will lie on the surface and will experience different forces from molecules in the bulk. The consequence of the finite size of the system is that the boundary conditions may affect the simulation results, especially when the system of interest is a homogeneous liquid or a solution. Usually, periodic boundary conditions (PBC) are adopted to reduce the surface effects. This technique involves simulating the system within a central cubic box surrounded by an infinite number of identical copies of itself. During the simulation, the molecules in the original box and their periodic images move exactly in the same way. Hence, when a molecule leaves the central box, one of its images will enter through the opposite side (Fig. 3.1). As a result, there are no physical boundaries

nor surface molecules. There are several shapes available in addition to the traditional cubic, such as dodecahedron, rhombic, and octahedral, which can be used depending on the system complexity, but they are all treated with periodic boundary conditions.

Even with fast computers, simplified approximations are needed to reduce the computational time to an acceptable level. The necessity of using a system of finite size, along with boundary conditions, also introduces artifacts. The fast multipole expansion method allows for relatively efficient handling of long-range interactions, and for periodic systems the Ewald summation technique is commonly implemented to compute the Coulomb interactions (Darden *et al.*, 1993); However, the most commonly used method to achieve a cost-effective computation is to use a spherical cutoff, which reduces the number of pair wise interactions by neglecting all interactions between particles separated by a distance larger than the cutoff. This cutoff differs for the chosen force field.

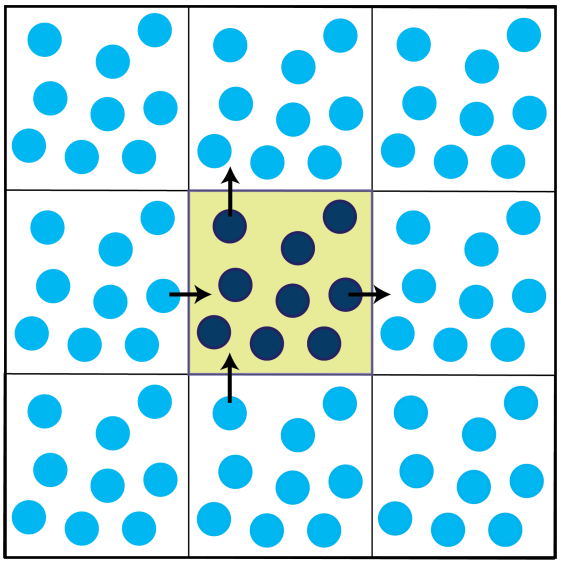


**Figure 3.1. Schematic depiction of periodic boundary condition in MD simulations.**The central box (colored yellow) represents the simulation box. Filled dark circles represent particles in the simulation box, and light blue circles represent their periodic image in other cells. Bold lines depict the movement of two particles near the boundary; as a particle leaves the simulation box, its image enters the box from the opposite end.

### 3.4.3 Constant Pressure Molecular Dynamics

When Newton's equations of motion are integrated, the total energy is conserved (adiabatic system), and if the volume is held constant, the simulation will generate a microcanonical ensemble (NVE). However, other statistical ensembles, such as canoni-

cal (NVT) and isothermal-isobaric (NPT) ensembles, represent the conditions under which biophysical experiments are performed than the standard microcanonical ensemble. Moreover, with the automatic control of temperature and/or pressure, slow temperature drifts that may occur due to force truncation errors are corrected. Further, rapid transitions to new desired conditions of temperature and pressure are more easily accomplished. Several methods for performing molecular dynamics at constant pressure have been proposed, ranging from ad hoc rescaling of atomic velocities to adjust the temperature to consistent formulations in terms of modified equations of motion that enforce the dynamics to follow the desired pressure constraint. Correcting the pressure in a simulation can be achieved by adjusting the inner virial through scaling of the inter-particle distances. The two most utilised methods are described next.

**Berendsen Coupling**: The pressure bath coupling method simulates a coupling of the system with an external pressure bath at the pressure $\rho_0$ and the interaction between this bath and the system is modulated by a barostat relaxation time constant $\tau P$. The pressure P is scaled to the reference pressure $\rho_0$ via an exponential law. For the strength of the coupling between the system and the "pressure bath", the same applies as for $\tau$ in the Berendsen thermostat: the larger $\tau P$, the weaker the coupling. While the Berendsen barostat and the thermostat are similar, they are also independent and fully separable (Berendsen *et al.*, 1984*a*).

The Berendsen barostat is known to suppress pressure fluctuations, leading to a non-canonical NPT ensemble. This may lead to inaccurate simulation properties, particularly in systems where these properties depend nonlinearly on the system's pressure. Moreover, in systems with complex structures, such as membrane proteins, it can introduce artifacts. Therefore, it is better suited for the initial equilibration of the system.

**Parrinello-Rahman** pressure coupling is where volume and shape are allowed to fluctuate and is based on the use of an extended Lagrangian, i.e., a Lagrangian that contains additional, artificial coordinates and velocities (Parrinello and Rahman, 1981). In most cases, the Parrinello-Rahman barostat is combined with the Nose-Hoover thermostat. The conventional Nose-Hoover algorithm only generates the correct distribution if there is a single constant of motion. The various methods for carrying out molecular dynamics at constant pressure are based on the same principles as the constant temperature scheme, with the role of temperature played by pressure and the role of atomic velocities

played by atomic positions.

## 3.5 Steered Molecular Dynamics

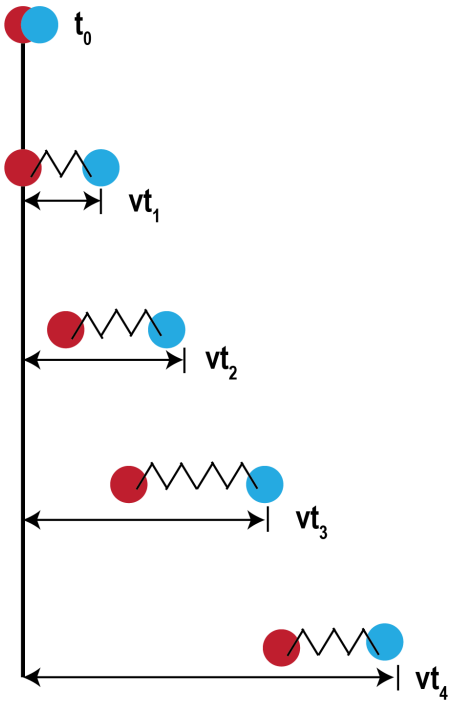



**Figure 3.2. Schematic depiction of constant velocity steered MD.** Pull group (blue) is connected to the particle (red) intended to be steered through an imaginary spring of constant $k$. During MD, constant velocity $v$ is applied on the pull group, which in turn displaces the original particle by a displacement of $vt$.

Classical molecular dynamics simulates systems such that the macroscopic properties remain constant over time. Simulations are therefore run for a long period of time so as to attain equilibrium. This equilibrium is described by the system's energy, temperature, and pressure having converged to constant values. Non-equilibrium molecular dynamics drives systems away from equilibrium, often by applying external forces or gradients.

Steered molecular dynamics (SMD) is a non-equilibrium molecular dynamics method employed to study the resistance of a molecule when it is moved in a certain direction. This resistance is expressed as a force. The external force applied to the system of interest enables the observation of changes in the molecule within the confines of simulation timescales. These simulations are typically used to estimate binding free energies of a ligand with a protein (estimating the unbinding of the ligand).

SMD utilizes a dummy particle that applies a steering force in order to traverse a reaction coordinate at a particular velocity. It is up to the user's discretion to either keep the force or the velocity constant in these simulations (Fig. 3.2).

## 3.6 Enhanced Sampling through Umbrella Sampling

Umbrella sampling of a chosen variable $\lambda$ leads to a uniform sampling of this variable in ideal conditions. A number of simulations $i$ with a modified potential are carried out, where the modified potential is described as:

$$H_i = H_0 + U_i\lambda \tag{3.26}$$

Here $U_i\lambda$ is the umbrella potential and $H_0$ is the Hamiltonian of the system.

After each simulation, the umbrella potential $U_i\lambda$, is updated such that a uniform sampling of the variable is expected. This is done based on the statistics from the previous simulations, $j = 1, 2, ..., i$.

For this purpose, the variable $\lambda$ is partiioned into bins with index k and $\mathrm{l} \in (\xi_k, \xi_{k+1})$, where $\xi_k$ denotes the boundaries between the bins.

For each of the simulations, the number of times $n_{j,k}$ in which the system is found in a particular bin is determined. Then an estimate $p_k^0$ for the probability of finding the unperturbed system in a particular bin is calculated by determining a self-consistent solution through the weighted histogram analysis method.

$$p_k^0 = \frac{\sum_j n_{j,k}}{\sum_j N_j f_j c_{j,k}} \tag{3.27}$$

$f_i$ is the scaling factor arising from the condition that the probability of finding the system in any of the bins should be equal to 1 and is given by:

$$f_i = \frac{1}{\sum_k c_{j,k} p_k^0} \tag{3.28}$$

where,

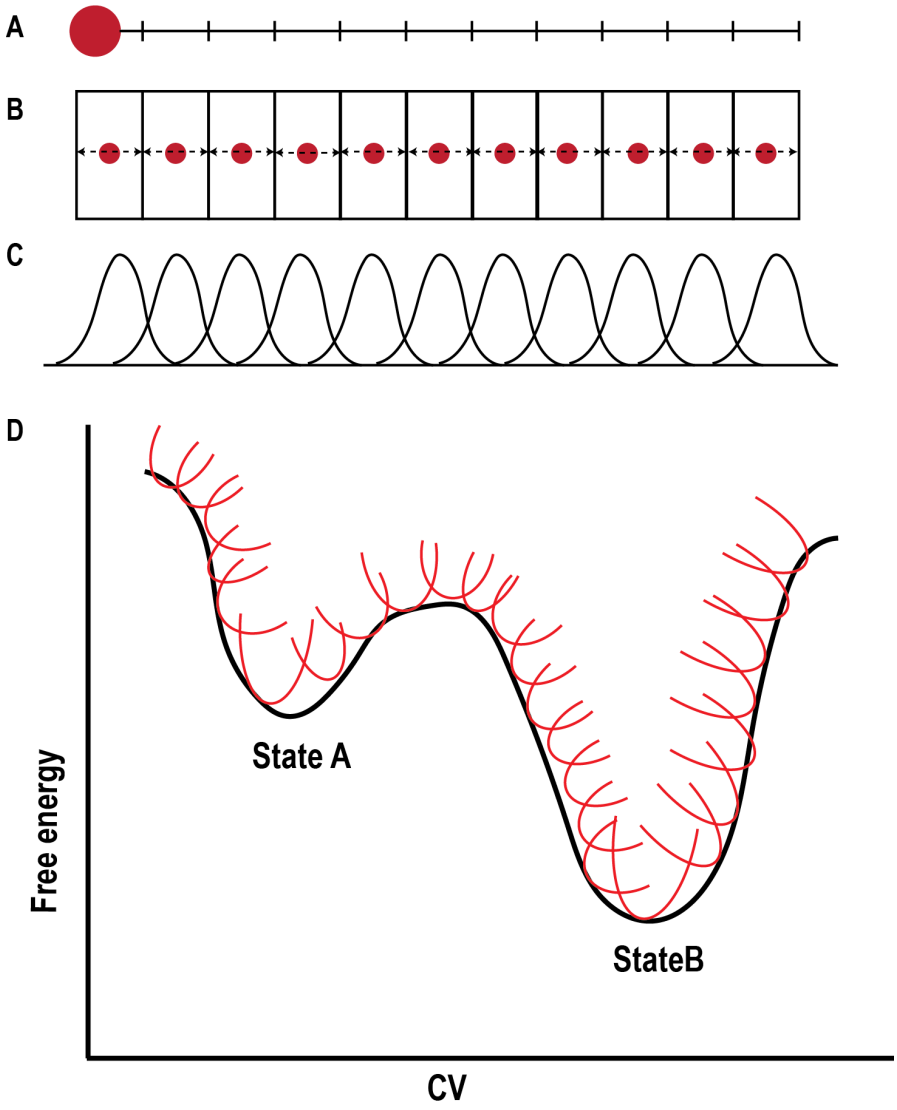



**Figure 3.3. Schematic depiction of constant umbrella sampling simulation. A.** Reaction coordinate is partitioned into bins of equal width. **B.** The pull group is retrained within each bin with a harmonic potential $U_i\lambda$. **C.** Each simulation for an umbrella bin is carried out independently, and the number of times a conformation is found in a particular bin is determined and used to generate a histogram. **D.** A free energy profile is obtained from the probability of finding the system in a particular energy state or bin.

$$c_{j,k} = e^{-U\lambda}(\xi_k)/RT \tag{3.29}$$

where R is the gas constant and T is the temperature of all simulations. While,

$$\xi_k = \frac{1}{2}\{\xi_k + \xi_{k+1}\} \tag{3.30}$$

is the midpoint of the bin $k$

$$N_j = \sum_k n_{j,k} \tag{3.31}$$

In an ideal umbrella sampling scheme, the simulations should sample the significant local minima and the transition state conformations without generating many unimportant conformations. To ensure this, it is important that the variable $\lambda$ or the reaction coordinate is suitable.

Transition states correspond to conformations with higher potential energies. Similarly, the unimportant conformations may have potential energies larger than those of the transition states or smaller than those of the local minima. Therefore, the potential energy of the three types of conformations is distinguishable, just as the reaction coordinate, $\lambda$.

Therefore, $\lambda$ can be equated to the potential energy $V$ of the system. And the Hamiltonian for the *ith* simulation is then given by:

$$H_i = H^0 + U_i(V) = K + V + U_i(V) \tag{3.32}$$

where K and V denote the kinetic and true potential energy of the system, $H^0$ is the Hamiltonian of the unperturbed system, and $U_i(v)$ is the umbrella potential applied in the *ith* simulation.

# CHAPTER 4

# Visualizing the tunnel within CETP to determine its function

## 4.1 Introduction

The structural folds in proteins enable them to carry out a variety of functions within the cell. These folds often generate spaces within protein structures that offer isolated environments favorable for binding ligands or substrates, facilitating enzyme catalysis, transport of small molecules, and conformational flexibility. The molecular makeup and physicochemical characteristics of these cavities are linked to the fundamental functions of proteins. For example, the binding pocket of chymotrypsin is surrounded by hydrophobic amino acids, which ensure effective substrate binding for catalysis (Ma *et al.*, 2005). Likewise, ion channels are selective regarding the ions they transport (Dudev and Lim, 2010), and lipid transfer proteins (LTPs) possess hydrophobic cavities to facilitate lipid transport in an aqueous environment (Wong *et al.*, 2017). Thus, a thorough characterization of all cavities within a protein—considering factors such as volume, depth from the surface, pH, electrostatics, propensity for hydrogen bonding, and conservation amongst homologues—can yield significant insights into the mechanisms of protein action and folding and help in designing complementary compounds that may act as activators or inhibitors (Nayal and Honig, 2006; Mason and Chen, 2018; Xue *et al.*, 2019).

Protein cavities can be classified according to their location, shape, and the number of openings into the following categories: *Pockets* are shallow indentations on the protein surface. *Clefts* are deeper indentations on the protein surface. *Channels* are pathways that link the protein's exterior to an internal location and usually associated with proteins that generate current; *Tunnels* are channels that do not generate current but still connect a point in the interior of the protein to the exterior. *Pores*, a specific type of tunnel that runs through the entire protein molecule and *Voids* are spaces within the protein that are closed at all ends. Table 4.1 provides an overview of the various types

| Type of cavity | Schematic | Functions | Examples |
|---|---|---|---|
| Pockets |  | Ligand binding | Protein Kinase C, agonist/antagonist pocket of GPCRs |
| Clefts |  | Enzyme catalysis, transport | Lipid binding proteins, LCAT, hydrolases |
| Channels/Tunnels |  | Enzyme catalysis, Transport (shuttle) | Lipid binding proteins, Pyruvate dehydrogenase, cytochrome c oxidase |
| Pores |  | Transport (membrane) | Ion channels, transporters, nicotinic acetylcholine receptor |
| Voids |  | Oligomerization | Chaperone, viral spike protein, HPeV protein complex |

**Table 4.1. The different types of cavities that can exist in a protein and the functions associated with them.**

of cavities found in proteins and the different functions ascribed to them. In this text, we refer to the term internal cavities as encompassing a category that includes pores, channels, tunnels, and voids (Fig. 4.1).

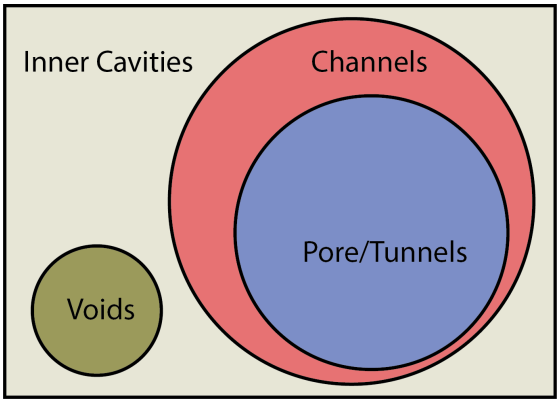



**Figure 4.1. Venn Diagram depicting the different types of inner cavities.**

Over the last two decades, several computational methods have been developed to predict, identify, and characterize protein cavities. These algorithms encompass a wide range of techniques that detect cavities by analyzing the geometric characteristics of the protein's molecular surface. These techniques can be further divided into the following categories:

*Sphere-based approaches*: These approaches involve modelling either the protein surface or the detection probe as spheres with a specific van der Waals radius (probe radius). The algorithms used for detection vary among the different tools available. For instance, Hole necessitates prior knowledge about the cavity's position and orientation to facilitate probe placement. This probe is then enlarged and shifted according to the direction specified by the user to measure the cavity's length and volume (Smart *et al.*, 1996). SURFNET renders the protein as a set of spheres with a user-defined van der Waals radius and then calculates the area and volume of the spaces encompassed by these spheres (Laskowski, 1995).

*Grid-based approaches*: These techniques position the protein within a three-dimensional grid of voxels and subsequently employ clustering or search algorithms to identify vacant voxels that signify empty regions within the protein. Examples of such methods are LIGSITE (Hendlich *et al.*, 1997), and PoreWalker (Pellegrini-Calace *et al.*, 2009*a*).

*Surface-based approaches*: These approaches examine the geometric characteristics of the surface, including curvature and surface fractal dimension, to detect cracks on

protein surfaces. Notable examples of these methods are CHUNNEL (Coleman and Sharp, 2009) and MSPocket (Zhu and Pisabarro, 2011).

*Tessellation-based approaches*: In these approaches, the molecular surface of a protein is segmented into triangles or polygons (tessellated). This creates a mesh or network of linked elements that encompasses the entire surface of the protein. Cavities are subsequently detected based on geometric or physico-chemical parameters such as solvent exposure, depth, and concavity. Examples of this include MOLE (Sehnal *et al.*, 2013), MolAxis (Yaffe *et al.*, 2008), CAVE (Buša *et al.*, 2010), and Fpocket (Le Guilloux *et al.*, 2009).

Many of the above-mentioned tools offer details about the residues surrounding a cavity and the cavity's volume or radius, but they often require human input *user-assisted cavity detection*, such as specifying the probe size or diameter, or identifying known seed residues that are part of existing cavities to initiate the empty voxel search. While the proteins that form channels might be well-annotated, as they contain two visible openings connecting the different cellular environments, an internal cavity may be formed by a complete void with no openings or multiple openings. Moreover, a protein might contain several internal cavities, and knowledge of residues/points lying in all such regions might not always be available. Therefore, this is a major limitation.

Additionally, each method for cavity detection comes with its own assumptions and, consequently, its own limitations. Sphere-based approaches struggle with proteins containing multiple cavities of varying sizes, as they necessitate frequent adjustments of probe sizes to identify cavities of different dimensions. In cases where the protein is rendered using spheres, smaller cavities may be merged or undetectable, depending on the chosen van der Waals radii. Grid-based methods are particularly sensitive to the initial grid settings, referred to as *grid-spacing sensitivity*, and the orientation of the protein, referred to as *orientation sensitivity*.

Tessellation-based approaches can sometimes mistakenly identify atoms or residues that border the cavity. These boundary issues, or *mouth opening ambiguity* are especially common with cavities that lead to the protein's surface such as tunnels, such that multiple surface residues are detected. Some geometric techniques require user intervention to commence cavity detection, making them less automated, while others use complex search algorithms that are time-consuming. Furthermore, most algorithms fail

to differentiate between two adjacent cavities, resulting in them being recognized as a single cavity.

Compounding this issue is the aspect of 'dynamicity': proteins within cells exhibit various movements that contribute to their biological functions. During these processes, protein-enclosed cavities undergo changes in their shape and size. To date, only a limited number of methods have been created to detect cavities that evolve over time. Notable examples include MDPocket, which tracks the time evolution of surface pockets (Schmidtke *et al.*, 2011), and CAVER 3.0, which focuses on the dynamic nature of channels and tunnels (Chovancova *et al.*, 2012). As a result, despite the existence of numerous robust cavity detection algorithms, much more work is required in this area to develop a single algorithm that can reliably identify all types of cavities. Table 4.2 summarizes the most frequently used cavity detection algorithms, their underlying algorithmic approach, the various features they offer to functionally characterize a cavity, and their limitations.

| Tools | Algorithm | Types of cavities detected | | | Features | | | Limitations | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Pockets / Clefts | Channels / Tunnels | Voids | Residue / Atom | Volume / Radius Profile | Hydrophobicity / Charge profile | Mouth Opening Ambiguity | User Assisted Detection | Grid Spacing Sensitivity |
| HOLE | Sphere | | * | | | * | | | * | |
| SURFNET | Sphere | * | * | * | * | | | | | |
| Fpocket | Tesselation | * | | | * | * | * | | * | |
| LIGSITE | Grid | * | * | * | * | | | * | | * |
| PoreWalker | Grid | | * | * | * | * | * | * | | * |
| Chunnel | Surface | | * | | | | | | | |
| MSpocket | Surface | * | * | * | * | | | * | * | |
| MOLE | Tesselation | * | * | * | * | * | * | * | * | |
| MoleAxis | Tesselation | * | * | * | * | * | | | | * |
| CAVE | Tesselation | | | * | * | | | * | * | |

**Table 4.2. A comparison table of the types of cavities detected, various features offered, and limitations of the different cavity detection tools**

CETP is a lipid transfer protein that belongs to the bactericidal permeability increasing factor (BPIF) superfamily. Notable members of this family include bactericidal permeability increasing protein (BPI), lipopolysaccharide binding protein (LBP), phospholipid transfer protein (PLTP) and Palate Lung and Nasal epithelium Clone (PLUNC) proteins. Each of these proteins contain two large barrel-shaped domains connected by a central $\beta$-sheet (BPI-fold). The internal cavity within these proteins is implicated in their lipid transfer function. Specifically, in the case of CETP, in addition to the CE binding cavities, a series of small, isolated hydrophobic cavities were observed (Zhang

*et al.*, 2012). It is hypothesized that the conformational changes in the protein bring these cavities together to create a continuous hydrophobic tunnel, facilitating lipid transfer. Therefore, a comprehensive characterization of the cavities possessed by CETP may yield meaningful insights into its mechanism of action. Given the existing challenges in protein cavity detection, a new algorithm called Characterization of Internal Cavities of Proteins (CICLOP) was developed, which combines the grid and tessellation-based approaches to identify internal cavities in proteins. This method was then used to characterize the cavity within CETP.

## 4.2 Approach

**Alignment of the Structure:** The user is offered two options for aligning their protein structure(s). In automatic mode, the protein is aligned by rotating the best-fit line to coincide with the positive Z-axis. This line minimizes the sum of the squares of the perpendicular distances of all the C$\alpha$ atoms. Only the C$\alpha$ atoms are used to minimize the computational overheads and requirements. Since all the C$\alpha$ atoms define the backbone of a protein structure, the best-fit line thus calculated gives a good initial approximation of the central pore axis. In the manual alignment, CICLOP assumes that the user-provided structure has the best-fit line (or the central cavity axis) aligned with either the negative or the positive Z-axis. The protocol for automatic alignment of the input structure has been rigorously tested for reproducibility, with the deviation from the mean of the number of residues detected by CICLOP lying within 1% (Fig. 4.2).

### 4.2.1 Identification and Mapping of the Inner Surface Residues:

The identification and mapping of inner surface residues by CICLOP is a six-stage process.

***Stage I:*** All the relevant information (such as coordinates of all atoms, the residues, and protein chains) is extracted from the given PDB file. All the atoms of the structure are subjected to a linear transformation, which places the protein in the first octant (for simplifying calculations) of the three-dimensional system, with coordinates of all atoms having either a positive value or a zero value.

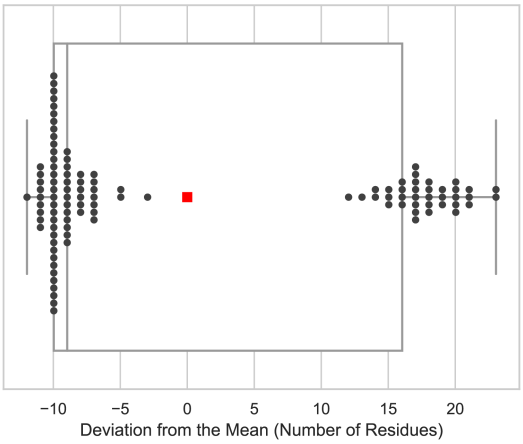



**Figure 4.2. Effect of the degree of rotation of the submitted structure imparted on CICLOP's performance in detecting residues on the inner surface.** Protein (PDBID:1AON) was rotated a full 360° and for each structure obtained at every 10°, residues that line the inner cavity were calculated using CICLOP. The box plot shows the deviations observed from the mean number of residues, detected on rotating this protein structure. The observed deviations in rotational variance across the three planes lie within ±0.05% with respect to the total inner lining residues.

***Stage II:*** The protein after the linear transformation is mapped to a three-dimensional grid of cubes similar to voxels (voxel is a unit of graphic information that defines a point in three-dimensional space). Each constituent voxel cube has a side length of 1 Å and has a volume of 1 $Å^3$. It may be noted that each voxel in the three-dimensional grid may either contain no atom, one atom, or more than one atom. The voxels containing no atoms are identified and marked empty. Each voxel cube from the complete grid contains references to cubes bounding it from all six faces, allowing this data to be represented as an undirected graph. Each node represents a voxel (the position of a group of atoms lying in close vicinity), and each edge represents an interface between the two voxels sharing a common face. To implement the algorithm, each voxel has been approximated as a point with X, Y, and Z coordinates. The 6 edges emanating out of this voxel (X, Y, Z) connect it to the voxels with coordinates (X+1, Y, Z), (X-1, Y, Z), (X, Y+1, Z), (X, Y-1, Z), (X, Y, Z+1), (X, Y, Z-1). Therefore, each voxel (X, Y, Z) is connected to its 6 adjacent voxels through these edges.

***Stage III:*** A breadth-first search algorithm is applied to look for empty nodes, starting from the node containing the coordinates of the geometrical center of the transformed protein. If the node is empty, it may be part of a cavity inside the protein. Therefore, this empty node is marked as visited, and edges emanating from this node are traversed.

The search continues in all six cardinal directions until no more empty nodes are encountered. The atoms surrounding the identified empty regions are selected and marked to be on the boundary surfaces for further analysis. This traversal aims to identify protein atoms that are adjacent to empty regions, including the surface surrounding inner cavities and the outer surface of the protein. Since the traversal happens over all the empty voxels, the search would identify the same protein atoms had it started from any other point but the geometrical center. This region physically represents the boundary surfaces of the protein (both the inner and the outer). This is done to reduce our search space for the atoms lying on the inner surface of the protein.

***Stage IV:*** The next step involves separating the inner surface from the outer surface. The algorithm for this step involves making 1 Å slices of the protein along the Z-axis. The Z coordinate is approximated to be the same for all the atoms in a slice. For every such slice made:

- The atoms marked to be on the boundary surfaces for the given Z slice are identified. The geometrical centre of these marked atoms is calculated.

$$X_{cen} = \frac{\Sigma atom.xcoord}{Total\ number\ of\ inner\ atoms} \tag{4.1}$$

$$Y_{cen} = \frac{\Sigma atom.ycoord}{Total\ number\ of\ inner\ atoms} \tag{4.2}$$

  Where $X_{cen}$ and $Y_{cen}$ are the X and Y coordinates of the geometric centre of the atoms marked to be on the boundary surfaces for the given Z slice.

- The mean radius at which these atoms lie from their geometrical centre is calculated. This represents a single value to denote the overall distribution of all the radii (calculated for every atom to the geometric center) in a given Z-slice.

$$R_{mean} = \frac{\Sigma\sqrt{(X_i - X_{cen})^2 + (Y_i - Y_{cen})^2}}{Total\ number\ of\ inner\ atoms} \tag{4.3}$$

  Where $R_{mean}$ is the mean distance at which these inner atoms lie for the given Z slice.

- The standard deviation of all the atoms (lying at a distance less than the $R_{mean}$ from the geometrical centre) is calculated from the $R_{mean}$.

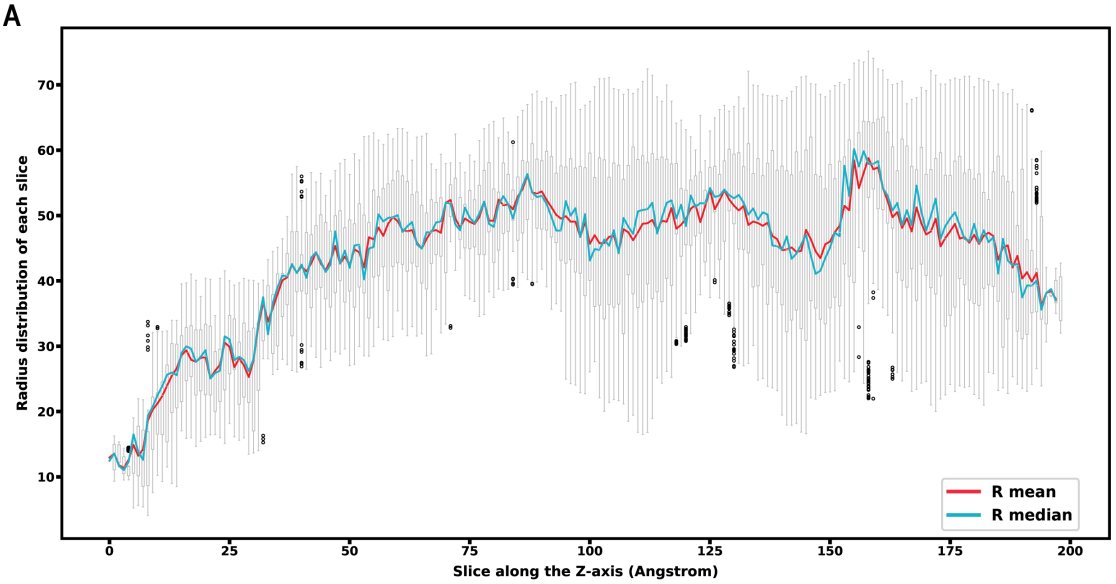
$$R_i = \sqrt{(X_i - X_{cen})^2 + (Y_i - Y_{cen})^2} \tag{4.4}$$

$$SD = \sqrt{\frac{\Sigma(R_i - R_{mean})^2}{Number\ of\ Inner\ atoms\ having\ R_i\ less\ than\ R_{mean}}} \tag{4.5}$$

Where $R_i$ is the distance from $(X_{cen}, Y_{cen})$ along the XY plane, $R_{mean}$ is the mean radius and $SD$ is the Standard Deviation of the $R_i$ from $R_{mean}$ $\forall$ atoms having $R_i \leq R_{mean}$

- Any atom lying outside the circle of $Radius = R_{mean}$ - (0.7) x $SD$ is unmarked and removed from the list of atoms lying on the inside lining. The value of 0.7 was chosen based on rigorous analysis performed on multiple proteins (Table 4.3).

A




B

| 2wvw | With standard deviation | Without standard deviation |
|---|---|---|
| *Recall* | 83.03% | 99.63% |
| *False Positivity Rate* | 0.49% | 25.28% |
| *Specificty* | 99.5% | 74.71% |
| *False Negativity Rate* | 16.96% | 0.36% |
| *Accuracy* | 92.79% | 84.89% |
| *Precision* | 99.13% | 73.08% |

**Figure 4.3. Optimizing the inner residue detection by CICLOP. A.** A box plot showing the distribution of radius values from the geometrical center of all the atoms (PDBID:1AON) that make up a Z-slice The $R_{mean}$ values (red) for each slice as well as the $R_{median}$ values (blue) have been overlaid. It can be seen that both $R_{mean}$ and $R_{median}$ values for each slice lie very close to each other. **B.** Comparison of the accuracy and precision performance of CICLOP with and without the use of standard deviation to filter out atoms detected on the cavity inner lining. The analysis was performed using the molecular dynamics simulation trajectory of PDBID:2wvw

The accuracy and precision values of CICLOP drop drastically without the use of standard deviation for filtering atoms detected to be in the inner cavity (Fig. 4.3 B). Further, it can be observed that as the SD cutoff decreases from 0.7 to 0, accuracy decreases,

while the precision of detection increases. While increasing the cut-off value from 0.7 to 1, accuracy increases, but precision decreases. A similar trend is evident in the case of recall and specificity as well as FPR and FNR (Table 4.3). Considering this trade-off between accuracy and precision of inner-residue detection, a cut-off of 0.7 is justified. Although the Z coordinates are dropped while calculating the radius of the cavity, these slices are later combined, and the Z-axis component is recovered.

Once an iteration of stages ***II***, ***III***, and ***IV*** is completed, the algorithm checks for any other empty voxel that may have been left unvisited in the 3D grid box. If found, the empty voxel is chosen and stages ***II***, ***III*,** and ***IV*** are repeated until all the empty voxels have been traversed at least once. Since there is a finite and fixed number of voxels (utilizing the formula shown below) in the grid box, the algorithm eventually terminates when all the empty voxels have been marked visited.

Total Number of voxels in a 3D box = ($X_{max}$ - $X_{min}$ -1) * ($Y_{max}$ - $Y_{min}$ - 1) * ($Z_{max}$ - $Z_{min}$ - 1)

Where $X_{max}$, $Y_{max}$, $Z_{max}$ are the maximum values of the X, Y and Z coordinates of protein atoms, while $X_{min}$, $Y_{min}$ and $Z_{min}$ are the minimum values of the X, Y and Z coordinates of protein atoms respectively.

***Stage V:*** Once all the atoms lining the inner surface have been identified, a new PDB file is created with the original coordinates, and the temperature factor column is modified such that 9999 is assigned to atoms lying on the inner surface while a value of 0 is allotted to all other atoms of the protein. This ensures that the user can effectively visualize the atoms lying on the inner surface identified by CICLOP.

***Stage VI:*** Finally, residues lining the inner surface are written in a separate file called 'residue.dat' as an output. This list contains all those residues for which even a single atom was detected on the inner surface.

To summarize briefly, instead of detecting the empty space in the protein that makes up the cavity, CICLOP aims to identify the protein atoms that surround these empty spaces (therefore efficiently and accurately identifying residues that make up the inner surface). This is performed by enclosing the protein in a 3D grid. Each voxel is marked as empty or non-empty based on whether any protein atom's coordinates lie inside that voxel. Each iteration of the algorithm traverses a connected component of empty vox-

| 1EK9 | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 97.05521 | 41.70213 | 58.29787 | 2.944785 | 82.94393 | 80.14184 |
| SD 0.1 | 95.09202 | 36.38298 | 63.61702 | 4.907975 | 83.64486 | 81.92389 |
| SD 0.2 | 93.86503 | 31.91489 | 68.08511 | 6.134969 | 84.50156 | 83.60656 |
| SD 0.3 | 91.65644 | 25.53191 | 74.46809 | 8.343558 | 85.43614 | 86.15917 |
| SD 0.4 | 88.22086 | 19.3617 | 80.6383 | 11.77914 | 85.51402 | 88.76543 |
| SD 0.5 | 83.68098 | 16.38298 | 83.61702 | 16.31902 | 83.72274 | 89.85507 |
| SD 0.6 | 80 | 11.48936 | 88.51064 | 20 | 83.17757 | 92.35127 |
| SD 0.7 | 76.44172 | 9.148936 | 90.85106 | 23.55828 | 81.7757 | 93.54354 |
| SD 0.8 | 72.02454 | 5.957447 | 94.04255 | 27.97546 | 80.14019 | 95.44715 |
| SD 0.9 | 66.25767 | 4.255319 | 95.74468 | 33.74233 | 77.1028 | 96.42857 |
| SD 1 | 61.10429 | 2.553191 | 97.44681 | 38.89571 | 74.45483 | 97.64706 |

| 2WVW | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 99.63 | 25.28 | 74.71 | 0.36 | 84.89 | 73.08 |
| SD 0.1 | 99.41 | 20.79 | 79.2 | 0.58 | 87.46 | 76.71 |
| SD 0.2 | 99.14 | 16.12 | 83.87 | 0.85 | 90.12 | 80.9 |
| SD 0.3 | 98.05 | 11.35 | 88.64 | 1.94 | 92.5 | 85.61 |
| SD 0.4 | 96.6 | 7.07 | 92.92 | 3.39 | 94.44 | 90.38 |
| SD 0.5 | 93.8 | 3.46 | 96.53 | 6.19 | 95.43 | 94.91 |
| SD 0.6 | 88.59 | 1.434 | 98.56 | 11.4 | 94.51 | 97.7 |
| SD 0.7 | 83.03 | 0.49 | 99.5 | 16.96 | 92.79 | 99.13 |
| SD 0.8 | 74.75 | 0.43 | 99.56 | 25.24 | 89.45 | 99.15 |
| SD 0.9 | 67.91 | 0.34 | 99.65 | 32.08 | 86.72 | 99.27 |
| SD 1 | 60.22 | 0.31 | 99.68 | 39.77 | 83.6 | 99.25 |

| 4ADS | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 99.5338 | 29.3911 | 70.6089 | 0.4662 | 83.06613 | 71.8452 |
| SD 0.1 | 99.223 | 23.4192 | 76.5808 | 0.777001 | 86.33935 | 76.14788 |
| SD 0.2 | 98.5237 | 18.44262 | 81.55738 | 1.476301 | 88.87776 | 80.10107 |
| SD 0.3 | 97.7467 | 13.52459 | 86.47541 | 2.253302 | 91.34937 | 84.48623 |
| SD 0.4 | 96.0373 | 9.42623 | 90.57377 | 3.962704 | 92.95257 | 88.4753 |
| SD 0.5 | 92.69619 | 5.913349 | 94.08665 | 7.303807 | 93.52037 | 92.19474 |
| SD 0.6 | 87.56799 | 3.395785 | 96.60422 | 12.43201 | 92.75217 | 95.10549 |
| SD 0.7 | 81.27428 | 1.75644 | 98.24356 | 18.72572 | 90.98196 | 97.2119 |
| SD 0.8 | 73.58197 | 1.17096 | 98.82904 | 26.41803 | 88.00935 | 97.93175 |
| SD 0.9 | 65.19037 | 0.87822 | 99.12178 | 34.80963 | 84.56914 | 98.24356 |
| SD 1 | 58.35276 | 0.58548 | 99.41452 | 41.64724 | 81.79693 | 98.68594 |

| 6V0C | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 99.24375 | 40.04276 | 59.95724 | 0.756254 | 73.4976 | 56.54624 |
| SD 0.1 | 98.77836 | 34.51436 | 65.48564 | 1.22164 | 76.96314 | 60.04243 |
| SD 0.2 | 98.19663 | 28.92486 | 71.07514 | 1.803374 | 80.42869 | 64.06072 |
| SD 0.3 | 96.68412 | 23.39646 | 76.60354 | 3.315881 | 83.53365 | 68.4514 |
| SD 0.4 | 94.41536 | 18.84545 | 81.15455 | 5.584642 | 85.73718 | 72.45536 |
| SD 0.5 | 91.09948 | 14.9664 | 85.0336 | 8.900524 | 87.13942 | 76.16732 |
| SD 0.6 | 86.96917 | 12.33965 | 87.66035 | 13.03083 | 87.4399 | 78.72565 |
| SD 0.7 | 81.55905 | 10.29322 | 89.70678 | 18.44095 | 86.91907 | 80.62105 |
| SD 0.8 | 74.98546 | 8.918754 | 91.08125 | 25.01454 | 85.55689 | 81.53068 |
| SD 0.9 | 68.58639 | 8.002443 | 91.99756 | 31.41361 | 83.95433 | 81.81818 |
| SD 1 | 62.07097 | 7.116677 | 92.88332 | 37.92903 | 82.29167 | 82.07692 |

| 6V0H | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 99.2986 | 29.46655 | 70.53345 | 0.701403 | 83.74656 | 74.01046 |
| SD 0.1 | 98.49699 | 23.7934 | 76.2066 | 1.503006 | 86.45546 | 77.76899 |
| SD 0.2 | 97.49499 | 17.2735 | 82.7265 | 2.50501 | 89.53168 | 82.6678 |
| SD 0.3 | 95.59118 | 11.09229 | 88.90771 | 4.408818 | 92.01102 | 87.92627 |
| SD 0.4 | 92.88577 | 6.77392 | 93.22608 | 7.114228 | 93.11295 | 92.05561 |
| SD 0.5 | 88.47695 | 3.725656 | 96.27434 | 11.52305 | 92.74564 | 95.25351 |
| SD 0.6 | 83.06613 | 1.862828 | 98.13717 | 16.93387 | 91.2764 | 97.41481 |
| SD 0.7 | 77.65531 | 0.677392 | 99.32261 | 22.34469 | 89.43985 | 98.97829 |
| SD 0.8 | 70.84168 | 0.508044 | 99.49196 | 29.15832 | 86.40955 | 99.15849 |
| SD 0.9 | 64.92986 | 0.42337 | 99.57663 | 35.07014 | 83.74656 | 99.2343 |
| SD 1 | 58.31663 | 0.084674 | 99.91533 | 41.68337 | 80.89991 | 99.82847 |

| 6VAM | | | | | | |
|---|---|---|---|---|---|---|
| | Recall | FPR | Specificity | FNR | Accuracy | Precision |
| SD 0 | 96.69157 | 37.17949 | 62.82051 | 3.308431 | 81.54206 | 75.75251 |
| SD 0.1 | 94.66382 | 29.61538 | 70.38462 | 5.336179 | 83.8785 | 79.3381 |
| SD 0.2 | 92.84952 | 23.07692 | 76.92308 | 7.15048 | 85.86449 | 82.85714 |
| SD 0.3 | 89.11419 | 17.5641 | 82.4359 | 10.88581 | 86.33178 | 85.90535 |
| SD 0.4 | 85.80576 | 12.69231 | 87.30769 | 14.19424 | 86.74065 | 89.03654 |
| SD 0.5 | 81.21665 | 9.102564 | 90.89744 | 18.78335 | 85.86449 | 91.46635 |
| SD 0.6 | 76.09392 | 6.025641 | 93.97436 | 23.90608 | 84.46262 | 93.81579 |
| SD 0.7 | 72.14514 | 4.102564 | 95.89744 | 27.85486 | 83.17757 | 95.48023 |
| SD 0.8 | 64.99466 | 3.076923 | 96.92308 | 35.00534 | 79.73131 | 96.20853 |
| SD 0.9 | 60.08538 | 1.923077 | 98.07692 | 39.91462 | 77.57009 | 97.40484 |
| SD 1 | 54.96265 | 1.666667 | 98.33333 | 45.03735 | 74.88318 | 97.53788 |

**Table 4.3. Standard deviation (SD) cut-offs ranging from 0 to 1 for various proteins to demarcate the inner residues from the outer residues.**The metrics used for evaluation include recall, FPR (false positive rate), specificity, FNR (false negative rate), accuracy, and precision.

els, starting with the cavity that has the geometric center of the protein, if one exists. CICLOP specifically aims to elucidate the inner cavities rather than those found on the exterior.

For the detection of other disconnected cavities, CICLOP searches for any other empty voxel that has not been encountered in the first connected component. Since, in case of disconnected cavities, there exist empty voxels that have not been encountered previously, a random empty voxel from this component is chosen as a starting point and mapping of inner surface residues is repeated until all the empty voxels (and hence cavities) are encountered. We finally define the inner surface atoms as those directly in contact with the empty void of the cavity and lining its surface. Fig. 4.4 depicts the entire approach of cavity detection by CICLOP as a concise flow-chart.

### 4.2.2 Calculation of the cavity diameter profile:

The calculation of the cavity diameter is based on the largest circle that can fit inside the cavity at the given length along the cavity axis. The maximum of the calculated diameters is taken to be the pore diameter for the particular cross-section. The diameter measurements are calculated for each 1 Å slice (Fig. 4.5) but recorded for every 3 Å block as most of the atoms of a protein have a mean van der Waals diameter of $sim$3.6 Å (Pellegrini-Calace *et al.*, 2009*b*), and therefore, the diameter is unlikely to change much at a smaller step size. The cavity diameter is computed by calculating the Voronoi diagram of all the atoms that are identified on the inner surface in the 3 Å block in order to obtain its Voronoi vertices. These Voronoi vertices are points that are equidistant from at least 3 atoms and hence can also possibly be the centers of the largest circle that can fit inside the contour formed by the inner surface atoms. Since the Voronoi vertices can sometimes lie outside the pore, only those vertices that lie inside the convex hull formed by the protein atoms marked to be on the inner surface for the particular slice are considered. For each such Voronoi vertex, the diameter of the largest circle that can fit inside the pore is calculated. The Voronoi diagrams are created for each slice along the Z-axis independently. Since each of these slices is a 2D projection of 1 Å thick cross sections of the protein onto a plane parallel to the XY plane (z = k, where k is a constant), the same plane also acts as the reference plane for the calculation of Voronoi diagrams and is also 2D in nature.

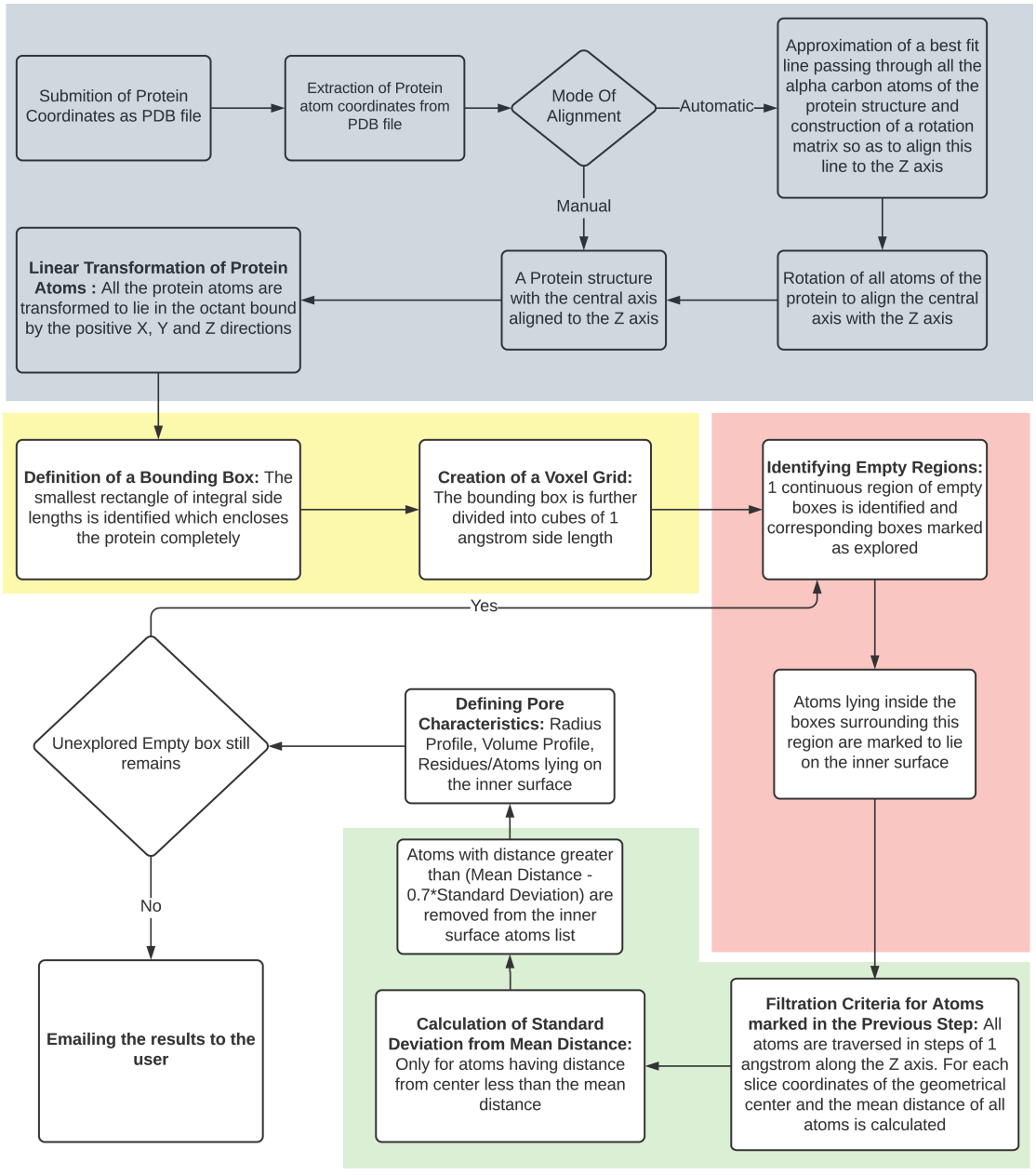



**Figure 4.4. A flowchart describing the functioning of the algorithm for detection of protein cavities by CICLOP**

If $X$ is a metric space with a distance function $d$. Let $K$ be a set of indices and $(P_k)_{k \in K}$ be an ordered collection of non-empty sites in space $X$. The Voronoi Region $(R_k)$, associated with the site $P_k$, is the set of all points in $X$ whose distance to $P_k$ is not greater than their distance to other sites in $X$.

$$R_k = \{x \in X \mid d(x, P_k) \leq d(x, P_j) \textit{ for all } j \neq k\} \tag{4.6}$$

Voronoi vertices are the points where three or more of the Voronoi regions $(R_k)$ intersect. The distance used in our study is the *Euclidean distance* that is calculated between 2 coordinates (x1, y1) and (x2, y2) as follows:

$$d[(x1, y1), (x2, y2)] = \sqrt{(x1 - x2)^2 + (y1 - y2)^2} \tag{4.7}$$

Similarly, a convex hull is calculated for all the atoms lying on the inner surface in the same block. For all Voronoi vertices lying inside the convex hull, the distance between the inner surface atom closest to the Voronoi vertex is marked as a possible radius. The maximum value of all such radii is doubled, and 3 Å is subtracted from that value to account for van der Waals radii from the opposite ends. The value so obtained is the diameter for the given block. This process is repeated for all three blocks of 3 Å and plotted as a diameter along the Z-axis, which is provided as output to the user.

### 4.2.3 Calculation of the cavity volume profile:

The cavity volume is calculated for every 3 Å block along the cavity axis. The total volume of a block is the sum of the volumes of the three individual 1 Å slices cut perpendicular to the cavity axis that make up the block. For every 1 Å slice, its cavity volume value is obtained by first calculating the geometric centre of all the atoms lining the inner surface of that slice. Taking the centre as the reference point, this list of atoms is sorted in a clockwise direction. On joining every unique pair of consecutive points with the centre, a triangle is formed, and the area of all such triangles is calculated and summed. The division of the entire circular slice into smaller triangles that all have the centre of the circle as a common vertex ensures that no two triangles have overlapping areas. The slice volume is measured by multiplying the height of the slice (1 Å in this

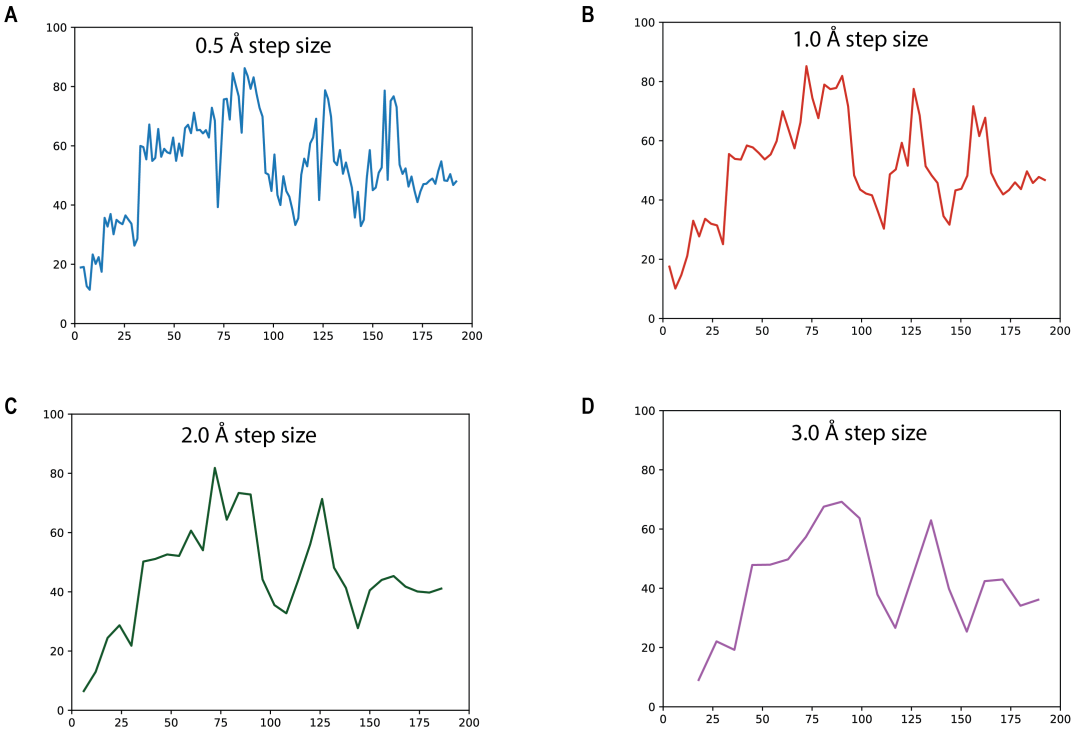



**Figure 4.5. Comparison of the diameter profile generated by CICLOP (using PDBID:1AON) at different step sizes.**The analysis highlights the preference of step-size based on the optimized granularity (1Å).

case) by the total area of the slice.

The area for each triangular region is calculated using the widely used *Heron's Formula* which gives the area of the triangle, $A$, in terms of its three side lengths such that:

$$A = \sqrt{s(s-a)(s-b)(s-c)} \tag{4.8}$$

where s is the semi-perimeter of the triangle given by:

$$s = \frac{1}{2}(a+b+c) \tag{4.9}$$

### 4.2.4 Other modules

**Hydrophobicity of protein cavity:** CICLOP traverses the protein cavity, along the cavity axis. For each 1 Å step, the residues lining the cavity are taken into consideration. The summation of their hydrophobicities, as indicated on the Kyte-Doolittle scale, is

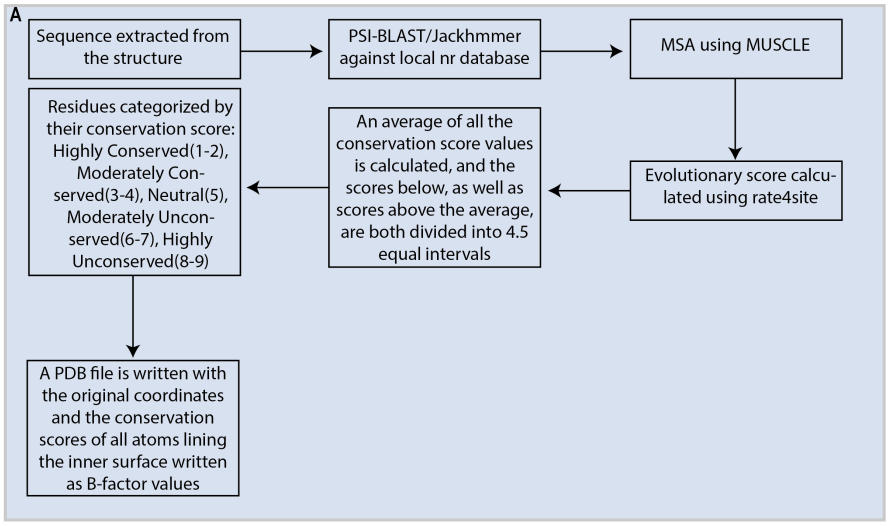
plotted as a function of the distance along the central axis of the cavity.

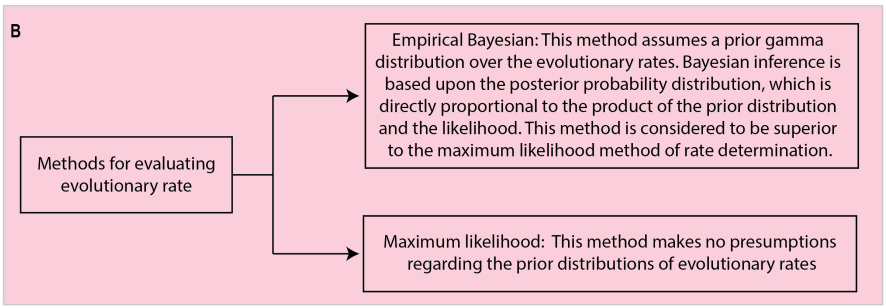
Hydrophobicity values are assigned at the residue level rather than the atomic level, adhering to the standard Kyte–Doolittle scale. Because hydrophobicity values are traditionally defined for entire residues, unlike atomic charges derived from DFT or MD force fields, this approach ensures alignment with established biochemical conventions. The mapping process follows a specific hierarchy to avoid over-counting. Namely, every atom is screened to determine whether it belongs to the inner cavity lining; those that do are assigned a specific B-factor (99999). Hydrophobicity is then mapped to a voxel only if it contains at least one such cavity-exposed atom. To maintain consistency, the hydrophobicity value is added only once per residue per voxel slice. If a single residue spans multiple voxels or extends across several Z-slices (for instance, if the backbone and side chain are separated), the hydrophobicity is recorded only in the first slice where a qualifying atom is encountered.

This method presents a notable limitation: it may not yield an exact representation of hydrophobicity distribution across all Z-slices, as values are not normalized across all voxels a residue occupies. However, because hydrophobicity scales are inherently relative and do not provide ”exact” physical quantities in the same way mass or charge do, this approach represents the most effective and reasonable solution for quantifying surface characteristics within the grid.

**Charge of protein cavity:** Similar to the previous plot, the central cavity is traversed along the central axis. For each Å step, the total number of positively and negatively charged amino acids is plotted as a function of the distance along the cavity axis. Amino acids LYS, ARG, and HIS are taken as positively charged, while ASP and GLU are considered to be negatively charged.

**Evaluating Residue Conservation and Calculation of the Conservation Scores:** Extracting the sequence from the structure, a basic local alignment search is performed using the local nr database. Subsequently, the multiple sequence alignment file is generated, and evolutionary scores are calculated and mapped onto the structure Fig. 4.6 depicts the modules and workflow for calculating conservation scores.




C

| Evolutionary Model | Description |
|---|---|
| JC amino acids | This model assumes that amino acids occur in equal frequency and are substitutedin equal probability. This method is chosen by default |
| Dayhoff Model | The empirical model of amino acid substitution developed by Dayhoff, Schwartzand Orcutt (1978) assumes that all sites in a protein evolve independently. |
| JTT Model | A new and improved version of the Dayhoff model developed by Jones, Taylor and Thornton. This uses a much larger dataset to construct a replacement matrix as proposed in the Dayhoff model |
| REV Model | The is the general reversible Markov process model of amino acid substitution. For this model, the probability of the first amino acid changing to the second is kept the same as thep robability of the second amino acid changing to first |
| WAG Model | A comparatively new model, offering a better statistical fit to the data. This method may provide more accurate estimates of phylogenetic trees than other existing models |
| LG Model | This model is an improvement to the WAG model. It incorporates the variability of evolutionary rates across sites in the matrix estimation and uses a larger and more diverse database than BRKALN (which was used to estimate the WAG model) |
| cpREV Model | A model similar to the REV model but optimised for chloroplast proteins |

**Figure 4.6. Workflow for calculation of conservation scores using CICLOP: A.** Summary of the basic pipeline for calculating evolutionary scores. All the unique chains from the input PDB file are extracted (in case the user does not submit a FASTA file for the same). A basic local alignment search is performed using the PSI-BLAST package Altschul *et al.* (1990) against the nr-database. The sequences are aligned using MuscleEdgar (2004). Alternatively, it is also possible to query the target sequence against the Swissprot databaseBairoch and Apweiler (1996) using jackhmmerJohnson *et al.* (2010) to obtain sequences distantly similar to the target. The rate4site methodPupko *et al.* (2002) is applied to the resulting MSA file to calculate the evolutionary scoresAshkenazy *et al.* (2016). **B.** Methods for estimating evolutionary rateMayrose *et al.* (2004). **C.** The seven substitution models offered by CICLOP used to infer the evolutionary conservation scoresChoudhuri (2014); Yang *et al.* (1998); Whelan and Goldman (2001); Le and Gascuel (2008); Adachi *et al.* (2000).

## 4.3 Methodology

### 4.3.1 Benchmarking CICLOP against other cavity detection tools

CICLOP was benchmarked against a few most commonly used methods for cavity detection. This section summarizes the methodology used to obtain results from the other methods and those of benchmarking.

For PoreWalker, each structure was uploaded on their web server, and the PDB file containing the marked inner atoms was downloaded. All the residues were extracted from the B-factor-loaded PDB using an *in-house* Python program.

For MOLE*online*, each structure was submitted to their web server with the default parameters. The option to "Ignore HETATOMs" was selected. The options "Merge Pores" and "Automatic Pores" were turned on. The results were downloaded and the inner residues extracted using an *in-house* Python script.

For CICLOP, in cases where conservation scores were calculated, the underlying evolutionary model was selected as 'LG' unless otherwise stated. All the conservation scores calculated for this study were obtained through the empirical Bayesian methodology, with no exceptions.

### 4.3.2 Molecular Dynamics Simulation for Quantitative Comparison:

All-atom MD simulations were performed using GROMACS (Abraham *et al.*, 2015), version 2020, to sample truly water-exposed inner residues. OPLS-AA force field (Kaminski *et al.*, 2001) was used to describe the system. SPC/E water model (Mark and Nilsson, 2001*b*) was used to solvate the protein, and sodium ions were added to neutralize the system. All simulations were performed at 298K using the modified Berendsen thermostat (Berendsen *et al.*, 1984*b*) for temperature control and the Verlet cutoff scheme for searching neighboring grid cells. Pressure coupling (coupling time 2.0 ps, isothermal compressibility 4.5e-5) using the Parrinello-Rahman scheme was also employed in all simulations, allowing the lateral and perpendicular pressures to be coupled independently to maintain a constant pressure of 1 bar. The simulations were performed under

periodic boundary conditions, with long-range electrostatic interactions treated using the Particle Mesh Ewald (PME) method. A grid spacing of 0.16 nm was employed, combined with fourth-order B-spline interpolation to compute the potential and forces between grid points. The LINCS algorithm was used to constrain all the bonds. The short-range interactions were cut off at 1.0 nm. The production run for each system was performed for a total of 10 ns, with 2 fs time step used for numerical integration of the equations of motion. All the starting structures were subjected to a minimization protocol of 50000 steps using the steepest descent algorithm, followed by equilibration runs in NVT and then NPT ensembles for 100 ps each.

Each frame of the protein-water simulation generated by the MD protocol described above was analysed separately using *in-house* scripts. All the residues having at least one atom in close proximity (less than or equal to 3.5 Å) to the water group lying inside the cavity were identified as inner residues for the frame in question. The distances for each protein atom from the water group identified were calculated using GROMACS trjorder. Finally, only those residues were considered to be truly inside the cavity, which appeared to be lying on the inner surface in at least 500 frames out of the 1000 frames extracted from the 10 ns simulation, using a time step of 10 ps (except for the 900 frames utilized for calculations for 6V0B).

The final list of residues thus generated served as the ground truth positive set, representing the residues that lie within, while the rest were categorized as the negative set for the experiment. Using the residue lists curated in the aforementioned step, true positives, false positives, true negatives, and false negatives were calculated for CICLOP, PoreWalker, MOLE*online*, and Caver Web. The PDB structures provided as input to the tools were extracted from the simulation, with the frame selected after the protein had reached a stable state. The root mean square deviation (RMSD) of the backbone atoms of the protein was chosen as a measure of the system's stability throughout the simulation length and was calculated using GROMACS' rms. The 4 ns time-step frame's structure was utilized for all other proteins (except 5 ns for 4nmd and 6 ns for 6vai, due to delayed stabilization). The following defines the ground truth and the evaluation metrics:

- Ground truth positive set: The inner residues identified to be in "saturated" contact with water molecules (in at least 5ns of simulation length).
- Ground truth negative set: All the residues that didn't qualify for the ground truth positive set.

- True Positives: Those residues detected by the tool in question as positive (inner lying residue) and also present in the ground truth positive set.
- False Positives: Those residues detected by the tool in question as positive (inner lying residue) but not present in the ground truth positive set.
- True Negatives: Those residues detected by the tool in question as negative (not an inner lying residue) and also present in the ground truth negative set.
- False Negatives: Those residues detected by the tool in question as negative (not an inner lying residue) but not present in the ground truth negative set.

Finally, for the complete trajectory, accuracy, precision, recall, specificity, false positivity rate, and false negativity rate were calculated for each of the tools using the following formulas:

$$Recall = \frac{TP}{TP + FN} \quad , \quad Specificity = \frac{TN}{TN + FP} \tag{4.10}$$

$$FPR = \frac{FP}{FP + TN} \quad , \quad FNR = \frac{FN}{TP + FN} \tag{4.11}$$

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad , \quad Precision = \frac{TP}{TP + FP} \tag{4.12}$$

FPR: False Positivity Rate, FNR: False Negativity Rate

### 4.3.3 Cavity characterization of proteins within the BPI family

Four proteins of the BPI/LBP family were considered: CETP (PDB ID: 2OBD), BPI (PDB ID: 1EWF), LBP (AF-P18428-F1), and PLTP (AF-P55058-F1). The structures were automatically aligned and submitted to the CICLOP web server. The results were visualized in Chimera (Pettersen *et al.*, 2004). For a dynamic characterization of their volume and hydrophobicity, an MD simulation was performed for 500 ns, and 150 frames post-stabilization were considered, spaced every 1 ns.

### 4.3.4 Identification of CETP cavity lining and residue interaction network

CICLOP was used to identify residues lining the inner cavity in the CETP simulation. A consensus list of residues lining the cavity was obtained, containing those residues

that lined the cavity for 80% of the simulation time. Hydrophobic interactions amongst the cavity lining residues and between PL acyl tails and cavity lining residues were calculated using MDAnalysis (Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016) within 5 Å through an in-house Python script. A hydrophobic interaction (non-directed) network was constructed, where the cavity-lining residues were represented as nodes and connected through edges based on their dynamic interactions throughout the simulation. Furthermore, the edge weight was assigned based on the persistence of a hydrophobic interaction, which was calculated as the number of frames in which a hydrophobic interaction exists between a given residue pair divided by the total number of frames. The network was visualized using Cytoscape (Shannon *et al.*, 2003).

### 4.3.5 Conservation profile of CETP cavity lining residues

Conservation Scores were calculated using CICLOP, where PSI-Blast was used for homologue detection with an e-value cutoff of 10. Empirical Bayesian was used as the rate inference method while JC amino acids were used as the underlying evolutionary model. The conservation scores computed were then mapped as a network parameter and visualized using cytoscape.

### 4.3.6 Volume profiles of CETP in presence and absence of PL plugs

Simulation frames for CETP with and without its PL plugs were extracted every 1 ns post-stabilization using gmx trjconv and aligned with respect to the last frame. The volume of the tunnel and hydrophobicity were calculated using CICLOP (Garg *et al.*, 2022).

## 4.4 Results

### 4.4.1 Protein cavity analysis using CICLOP

CICLOP maps and identifies the different residues that line the inner surface of any cavity-containing protein. It is implemented as both a web server and an executable.

Each job on the server requires the user to provide their email address along with the PDB file(s) of the proteins they wish to characterize. Using the input from the PDB file format, the method operates in two modes of protein alignment, normal to the cavity axis. In the automatic mode, the algorithm rotates the input structure such that its central cavity axis lies along the Z-axis while in the case of manual mode, the same is assumed. The resultant B-factor-loaded PDB output file marks the detected inner surface residues of the cavity. This is followed by estimating the total cavity diameter and volume, which are calculated by summing the areas enclosed by all the inner lining atoms. Both the diameter and volume are plotted along the length of the cavity and are provided to the user. Moreover, the user is also provided with a PQR-formatted file that can be used to visualize the volume occupied by the internal cavity in 3D space.

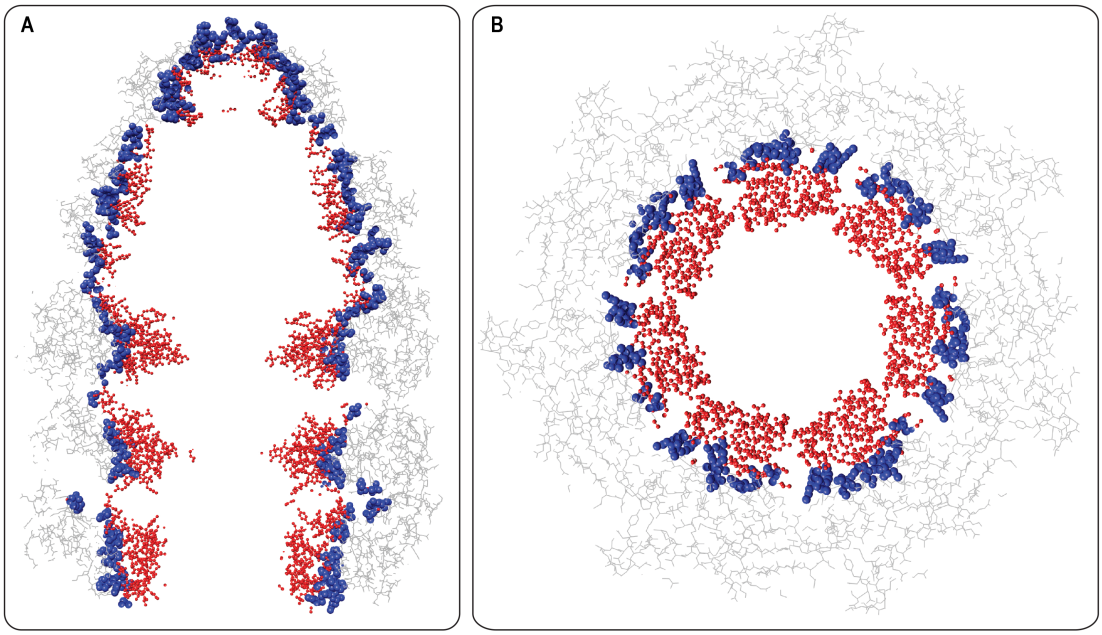



**Figure 4.7. Resolution of cavity detection. A.** Longitudinal section and **B.** Transverse section of the asymmetric GroEL/GroES/ADP complex (PDBID: 1AON). The bulk protein residues(gray wires) with residues that have 25-75% of their atoms detected to lie on the inside surface (blue). Residues with 100% atoms lining the cavity are marked in red.

To comprehensively annotate protein cavities, CICLOP offers an extensive suite of user-friendly features and modules. Our tool can characterize internal cavities at both the residue and atomistc resolution (Fig. 4.7). In addition to the detection of cavities (Fig. 4.8 A), CICLOP computes conservation attributes of cavity-lining residues, which are normalized for comparison between proteins (Fig. 4.8 B). Furthermore, the distribution of inner residues based on their conservation score along the axis of the detected cavity (Fig. 4.8 C) is also provided. Additionally, CICLOP assigns the secondary struc-

ture for each residue lining the cavity(Fig. 4.8 D).

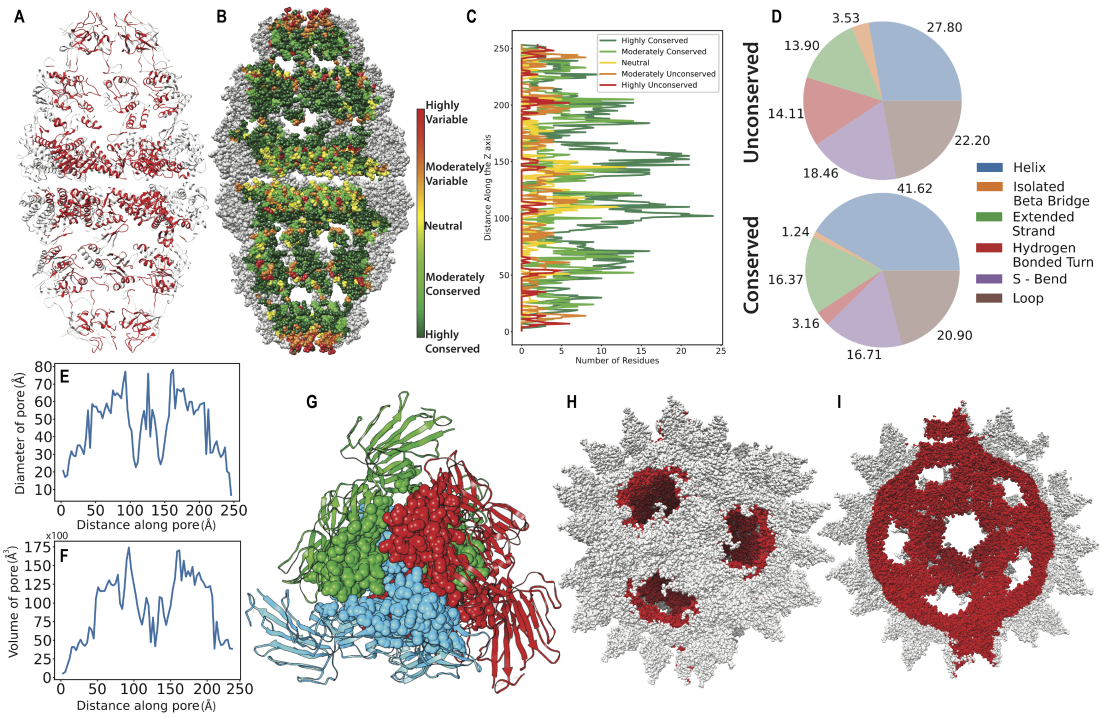



**Figure 4.8. Application of CICLOP. A.** Identification of the residues lining the inner surface of WT human mitochondrial chaperonin (ADP:BeF3)14 complex (PDB ID: 6HT7). **B.** Conservation of the identified residues; marked in the range of highly variable (red) to highly conserved (dark green). 57.83 & 16.59% of the residues lining the cavity are highly & moderately conserved, respectively. **C.** Conservation profile as a function of Z-axis **D.** Secondary structure assignment of the conserved and unconserved residues as detected by CICLOP. **E.** Diameter profile of the cavity of the protein as a function of Z-axis distance. **F.** Volume profile of 6HT7 as a function of Z-axis distance (Total cavity volume = 738,319.197 Å$^3$). **G.** Top view of the oligomerization interface residues (forming a cavity), of the three chains (highlighted in red, blue and green) of the alpha-coronavirus spike glycoprotein (PDB ID: 6IXA), as detected by CICLOP and represented as spheres. **H.** Outer surface of Human parechovirus (HPeV) protein complex (PDB ID: 4UDF) with inner surface marked red and **I.** longitudinal section of the inner surface (red) as detected by CICLOP.

This method also generates a diameter and volume profile of the cavity detected in the input structure. These profiles are provided along with the output summary file. Fig. 4.8 E, F show the diameter and volume profile of the multi-subunit chaperone complexes, human and bacterial in origin, respectively. Additionally, CICLOP facilitates the functional characterization of the cavity by providing a charge distribution analysis, as well as a hydrophobicity plot along the cavity's length. CICLOP can also detect cavities at the interface of multimers, such as those formed by the homo-trimeric arrangement in the spike protein of the SARS-CoV-2 virus (Fig. 4.8 G), consequently allowing for the identification of oligomerization domains in these proteins. The robustness of CICLOP

is highlighted by its ability to effortlessly characterize proteins as large as the human parechovirus envelope (HPeV) containing 302,100 atoms arising from 38,580 residues (Fig. 4.8 H & I).

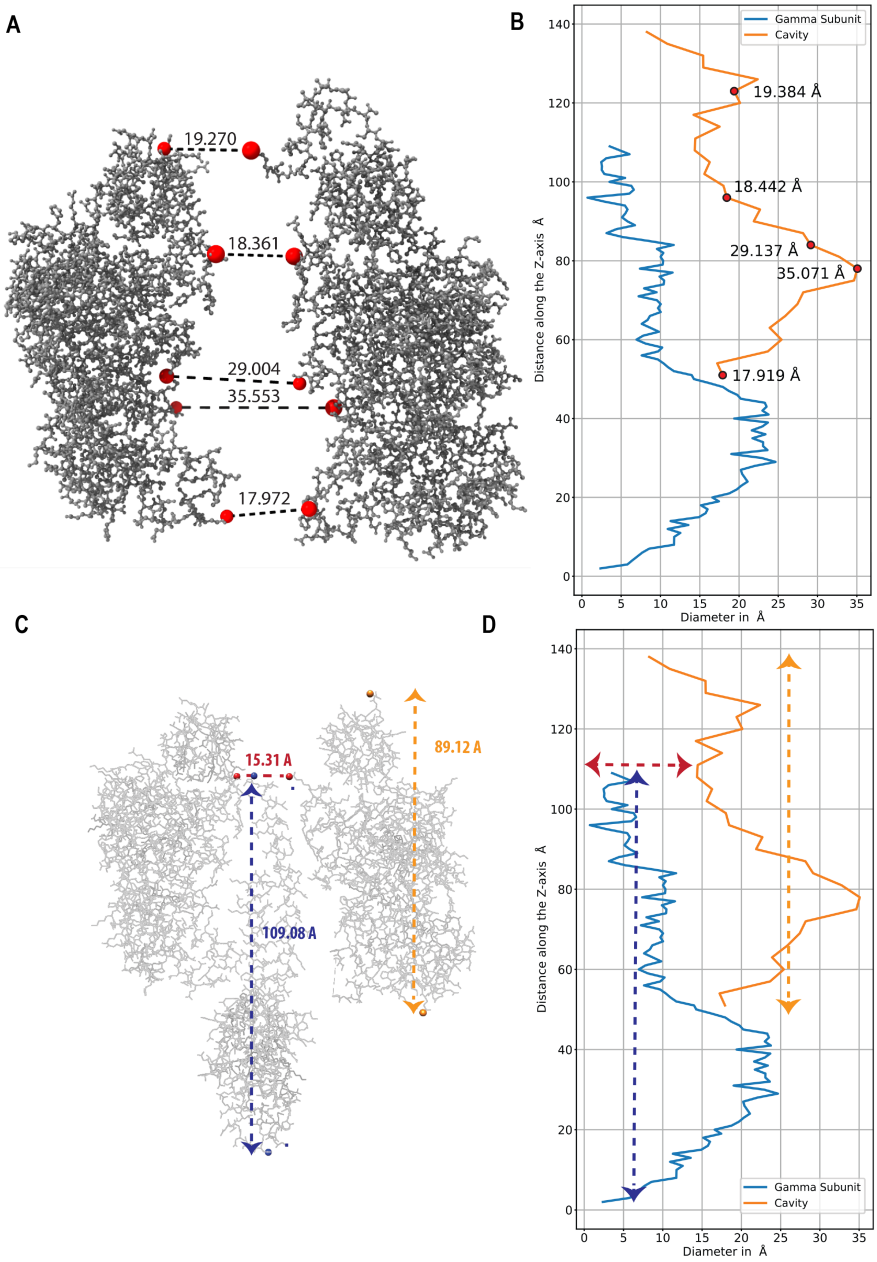



**Figure 4.9. Validation of the size of the cavity determined by CICLOP through manual annotation (using UCSF Chimera) A.** Validation of the diameter profile generated using CICLOP through manual annotation (using UCSF Chimera). **B.** Diameter profile generated by CICLOP showcasing the distances between atoms lying on the opposite ends of chains A and F of State 1A (left). **C.** Validation of the size of the cavity. Distances between the atom with the minimum and maximum Z-axis value are marked (in yellow) for chain B (representative of the cavity) as well as chain G ( $\gamma$ subunit) of state 1A. **D.** The corresponding values are indicated in the diameter profile generated by CICLOP.

Next, to validate the diameter profile generated by CICLOP, we characterized a

unique cavity, that of the F1 domain of the bovine mitochondrial ATP synthase. ATP synthases consist of two functional domains. The F1 domain is formed by a hexameric cap that fits onto a rotor (Jonckheere *et al.*, 2012). We extracted (from the crystal structure) the measurement of the distance between pairs of atoms of the two chains (that are roughly diametrically opposite to each other), which encompasses the 'crown' (Fig. 4.9 A), and compared it with the diameter profile created by our tool. Fig. 4.9 B illustrates the high-resolution diameter profile generated by CICLOP. The diameter for each slice obtained by CICLOP relative to the cavity's physical measurement falls within the error limits (mean error = 0.053). Similarly, Fig. 4.9 C & D depict the length and width of the cavity inferred from the diameter profile. These values align well with the physical measurements taken on the structure and explain the rotor's tight fit within the cavity. Together, these results highlight the sensitivity of diameter measurements computed by our tool and how they precisely match the physical measurements of the protein structure obtained from the crystal structure.

### 4.4.2 Benchmarking CICLOP against other existing tools.

CICLOP has been tested on a large subset of cavity-containing proteins varying in size and shape. Cavities that are curved, disconnected, or even ramified are efficiently detected by this method.

Next, CICLOP's performance was evaluated against PoreWalker & MOLE*online*, two of the widely used cavity detecting tools. A set of four proteins [a cytosolic chaperone (1AON), a group II chaperonin (3LOS), Nicotinic AcH receptor (2BG9), and Alpha-hemolysin (7AHL)] from different organisms varying in size, sub-cellular localization, cavity conformation and function were chosen and used as an input for CICLOP and the other methods. Using only the default parameters for each tool, the inner residues for each input protein were obtained. These were then mapped onto the protein structure (coloured red) and have been displayed in Fig. 4.10. It was observed that residues detected by CICLOP formed a continuous contour along the cavity and are in line with the protein's overall structure compared to the other methods (Fig. 4.10 A & B). Specifically, MOLE*online* was unable to detect a majority of residues lining the cavity (Fig. 4.10 C), and both PoreWalker and MOLE*online* had boundary issues or mouth opening ambiguity (defined as residues visibly on the outer surface but detected to be on the inner

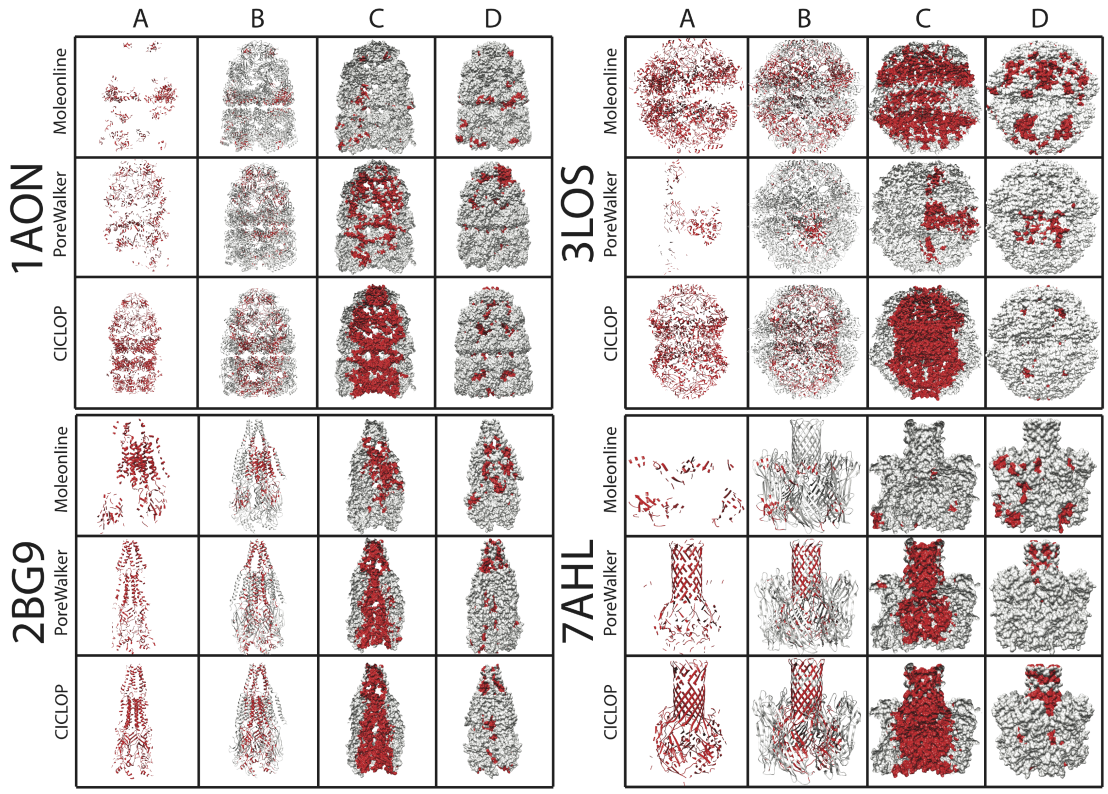
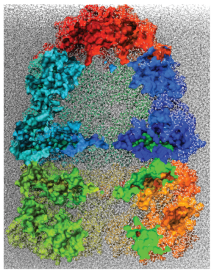


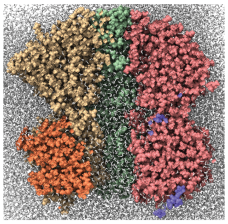

**Figure 4.10. Qualitative comparison of detection of inner-lining residues within a protein cavity with other methods.**The residues identified by various methods are marked in red. **A.** The inner surface contour as detected by the different methods. **B.** The inner surface marked (red) for the protein in ribbon representation for different methods. **C.** A longitudinal section of the inner surface (red) of the proteins in surface representation for different methods. **D.** The complete surface as seen from the outside for the corresponding figure in panel c for different methods.

surface by a method) (Fig. 4.10 C & D). Additionally, CICLOP was efficiently able to detect the cavity of a completely closed chaperonin (3LOS) while detections by both PoreWalker & MOLE*online* were majorly misplaced.



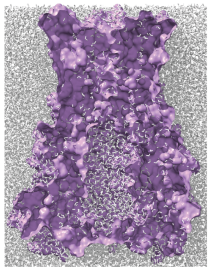
1AON

| | TP | FP | TN | FN | Accuracy | Precision |
|---|---|---|---|---|---|---|
| CICLOP | 80.31% | 5.51% | 94.48% | 19.68% | 89.07% | 90.01% |
| PoreWalker | 29.44% | 5.08% | 95.15% | 70.55% | 70.04% | 78.16% |
| MoleOnline | 6.69% | 9.10% | 90.89% | 93.30% | 58.71% | 31.25% |



1TF7

| | TP | FP | TN | FN | Accuracy | Precision |
|---|---|---|---|---|---|---|
| CICLOP | 85.58% | 5.01% | 94.98% | 14.41% | 91.52% | 90.85% |
| PoreWalker | 17.88% | 3.97% | 97.16% | 82.11% | 68.00% | 72.30% |
| MoleOnline | 31.64% | 11.38% | 88.67% | 68.35% | 67.69% | 61.79% |



6V0B

| | TP | FP | TN | FN | Accuracy | Precision |
|---|---|---|---|---|---|---|
| CICLOP | 72.71% | 0.69% | 99.30% | 27.28% | 85.22% | 99.15% |
| PoreWalker | 29.62% | 0.55% | 99.44% | 70.37% | 62.48% | 98.36% |
| MoleOnline | 44.69% | 14.44% | 85.55% | 55.30% | 63.92% | 77.68% |

**Figure 4.11. Validation of the inner residues detected by CICLOP using MD simulation.** A snapshot of the longitudinal section of the simulated system of 1AON, 1TF7 & 6V0B is shown in surface representation while the water molecules are shown in grey. The table displays the accuracy and precision of CICLOP and other methods in detecting residues lining the inner cavity of 1AON, 1TF7 & 6V0B.

To further verify the validity of the residues detected on the cavity's inner surface, we performed an all-atomistic molecular dynamics simulation. The purpose of the MD simulation experiment was to sample residues in a protein PDB that were truly water-exposed. For this, the dynamics of bulk water were sampled, while ensuring the protein structure didn't drastically alter during the course of the simulation. The analysis was performed after the protein had been stabilized in the aqueous medium. Several protein-water systems were set up and simulated for 10 ns at 298K to saturate the contacts between the protein and water. Simulations were performed on proteins varying in size cavity conformations, and physiological functions, and CICLOP consistently performed well, detecting inner residues with an accuracy of 85.22% (6V0B) - 91.52% (1TF7) and precision (90.01% (1AON)- 99.15% (6V0B) (Fig. 4.11).

In order to quantitatively measure the efficiency of each method in detecting inner residues, the list of true positives obtained from these all-atomistic molecular dynamic simulations was extended to the list of residues lining the cavity obtained by PoreWalker & MOLEonline. PoreWalker gave an accuracy ranging between 62.48% (6V0B) - 70.04% (1AON) and precision ranging between 72.30% (1TF7) - 98.36% (6V0B) (Fig. 4.11). MOLE*online* performed similarly, giving an accuracy between 58.71% (1AON) - 67.69% (1TF7) and precision between 31.25% (1AON) - 77.68% (6V0B) (Fig. 4.11). These results highlight that the performance of each of these methods fluctuates significantly with the input structures and, hence, the size, structure, and cavity conformation of the input protein. While a certain amount of structure-specific performance can be anticipated for each method, CICLOP is consistent in the magnitude of its detection accuracy. Together these results indicated that CICLOP allows for an in-depth characterization of protein cavities, independent of size and conformation. This tool is built on a novel algorithm that imparts it with unprecedented accuracy and reproducibility, allowing it to outperform its predecessors.

### 4.4.3 CICLOP's utility in characterizing cavities within BPI/LBP family

Next, the cavity within four proteins belonging to the BPI/LBP family —BPI, CETP, LBP, and PLTP—was identified using CICLOP. We observed that all proteins in this family contained a continuous cavity that extended along the entire length of the protein. Irrespective of the linear LBP or the bent shape of CETP and PLTP, a continuous cavity of length 108 Å was detected. Considering the proteins share a similar fold containing two hollow beta barrels connected by a central beta strand-rich region, this continuous tunnel was expected. (Fig. 4.12 A).

Further, the volume of the tunnel was calculated using CICLOP. To account for the dynamic fluctuations in the cavity due to changes in protein conformations, each of the four proteins was simulated for 500 ns. A total of 150 frames, sampled every 1 ns post-stabilization, were used for cavity identification and calculation of its volume (Fig. 4.12 B). It was observed that the cavity within CETP had a maximum volume, especially in its C-barrel and neck region. BPI and PLTP both had a volume of ~350 $Å^3$ throughout the tunnel, which increases in the C-barrel region of BPI. LBP had the smallest cavity

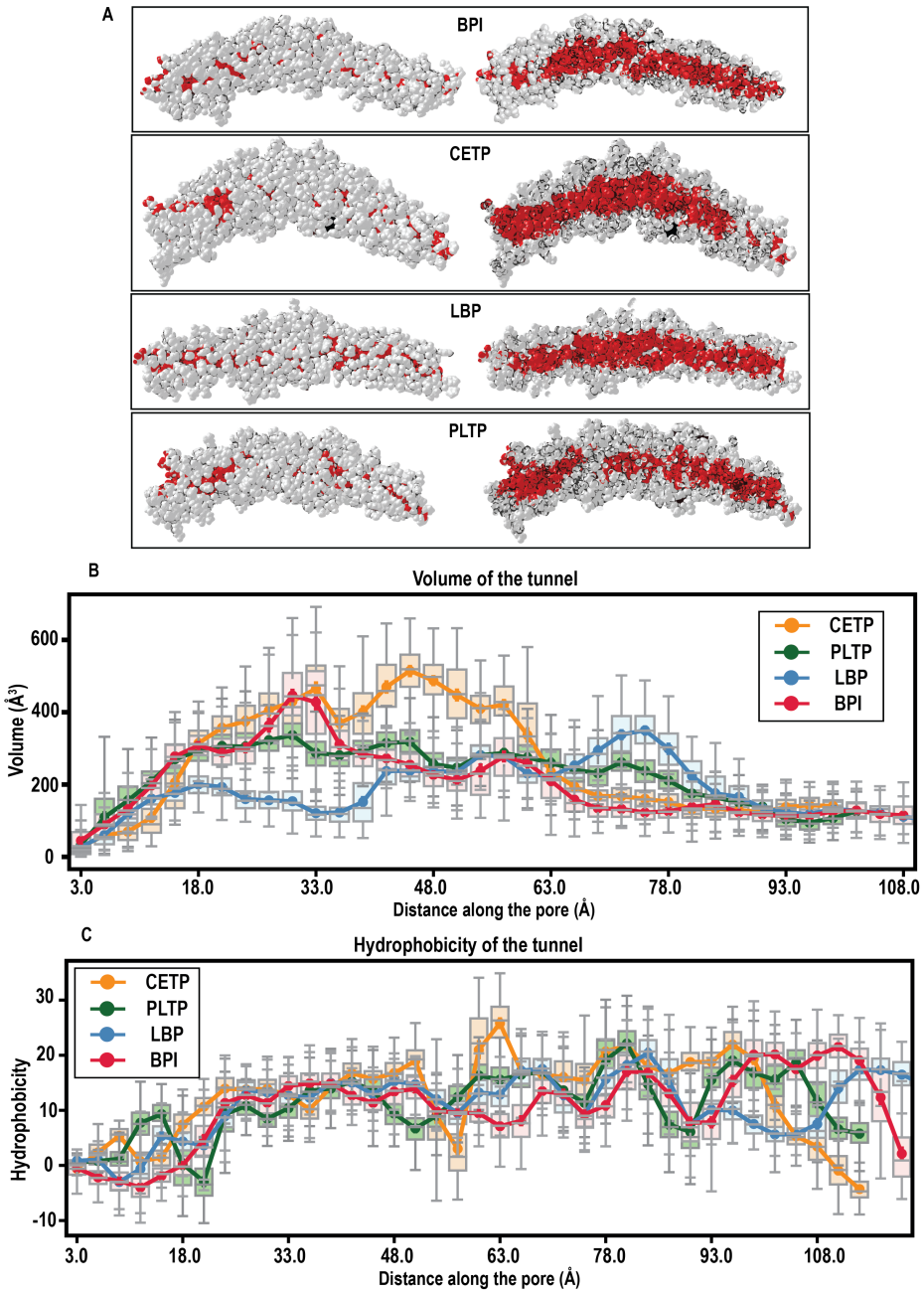



**Figure 4.12. Characterizing notable members of the BPI/LBP family of proteins.** **A.** cavity within proteins of BPI/LBP family. The bulk protein residues are rendered as white spheres, while residues lining the inner cavity of BPI, CETP, LBP and PLTP are shown as red spheres. A horizontal section of the protein is shown for a better visualization of the cavity. **B.** Volume and **C.** hydrophobicity profile of BPI/LBP family proteins. Simulation frames post stabilization were aligned such that the tunnel within BPI, CETP, LBP & PLTP aligned along the z-axis. These structures were used to calculate volume and hydrophobicity using CICLOP. The box plot illustrates the distribution of volume and hydrophobicity values calculated over 150 frames, spaced 1 ns apart (post-stabilization). Solid lines depict the median value of the distribution for the distance along the tunnel axis.

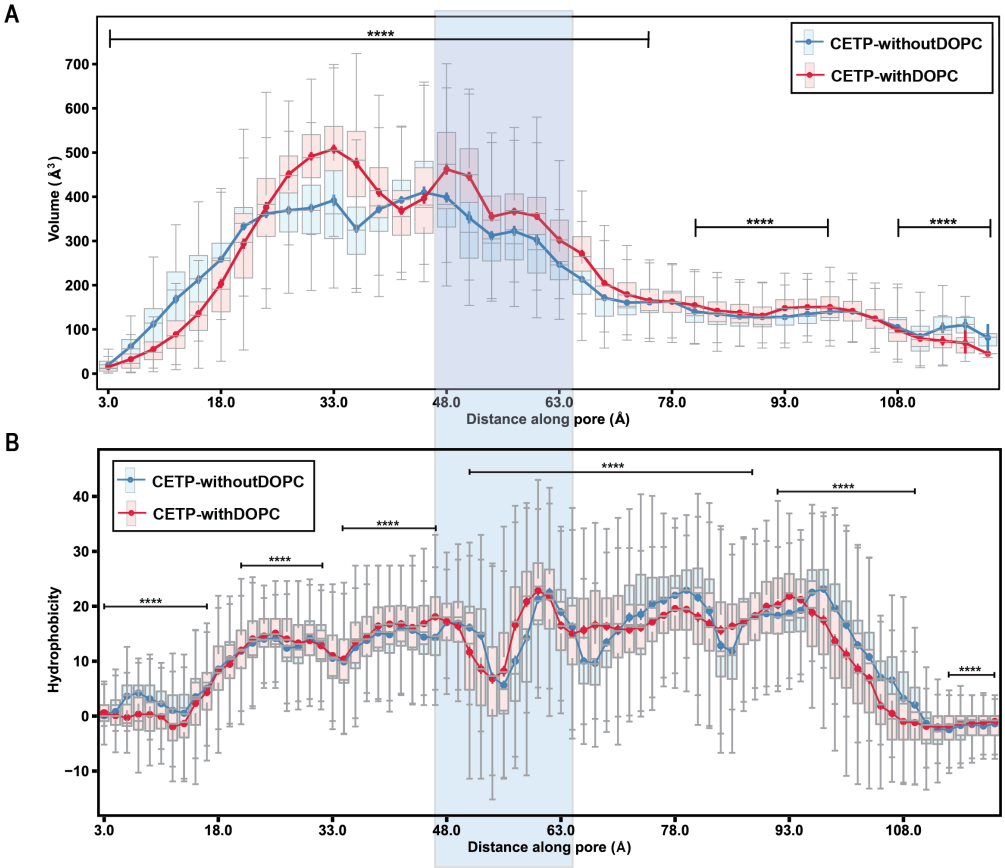



**Figure 4.13. Characterizing the CETP tunnel. A.** volume profile and **B.** hydrophobicity profile of CETP in the presence and absence of PL plugs from C- to N-barrel (left to right). Simulation frames post stabilization were aligned such that the CETP tunnel aligned along the z-axis. These structures were used to calculate volume and hydrophobicity using CICLOP. The box plot illustrates the distribution of volume and hydrophobicity values calculated across 150 frames (post-stabilization). Solid lines depict the median value of the distribution for the distance along the tunnel axis. Mann-Whitney U test was used to calculate the statistical significance of the difference between 1-ns intervals (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

volume amongst the BPI/LBP family proteins of ~200 $Å^3$, which increases to ~380 $Å^3$ towards the end of N-barrel opening (Fig. 4.12 B). Further, the hydrophobicity profile along the cavity axis, which is calculated based on the side chains of residues lining the cavity, showed that all four proteins have a predominantly hydrophobic tunnel (Fig. 4.12 C).

### 4.4.4 Characterizing the cavity within CETP in the presence of PL plugs

To decipher the role of PL plugs in maintaining CETP cavity architecture, the structure of CETP (PDB ID: 2OBD) was simulated with and without its PL plugs. Structures were sampled every 1 ns post stabilization, and they showed a significant increment in the volume of the tunnel within the neck and the adjacent C-barrel region of CETP in the presence of PLs (Fig. 4.13 A). The tunnel hydrophobicity, computed based on the hydrophobicity score of the side chains exposed within the tunnel, remained unchanged for these conformations of both systems (Fig. 4.13B). Therefore, PL plugs altered the cavity architecture by increasing the cavity volume.

Next, using the list of inner cavity lining residues identified by CICLOP in the last frame of the simulation, a post-stabilization hydrophobicity interaction network was constructed. The weighted network depicts the CETP cavity architecture. Each node in the network represents a cavity lining residue, and its size is proportional to the number of connections it makes with the other nodes. These nodes are connected to other nodes through edges, representing hydrophobic interactions < 5 Å. The thickness of each edge depicts the persistence of a pair of hydrophobic interactions throughout the simulation. F265, M412 had the highest degree of forming hydrophobic interactions with 16 other nodes, closely followed by I413, F93 & F263, which formed connections with 15 other nodes. The two PLs (DOP487 and 488) were observed to act as hub nodes, having a very high degree, and formed a maximum number of hydrophobic interactions (> 30) within the tunnel (Fig. 4.14). Further, the betweenness centrality of each node was calculated to highlight those nodes that are positioned on the shortest paths connecting different regions of the protein. These regions often maintain connectivity and facilitate interaction within the hydrophobic network. F35, V30, F93, I183, I418, F471, and the

two PL plugs (DOP487 and 488) were observed to have high betweenness centrality (Fig. 4.14).

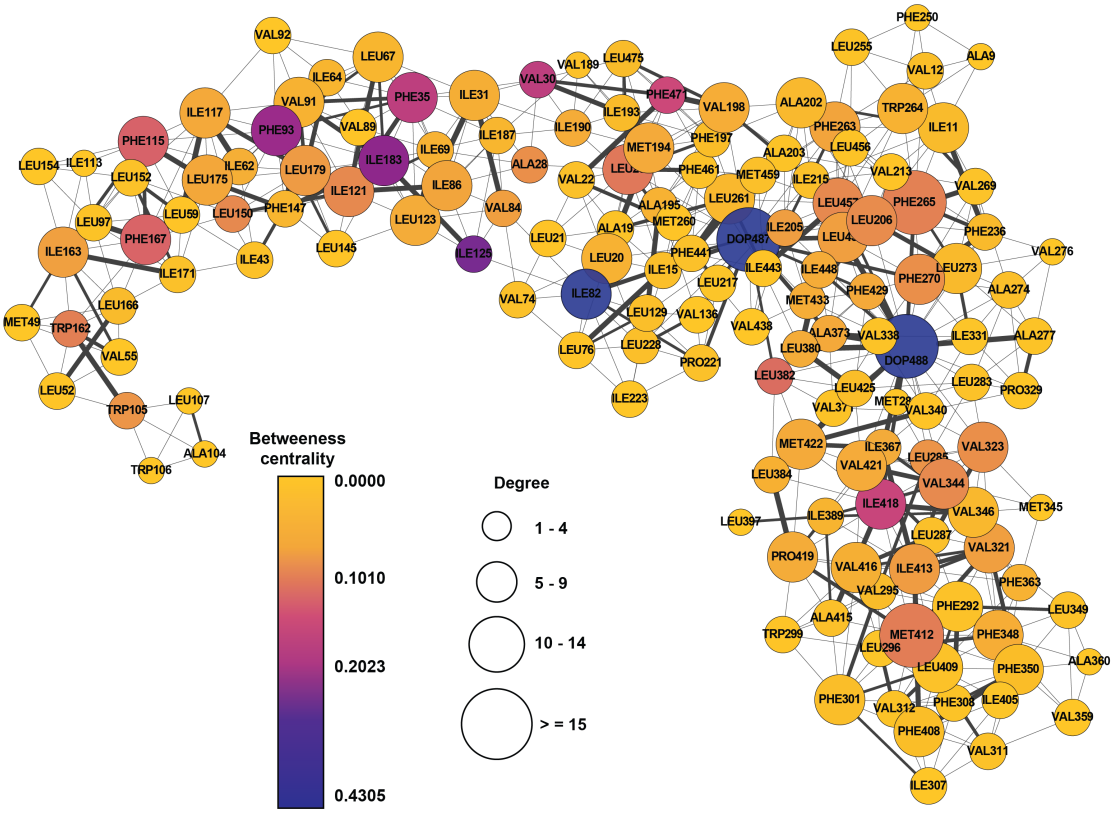



**Figure 4.14. Tunnel architecture of CETP.**
A weighted network depicting hydrophobic interactions within 5 Å in the CETP tunnel. Each node in the network represents a cavity lining residue, and its size is proportional to the number of connections it makes. Each connection or edge depicts hydrophobic interactions, and its thickness represents the persistence of interactions between pairs of nodes throughout the simulation.

## 4.5 Discussion

Advances in biophysical techniques have led to remarkable insights into the structural and dynamical properties of proteins, allowing for an understanding of their morphology, nature, and mechano-functional roles (Dobson, 2019). Despite all these efforts, internal protein cavities, a predominant feature of most proteins, have been largely overlooked due to their relative inaccessibility, which makes them experimentally challenging to characterize and study. Computational approaches can offer an ingenious solution to bridge this gap by speedily and accurately identifying residues lining the cavity, a prior step in their functional characterization. While most methods work well for identifying the location of a cavity, an accurate characterization of atoms/residues that surround

internal cavities is not given a priority. Considering the physico-chemical properties of cavity lining residues impart functionality to the cavity, their accurate detection and characterization are a priority for CICLOP. Therefore, in addition to the diameter and volume of the cavity along its length, the overall nature of the cavity is discerned while also reflecting upon its secondary structure makeup and evolutionary conservation.

Cavity identification by CICLOP is automatic, requiring no user intervention, such as seed residues or active cavity residues, to initiate the cavity search. This, in turn, makes it better than methods like HOLE (Smart *et al.*, 1996), MOLE (Sehnal *et al.*, 2013), and CAVE (Buša *et al.*, 2010). Furthermore, the algorithm is able to avoid pitfalls related to boundary issues, making it superior to PoreWalker (Pellegrini-Calace *et al.*, 2009*a*) and MSpocket (Zhu and Pisabarro, 2011), which allow for a distinction between the cavity openings and the protein boundary. The algorithm does not suffer from issues due to protein orientation, which has been rigorously tested. However, the user is still provided with a choice for manual alignment in case a specific orientation is desired.

Additionally, grid-based methods are known to suffer from grid spacing sensitivity, such that a certain list of cavities is only identified at a particular grid spacing while others are missed (Simões *et al.*, 2017). This impacts their accuracy in cavity detection, as well as the space and time complexity. CICLOP's algorithm is optimized to detect all empty voxels; therefore, it continues running until all empty spaces have been identified. Moreover, it has been optimized at a grid dimension of 1.0 Å and has been validated and benchmarked extensively to show that no cavity is missed at this grid dimension. Even without fluctuating the grid dimension spacing, CICLOP outperforms PoreWalker (Pellegrini-Calace *et al.*, 2009*a*), Hole (Smart *et al.*, 1996), and MOLE (Sehnal *et al.*, 2013)for larger proteins containing more than 103,824 atoms in terms of computation time without compromising on the accuracy of detection.

One major limitation of the tool is that it can only detect internal cavities, like channels, tunnels, pores, and voids. It tends to perform poorly on small globular proteins with surface cavities or clefts. Additionally, CICLOP allows for visualization of the cavity and does not automatically determine the number of cavities identified, like in MOLE (Sehnal *et al.*, 2013) or Fpocket (Le Guilloux *et al.*, 2009). Rest assured, the discontinuous nature of cavities is preserved during identification, and the same can be verified by direct visualization or through the diameter profile.

Characterization of cavities within proteins of the BPI/LBP family showed that all four of BPI, CETP, LBP, and PLTP contain a continuous hydrophobic cavity that extends through the protein, while the cavity within CETP is the largest. Considering CETP is implicated in the exchange of two neutral lipids (CE and TG) between lipoproteins, the hydrophobic nature and the larger size of the cavity are justified. Further, the presence of PL plugs increases the volume of the cavity within the neck region. More importantly, these PLs extensively interact with tunnel residues through their acyl tails, resulting in their high betweenness centrality. These PLs may act as hub nodes, playing a crucial structural and functional role in maintaining the CETP tunnel's hydrophobic architecture.

In summary, CICLOP's ability to quantitatively and qualitatively characterize the internal cavity of proteins has been demonstrated. The cavities identified by CICLOP have been benchmarked against existing tunnel detection methods, revealing a more sensitive, precise, and accurate performance by CICLOP. The atomistic detail provided by CICLOP can have significant application in the field of structural biology in evaluating protein structures, identifying solvent-accessible surfaces, protein oligomerization studies, understanding chaperone-assisted protein folding, as well as functional characterization of channels. The method is available at https://ciclop.raylab.iiitd.edu.in .

# CHAPTER 5

# Mechanism of Triglyceride Transfer

## 5.1 Introduction

Cholesteryl ester transfer protein (CETP) is the only known protein that mediates the transfer of cholesteryl esters (CEs) from high-density lipoproteins (HDLs) in exchange for triglycerides (TGs) from low-density lipoproteins (LDLs), intermediate-density lipoproteins (IDL), or very low-density lipoproteins (VLDLs) (Morton and Zilversmit, 1983). This hetero-exchange results in CE depletion and TG enrichment of HDL, thereby leading to its rapid catabolism (Tall, 1993; Lewis and Rader, 2005). To date, two models have described the mechanism of neutral lipid transfer through the CETP tunnel. CETP may act as a shuttle, carrying lipids from donor to acceptor lipoproteins (forming a binary complex) (Sacher *et al.*, 2023). Because of the protein's characteristic concave curvature, it was hypothesized that CETP bound the convex surface of the lipoproteins, such that neutral lipid substrates entered through the central openings as opposed to the terminal. As CETP detaches from the surface of lipoproteins, it takes two phospholipids from the surface of lipoprotein to seal its tunnel according to this model (Qiu *et al.*, 2007).

Alternatively, CETP may form a bridge between two lipoproteins simultaneously (forming a ternary complex) (Sacher *et al.*, 2023). However, upon incubation of CETP with HDL and VLDL, both binary and ternary complexes were observed (Zhang *et al.*, 2012, 2015) showing CETP penetrating through the surface of lipoprotein (penetration depth: HDL (~ 30 Å ± 10 Å), LDL (~ 25 Å ± 10 Å) and VLDL ( ~ 20 Å ± 10 Å). Moreover, MD simulations confirmed the partial penetration of upright CETP into HDL. Further, these simulations showed that CE could indeed migrate from the core of HDL to the terminal opening of CETP (Cilpa-Karhu *et al.*, 2015). It was also proposed that binding of CETP to one lipoprotein may initiate a conformational change that elicited binding to the second lipoprotein with increased affinity (Zhang *et al.*, 2017). This tunnel mode of transfer, however, could not explain the presence of PLs plugging the central pores, especially since the PL-bound orifices lay 15 Å away from the surface of HDL

in the CETP-HDL complex (Zhang *et al.*, 2012). The entry of lipids into the tunnel is proposed to be guided by a concentration gradient (Zhang *et al.*, 2015) while CEs and TGs in the surface of lipoproteins may themselves compete amongst each other to gain entry into the tunnel (Cilpa-Karhu *et al.*, 2015; Morton and Greene, 2003*a*). What regulates the movement of lipid once it enters the CETP tunnel hasn't been studied to date, although the role of simple diffusion is presumed. No additional proteins have been reported to facilitate lipid entry into the tunnel.

In the absence of biophysical techniques that allow for direct visualization of lipid transfer, MD simulation-based approaches have been widely used to study the mechanism of lipid transfer mediated by CETP. All of these studies have notably focused on the movement of CE through the tunnel. Atomistic simulations of an upright CETP-HDL complex showed that the immobilization of CETP on the HDL surface is promoted by POPC, apoA1, and the presence of E2 fatty acyl esters (Cilpa-Karhu *et al.*, 2015). Next, its penetration is guided by Trp-rich flaps at the N-terminal barrel region, and CE migrates from the HDL core to the mouth of the N-barrel. Upon reaching the opening of the tunnel, CE enters the tunnel through an aromatic-rich region consisting of Phe (Cilpa-Karhu *et al.*, 2015). Interestingly, Cilpa et al. hypothesized that CE bound in the tunnel stabilizes it, and that they may be exchanged or remain bound, while CE's from the HDL core are passing through the tunnel, implying the tunnel is wide enough to accommodate a wide variety of lipids (Cilpa-Karhu *et al.*, 2015). The authors also commented on the role of bound phospholipids as an essential structural component that is not exchanged (Cilpa-Karhu *et al.*, 2015). While insightful, this study assumed that CE molecules are transferred from the core of HDL; however, it has been shown that the lipid transfer rates of CETP are dependent on the concentration of surface-available CE and TG (Morton and Greene, 2003*a*). Further, the surface-oriented neutral lipids can reduce the surface tension of lipid droplets (Kim *et al.*, 2021). Moreover, this study assumes a simplistic structure of spherical HDL (sHDL), which contains a single apoA1 molecule, whereas sHDL is actually associated with more than three apoA1 molecules, making CETP penetration more sterically complex. Lastly, the authors used atomistic simulations to show that CE from the core diffused to the mouth of N-barrel; however, its complete traversal through the tunnel was not observed, as slow diffusion of lipids requires large-scale MD simulation. And so whether CETP acts through the tunnel or shuttle mode remains unanswered.

Due to the slow diffusion of lipids in classical MD simulation, in order to visualize lipid transfer at MD time scales, steered molecular dynamics is frequently used. The approach typically involves the application of a steering force that pulls CE into the protein through one end and moves it through the tunnel to exit through the other end. A simplified ternary complex, consisting of CETP with its distal ends penetrating into two opposite phospholipid monolayers, has been simulated (Lei *et al.*, 2016) to envelop CE and TG pools, respectively. Using steered molecular dynamics (SMD), these authors pulled a CE molecule to the surface of the pore, which led to conformational rearrangements in CETP, allowing CE traversal through the entire CETP tunnel (Lei *et al.*, 2016). During CE movement through the tunnel, it was found to form physical contacts with F115, R158 and F167 in the N-barrel domain that attracted the CE molecule allowing it to enter the tunnel; I15, L23, A202, I205, L206, F263, F265, M433 in the neck region which acts as an energy barrier being the narrowest region; and F301, M412 in the C-barrel region that allowed for the rotation of the sterol ring and its eventual exit from the tunnel (Lei *et al.*, 2016). However, the study did not consider the role of bound phospholipids during this traversal. The study also used a large velocity for CE movement, as it exited through a 13 nm protein is just 0.5 ns. These faster dynamics overlooked atomistic details about the movement of CE across the tunnel. Several mutations in the neck region have also been reported to alter lipid transfer. Mutants I15W, M433W affected both CE and TG transfer. Moreover, F265R was not secreted, suggesting they did not fold properly (Qiu *et al.*, 2007).

In another study, CE was pulled through CETP, with the two PLs plugging the central holes (Keshavamurthy *et al.*, 2023). Here, CE was found to interact with L52, V55, W162, I163, I113, F115, F93, F35, F167, L150, L152, F147 and L177 during its entry through the N-barrel, and F429, A373, V323, V421 in the neck region and F292, A319, F350 and L409 in the C-barrel region (Keshavamurthy *et al.*, 2023). Further, the PL plugs were observed to assist in the conversion of CE from bent to linear form as it moved through the neck region (Keshavamurthy *et al.*, 2023). Classical molecular dynamic simulation of PL-bound CETP suggests that the states of bound PLs can influence the conformation of CETP and the cargo enclosed within it (Revanasiddappa *et al.*, 2018; Keshavamurthy *et al.*, 2023); Alternatively, to deduce the mode of action of CETP inhibitor, torcetrapib, CE movement was studied in the presence of the inhibitor (Dixit *et al.*, 2019). It was observed that due to the physical obstruction of the tunnel caused

by the inhibitor, a high energy barrier is created, leading to CE re-routing its path (Dixit *et al.*, 2019).

While some efforts have been made to understand the path taken by CE through CETP (Cilpa-Karhu *et al.*, 2015; Lei *et al.*, 2016; Keshavamurthy *et al.*, 2023), the traversal of TGs through the tunnel has been largely overlooked to date. Previously, it has been shown that CETP can accommodate two TG molecules in the tuning fork conformation (Chirasani and Sanjib, 2017). However, the study only considered two static TG molecules in the tunnel without PL plugs. Therefore, our objective was to bridge this gap by elucidating the mechanism of TG transfer through CETP and highlighting the role of PL plugs in this process. We believe that an in-depth understanding of the lipid transfer mechanism of CETP can streamline the development of potent therapeutic agents.

## 5.2 Methodology

### 5.2.1 Atomistic simulations of CETP with and without its PL plugs

MD simulations were performed in triplicate with the GROMACS v.2022.1 suite (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Lindahl *et al.*, 2001). The system, consisting of CETP with PLs (PDBID: 2OBD) or CETP without PLs, was represented by CHARMM36 (Best *et al.*, 2012; Klauda *et al.*, 2010). Water was modeled using the TIP3P representation (Jorgensen *et al.*, 1983). Each of the three starting conformations was placed in a dodecahedron box that was large enough to contain the system with at least 1.0 nm of solvent on all sides. Periodic boundary conditions were used, and long-range electrostatic interactions were treated with the particle mesh Ewald method (Darden *et al.*, 1993) using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to calculate the potential and forces in-between grid points. The real space cutoff distance and van der Waals cutoff were set to 1.2 nm. The bond lengths were fixed using the LINCS algorithm (Hess *et al.*, 1998), and a time step of 2 fs for numerical integration of the equations of motion was used. Coordinates were saved every 10 ps. Pressure coupling was performed by employing Berendsen barostat (Berendsen *et al.*, 1984*a*) using 1 bar as the reference pressure and a time constant of 2.0 ps with compressibility of 4.5e-5

bar using the isotropic scaling scheme. Fifteen positive counter-ions (Na+) were added by replacing fifteen water molecules to produce a neutral simulation box. All starting structures were subjected to a minimization and equilibration protocol recommended by CHARMM. Three independent trajectories, each 650 ns in duration at 300 K, were performed for CETP -with PLs. For CETP without PL, simulations were carried out for 600 ns at 300 K. The simulation setup is summarized in Table 5.1.

| System | Number of protein atoms | Number of lipid atoms | Number of water atoms | System size | Number of replicates | Time (ns) |
|---|---|---|---|---|---|---|
| CETP – without PL | 7,440 | 0 | 211,860 | 219,315 | 3 | 650 |
| CETP – with PL | 7,440 | 276 | 211,692 | 219,423 | 3 | 600 |
| R201A – CETP | 7,426 | 276 | 250.719 | 258.425 | 1 | 700 |
| R424A – CETP | 7,426 | 276 | 250,707 | 258,437 | 1 | 700 |

**Table 5.1.** Table summarizing the MD simulation setup

### 5.2.2 Steered MD simulation for pulling TG through CETP

Following the conventional MD run, the last frame of the trajectory was used as the starting point for SMD simulations. One TG (10:0/10:0/10:0) was manually placed at the opening of the C-barrel of CETP (Zhang *et al.*, 2012). TG was represented by improved CHARMM36 parameters that were compatible with particle mesh Ewald schemes for treating Lennard–Jones interactions (Campomanes *et al.*, 2021). This system was placed in a box with a length twice that of the total pulling distance. The system was minimized for 5000 steps using the steepest descent algorithm to remove atom–atom overlaps and subsequently equilibrated using the NPT ensemble for 1 ns. SMD simulations were performed independently on each of the two systems, both in the absence and presence of PL plugs, to elucidate the path of TG transfer. TG was pulled through the CETP tunnel using constant-velocity SMD simulations. The pulling vector was defined on the basis of the center of mass of the steered group, i.e., the tail region of TG and the center of mass of CETP tunnel residues in the direction of the N-barrel. Considering the asymmetric shape of CETP, its center of mass was dynamically switched from the C-barrel to the neck and finally the N-barrel, depending on the location of the steered group. Therefore, the entire pull simulation through the CETP tunnel was executed in four phases (Fig. 5.1).

| System | Number of protein atoms | Number of lipid atoms | Number of water atoms | System size | Number of replicates |
|---|---|---|---|---|---|
| CETP – without PL | 7,440 | 101 | 207, 336 | 214,877 | 6 |
| CETP – with PL | 7,440 | 276 + 101 | 207, 009 | 214,826 | 6 |
| R201A – CETP | 7,426 | 276 + 101 | 206, 988 | 214,807 | 3 |
| R424A – CETP | 7,426 | 276 + 101 | 206, 982 | 214,801 | 3 |

**Table 5.2.** Table summarizing the steered MD simulation setup

Constant velocity simulations were performed by implementing a force constant of 1000 kJ-mol-1-nm-2. The strength of the spring constant determines its stiffness and the force applied to the pull groups during SMD. An optimal value was determined based on the trade-off between stiffness and simulation speed. For each system, TG was pulled with a velocity of 1 nm/ns. This rate was chosen to obtain an optimal balance of accuracy and computational speed. For each system, SMD runs were performed in six replicates to sample the majority of paths. The steered MD simulation setup is summarized in Table 5.2.

### 5.2.3 Analysis

**Bend angle and normal mode analysis** The average bent angle between the two barrels was computed using gmx angle. To deduce the dominant internal modes of motion, the overall rotation and translation of CETP were removed by performing a least-squares fitting of each frame of the trajectory to the final structure of the simulation. The covariance matrix of atomic coordinates across the trajectory was built and diagonalized using gmx covar. The two principal components were obtained using gmx anaeig. The eigenvalues obtained were converted to nmd using Prody (Bakan *et al.*, 2011; Zhang *et al.*, 2021) and visualized using normal mode wizard in VMD (Humphrey *et al.*, 1996).

**Phasic movement of CETP barrels** The angle between the C-barrel and N-barrel domains with respect to the neck region was calculated using gmx angle. The Augmented Dickey Fuller (ADF) test was performed on this time series output to check for its stationarity. Briefly, a time series is considered stationary if its statistical properties

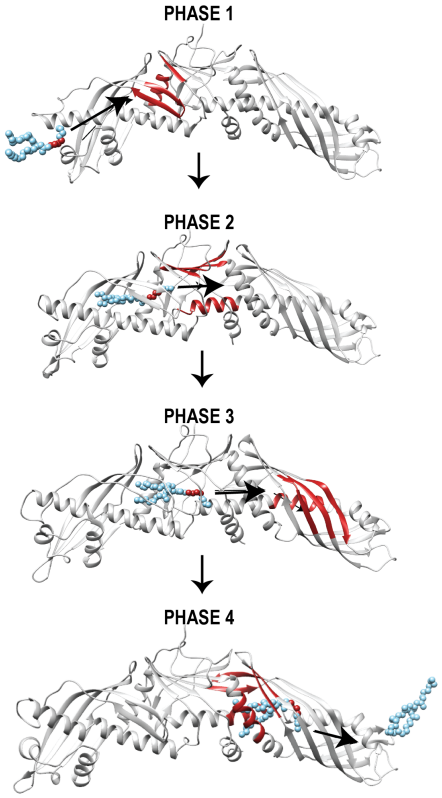



**Figure 5.1. Schematic depicting the four phases in which TG was steered through the CETP tunnel in the absence or presence of PL plugs.** TG was placed at the entrance of the C-barrel domain. A pull force was applied on the center of mass of the termini atoms of TG (highlighted in red) to pull it towards the center of mass of CETP tunnel residues in the direction of the N-terminal barrel (red). Briefly, edge of C-barrel (Phase 1), edge of neck (Phase 2), edge of N-barrel (Phase 3), and subsequently out of CETP through the N-barrel end (Phase 4) by applying a push force instead of a pull force.

remain constant over time. The ADF test assumes that a time series is not stationary owing to the presence of a unit root (Null hypothesis) or alternatively, the time series is stationary owing to the absence of a unit root (Alternative hypothesis). Subsequently, a 1-ns moving average was calculated for the movements of the C-barrel and N-barrel. The cross-correlation function was calculated as a measure of similarity or coherence between the movement of two barrels with respect to each other. Additionally, a window correlation was calculated for window sizes of 1 ns. To obtain the minimum-phase response and represent the frequency of this time series as the rate of phase change, the Hilbert transformation was performed as follows:

The analytical signal $x_a$ (t) of signal x (t) is:

$$x_a = F^{-1}(F(x)2U) = x + iy \tag{5.1}$$

Where F is the Fourier transform, U is the unit step function, and y is the Hilbert transform of x. Hilbert transformation of this time series was analyzed by zeroing out the negative half of the frequency spectrum, turning this real-valued signal into a complex signal. From the Hilbert-transformed signal, the phase angle for the movements of the C-barrel and N-barrel at each time point was calculated. The number of frames, in which the phase angles of the two barrels lay in the same phase (overlapped) or opposite phases (no overlapping), was recorded.

**Path traversed by TG through CETP tunnel and residues that make contact** For each SMD run in the absence or presence of PLs, the position of the C1 carbon of TG was traced throughout the simulation trajectory. The structures were visualized and rendered using UCSF Chimera (Pettersen *et al.*, 2004) to trace the path taken by TG. Contact analysis was performed using CONAN (Mercadante *et al.*, 2018) to identify CETP residues within the radius of 0.3–0.6 nm of TG as it traversed the tunnel. Residues within this contact radius for > 1 ns were extracted for each SMD run. Residues that were either common in all six runs or found in at least four of the six runs were reported.

**TG Fall out analysis** A simulation frame from the previously carried out SMD simulations was extracted such that TG had entered the tunnel from the C-barrel mouth and reached its end (end of Phase 1). Starting from this structure, a random seed in the

presence and absence of PL plugs was set to generate random velocities, and forty-three replicate simulations were performed where TG was pulled towards the neck (Phase 2). For each of these replicates, the distance between the center of mass of TG and an arbitrary vector passing through the CETP tunnel was calculated and plotted to show the path taken by TG through the neck.

**Power-stroke-like movement of PL acyl tails** The distance traversed by TG (center of mass of the glycerol head group) and the center of mass of residues in the N-barrel region was calculated using gmx distance. The instantaneous velocity, which is defined as the rate of change of velocity for a very small-time interval $\Delta t = 10ps$, was calculated for simulations in the presence or absence of PLs. These values were de-noised based on the distribution of instantaneous velocities obtained throughout the SMD replicates.

**Side chain movement of Phe residues lining the neck** For the three Phe residues lining the neck, the torsion angle of $\chi_2$ side chain was calculated using dihedral module of MD Analysis suite (Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). The side chains frequently undergo switching between rotamers; however, we were interested in the deviation of the side chain angle irrespective of the rotameric state. Two rotameric states were identified by clustering the torsion angle of $\chi_2$ side chain throughout simulation. Rotamericity of Phe residues substantially varies depending on their position in a protein (Schrauber *et al.*, 1993). Therefore, the mean of the two rotamers exhibited by the Phe residues was considered separately. These means were subtracted from the respective values across time. The deviation in the change of $\chi_2$ irrespective of the rotamer switch was then calculated.

**Salt bridge analysis** Salt bridges between the head group of PL and charged amino acids surrounding the ventral openings (Lys, Arg, Glu, Asp) were calculated using an in-house Python script. Salt bridge distances were computed between the basic nitrogens and acidic oxygens in the PL headgroup or the amino acids surrounding the PL. A salt bridge was said to exist if the distance between the two groups was smaller than 6 Å(Luo *et al.*, 1999; Beckstein *et al.*, 2009).

**Computational analysis of mutants** Simulation stabilized WT-CETP structure was mutated to R201A and R424A using the rotamer module in Chimera (Pettersen *et al.*, 2004). The mutant structures were equilibrated using the following MD protocol. The system, consisting of mutant CETP with PLs, was represented by CHARMM36 (Best *et al.*, 2012; Klauda *et al.*, 2010). Water was modeled using the TIP3P representation (Jorgensen *et al.*, 1983). Each starting conformation was placed in a dodecahedron box that was large enough to contain the system with at least 1.0 nm of solvent on all sides. Periodic boundary conditions were used, and long-range electrostatic interactions were treated with the particle mesh Ewald method (Darden *et al.*, 1993) using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to calculate the potential and forces in-between grid points. The real space cutoff distance and van der Waals cutoff were set to 1.2 nm. The bond lengths were fixed using the LINCS algorithm (Hess *et al.*, 1998), and a time step of 2 fs for numerical integration of the equations of motion was used. Coordinates were saved at every 10 ps. Pressure coupling was performed by employing Berendsen barostat (Berendsen *et al.*, 1984*a*) using 1 bar as the reference pressure and a time constant of 2.0 ps with compressibility of 4.5e-5 bar using the isotropic scaling scheme. A neutral simulation box was obtained by adding sixteen positive counter-ions (Na+) by replacing sixteen water molecules in each of the two systems. All starting structures were subjected to a minimization and equilibration protocol recommended by CHARMM. Independent trajectories for each mutant were simulated for 700 ns at 300 K. Simulation-stabilized structures for the two mutants were used as the starting point for SMD simulations, and the system was prepared as described previously.

**Root mean square fluctuation (RMSF) and stability of the hydrophobic plane** RMSF of the PL heavy atoms was calculated while the TG moved through the tunnel using gmx rmsf. The plane formed by the acyl tails of PL was measured by calculating the vectors formed between C29 and C218 (v1) and C39 and C318 (v2). The angle between the two vectors depicts the overall conformation of PL tails in the tunnel, such that an angle close to 0° represents that the two tails of PL are parallel to each other.

### 5.2.4 Umbrella sampling simulations for potential of mean force calculation

To calculate the potential of mean force (PMF), umbrella sampling simulations were performed on the path derived from SMD simulations, both in the absence and presence of PL. Starting from the initial position of TG at the mouth of the C-barrel, the path traversed through the tunnel was sampled at 0.1 nm intervals, resulting in approximately 50 windows. A force constant of 500 kJ-mol-1-nm-2 was used. To ensure sufficient overlap of histograms along the reaction coordinate, a force constant of 5000 kJ mol-1 nm-2 was used for certain windows to ensure sufficient sampling in specific regions of the reaction coordinate. These restraining umbrella potentials ensured the sampling of every region of the conformational space along the reaction coordinate, which would not otherwise be accessible during direct sampling. Each window was thoroughly equilibrated for 1 ns and subjected to a production run for 50 ns or 100 ns to ensure convergence of PMF. Hence, a series of histograms containing a biased distribution of the reaction coordinate from each window was obtained. These histograms were unbiased and combined to produce the PMF using the weighted histogram analysis method (WHAM) (Kumar *et al.*, 1992). The last 30 ns of the trajectory were used for calculation in the case of the system with PLs, while the last 20 ns were used for the system without PLs. The first and last bins of the reaction coordinate were assumed to be neighbors to generate a periodic PMF. The convergence of PMF was ensured by checking the overlap between individual histograms representing the distributions of sampled configurations and by calculating the deviations between individual PMFs calculated for different blocks of the simulation. An average PMF profile was generated by considering the mean potential at each point along the reaction coordinate. Table 5.3 summarizes the umbrella sampling simulations performed in this study.

| System | System size | Number of windows | Duration of each window | Total time (µs) |
|---|---|---|---|---|
| CETP – with PL | 214,826 | 48 | 50 ns | 2.4 **µs** |
| CETP - without PL | 214,877 | 51 | 100 ns | 5.1 **µs** |

**Table 5.3.** Table summarizing the umbrella sampling simulations

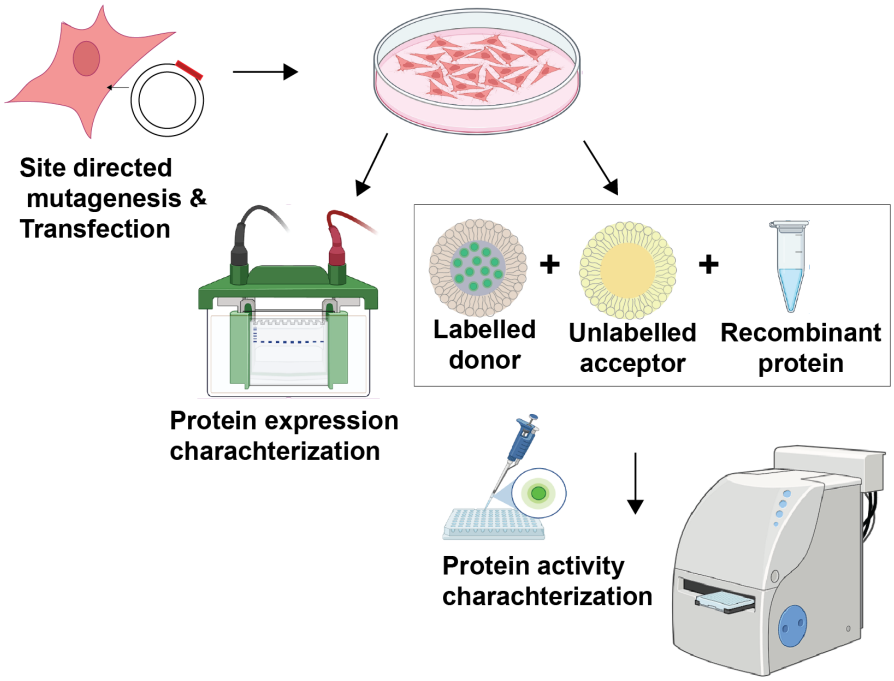



**Figure 5.2.** Schematic depicting the study design for experimental validation.

### 5.2.5 Site-directed mutagenesis & *in-vitro* characterization of mutants

**Cloning wild-type CETP into expression vector and preparation of mutants** WT-CETP plasmid was a kind gift from Thea Bismo Strøm, Department of Medical Genetics, Oslo University Hospital, Norway. Briefly, RNA isolated from immortalized human hepatocytes was converted to cDNA, amplified using polymerase chain reaction, and TA-cloned (Invitrogen, Waltham, MA, USA) into pcDNA3.1 V5/His TOPO TA vector. Due to the differences in activities between tagged CETP and WT-CETP, a stop codon was inserted between the two tags to obtain pcDNA3.1-WT-CETP-V5, which exhibits activity similar to that of WT-CETP (Ølnes *et al.*, 2023). The plasmid containing the WT-CETP sequence had p.Ile422 (c.1264G>A), a neutral genetic variant (allele frequency, 0.62). The lipid transfer activity, synthesis, and intracellular transport of both variants (WT-CETP and I422V) have been tested, showing no significant difference (Ølnes *et al.*, 2023).

Using pcDNA3.1-WT-CETP-V5 as a template, five mutants were generated using a QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. Plasmids were isolated using a Plasmid Midi Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Plasmid integrity

was verified by Sanger sequencing. The sequences of the oligonucleotides used are provided in Supplementary Table 11.

**Cell culture and transfection** HEK 293-F cells (6 × 105 cells/well) were seeded in a 6-well plate (Nunc, Villebon-sur-Yvette, France). Cells were cultured in Dulbecco's Modified Eagle Medium (Gibco, Villebon-sur-Yvette, France) supplemented with 10% fetal bovine serum (Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France) and 1% penicillin-streptomycin antibiotic (Gibco, Villebon-sur-Yvette, France). Cells at 50% confluence were transfected with 1 $\mu$g DNA using a jetPRIME transfection kit (Sartorius Polyplus SAS, Illkirch-Graffenstaden, France). After 2 h, cells were washed with phosphate-buffered saline (Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France) and were incubated in Freestyle 293 expression media (Gibco, Villebon-sur-Yvette, France). Cells were harvested 48 h post-transfection and were lysed using RIPA lysis buffer (NaCl 150mM, 1% d'IGEPAL CA-630, 0.5% Sodium deoxycholate, 0.1% Sodium Dodecyl Sulfate, Tris 50mM, pH 8) containing a protease inhibitor cocktail (Protease and phosphate inhibitor, Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France). The spent medium was centrifuged at 1200 rpm for 5 min to remove cell debris. Protein concentrations of the lysates and supernatants were determined using RC DC Protein Assay (Bio-Rad Laboratories, Inc., France).

**Western Blotting** Equal amounts (30 $\mu$g) of protein samples were separated by 10% sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Prestained Opti-Protein Marker (Applied Biological Materials Inc., Richmond, BC, Canada) was used as a molecular-weight marker. Separated proteins were blotted on a polyvinyl difluoride membrane (Amersham Hybond, 0.2mm. Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France). The membrane was blocked using skimmed milk for 1 h at 25 °. Subsequently, the membrane was incubated with an anti-V5 antibody (Invitrogen, Villebon-sur-Yvette, France, R96025) (1:1500 for supernatant and 1:5000 for lysate) for detecting V5-tagged CETP, or an anti-$\beta$-actin antibody (Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France A5441), which was used as an internal control for cell lysate. An anti-mouse IgG conjugated with horseradish peroxidase (Goat anti-mouse IgG(H+L) secondary antibody HRP, Thermo Scientific Cat . 32430, Villebon-sur-Yvette, France) was used as a secondary antibody. Immunoreactive protein bands were detected using Immobilon

Western Chemiluminescence HRP substrate (Millipore) and captured on Fusion Fx (Vilber Lourmat, Marne-la-Vallée, France). CETP expression was analyzed using ImageJ v.1.54p and normalized with respect to $\beta$-actin expression (Schneider *et al.*, 2012).

**CETP activity assay** Lipid transfer activity of CETP was measured using 4 $\mu$g protein from each supernatant (WT or mutant), using a CETP Activity Assay Kit (MAK106; Sigma Aldrich Chimie, Saint-Quentin-Fallavier, France) according to the manufacturer's instructions. Briefly, fluorescently-labelled donor particles were incubated with the supernatant and acceptor particles in a sealed black flat-bottomed 96-well plate (View plate 96 black, clear bottom, Revity) for 3 h at 37 °C in a water bath. Fluorescence intensities were measured using excitation and emission wavelengths of 465 and 535 nm, respectively, on a spectrofluorometer (SAFAS Xenius, Monaco) at 400 V.

Fig 5.2 summarizes the experimental design of the validation experiment employed in the study.

## 5.3 Results

To decipher the role of PL plugs in the lipid transfer, the structure of CETP (PDB ID: 2OBD) was simulated with and without its DOPC plugs. During the unrestrained simulation, we observed that the RMSD values began to plateau at values 2-3 Å for the triplicate CETP simulations without PL (Fig 5.3 A) and 3-5 Å for the triplicate CETP simulations in the presence of PL (Fig 5.3 B).

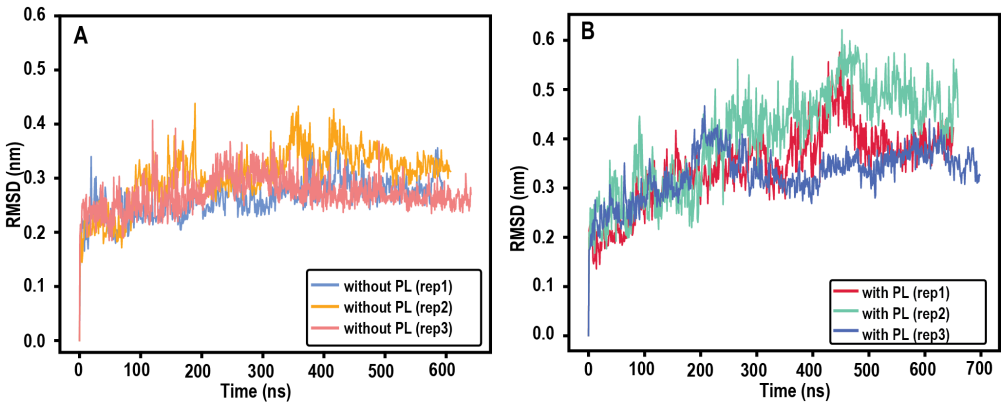



**Figure 5.3. Root mean square deviation of CETP A.** without and **B.** with its PLs

### 5.3.1 PL plugs synchronize the movement of CETP barrels

In the crystal structure, the two barrels of CETP appear bent, providing it a boomerang shape. Therefore, the average tilt of the CETP barrels post stabilization in the simulations was calculated and observed to differ drastically from the crystal structure (72.9°). However, within the two simulated systems, the angle only differed slightly (47.5° and 45° in the absence and presence of PLs, respectively) (Fig. 5.4 A, B). Next, normal mode analysis (NMA) along the dominant modes of motion was performed to study the large-amplitude motions exhibited by CETP. In addition to calculating the vibrational modes of protein motion, NMA is useful in determining the conformational states that are accessible to a molecule about an equilibrium position (Bauer *et al.*, 2019). The terminal regions of the barrel exhibited large movements towards each other irrespective of the PL-plug status of CETP (Fig. 5.4 C, D). However, in the presence of PLs, the neck and its adjacent regions, in addition to the barrel ends, exhibited small movements (Fig. 5.4 D). Both the reduction in bend angles and the motion of the barrels may impart an overall linear conformation to CETP in the presence of PLs. This is in contrast to the bent shape observed in the absence of PLs.

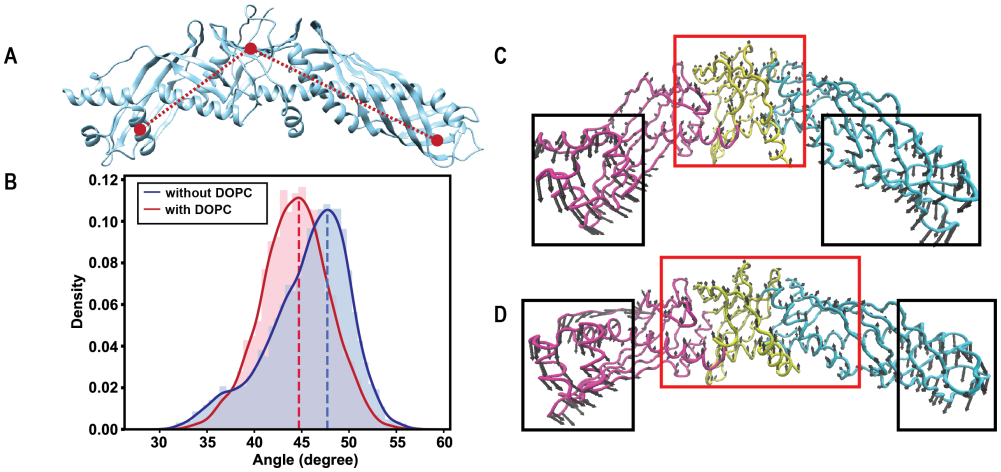



**Figure 5.4. Characterization of the bend angle of CETP. A.** The vectors used to define the average bend angle of CETP. **B.** Average bend angle of the two barrels with respect to the neck of CETP. Normal model analysis depicting the two dominant modes of motion in **C.** absence and **D.** presence of PL plugs. The length of the arrows depicts amplitude, and the arrowheads depict the direction of motion. The protein is colored according to its three domains: C-barrel (pink), Neck (yellow), and N-barrel (blue). The terminal regions of the barrels (highlighted in a black box) move towards each other, irrespective of the CETP PL status. The neck and its adjacent region (highlighted in red box) move in the opposite direction. The amplitude of movement of the neck region and its direction of movement differ when the PL plugs are present, possibly counterbalancing the overall motion.

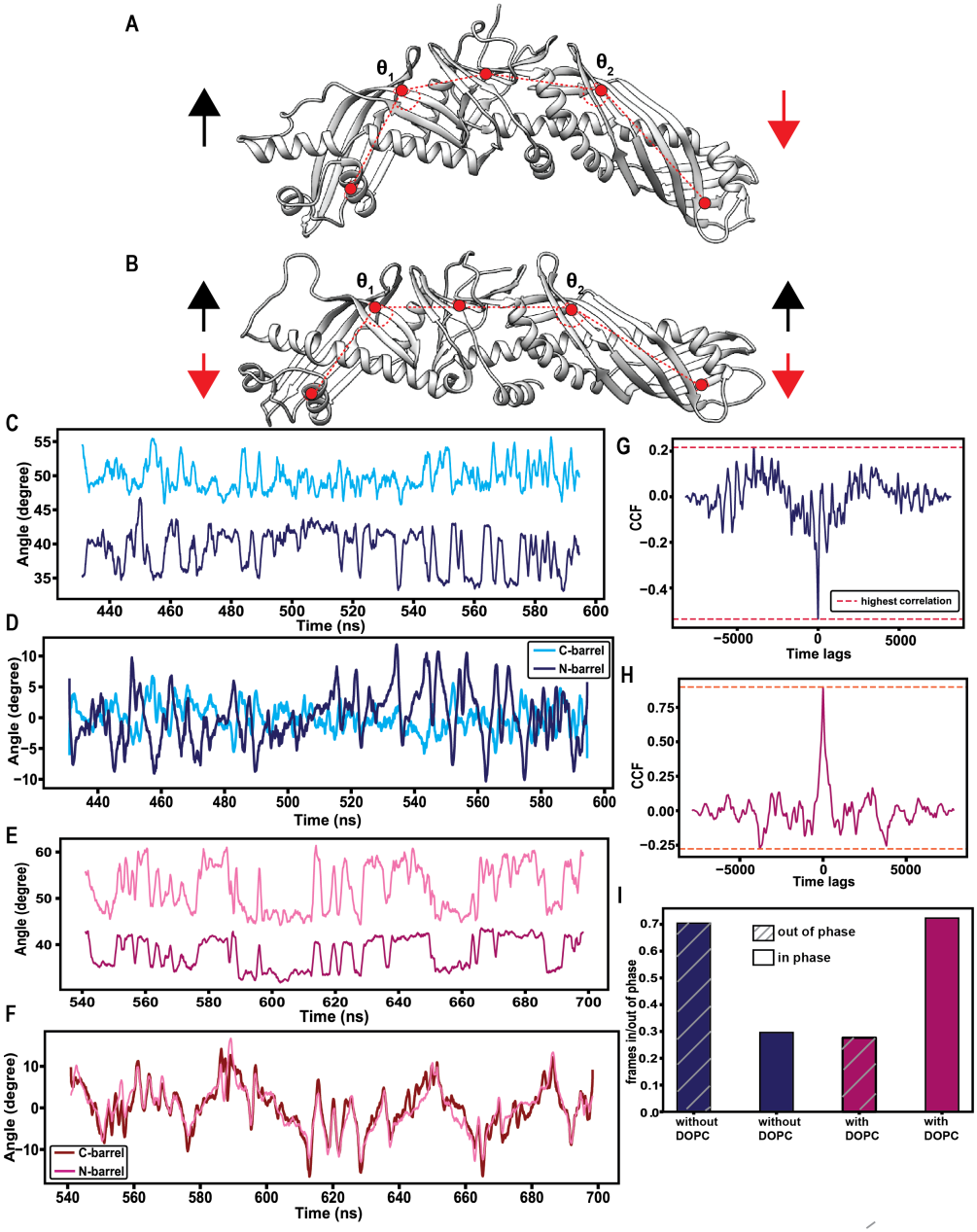



**Figure 5.5. Presence of PL plugs results in a switch in synchronicity of oscillations of CETP barrels.** Schematic representation of the vectors used to compute the movement of barrels. The graphic depicts the movements of two barrels **A.** out-of-phase and **B.** in-phase. Dynamic movements of the C-barrel and N-barrel domains of CETP in the **C.** absence of PL plugs. **D.** Rate of change of phase angle of the time series after Hilbert transformation in the absence of PL plugs. Movements of the C-barrel and N-barrel domains of CETP in the **E.** presence of PL plugs. **F.** Rate of change of phase angle of the time series after Hilbert transformation in the absence of PL plugs. Cross-correlation between movements of the C-barrel and N-barrel in **G.** absence and **H.** presence of PL plugs.**I.** Duration of simulation, in which the two barrels remain synchronous (in-phase) or asynchronous (out-of-phase) in the presence or absence of PL plugs, respectively. The duration is depicted as a ratio of the number of frames observed to be in/out phase across triplicate simulations to the total number of frames.

The differential motion of CETP in the presence and absence of PLs was further investigated to elucidate the rhythmicity within barrel motions. To this end, the angle formed by the N-barrel and C-barrel with respect to the neck across simulation time was calculated between the vectors defined by the barrels (Fig. 5.5 A,B). The N-barrel was found to oscillate between 35–40°, irrespective of the PL status. In contrast, the angle of C-barrel oscillations showed mean fluctuations of about 10 ° (between 47–57 °) in the presence of PLs compared to 2° mean fluctuation (between 50–52 °) in the absence of PLs (Fig. 5.5 C, E). Cross correlation between the movement of N-barrel and C-barrel showed that the movement of barrels was negatively correlated in the absence of PLs (Fig. 5.5 G) while a positive correlation was observed when the PL plugs were present (cross correlation coefficient, 0.8 at 0 lags) (Fig. 5.5 H). This implies that the movements of N-barrel and C-barrel in CETP with PLs are temporally similar, and the movements of the two barrels are synchronized with each other.

Movements of the barrels about a mean position can be considered a periodic function [x(t)] as they return to their respective equilibrium positions at regular intervals; however, the speed of barrel displacement (represented by frequency in a periodic signal) and exact angle of displacement of the barrels (represented by the phase of the periodic signal) may vary. To determine these quantities, a Hilbert transformation [ $x^{(}t)$] of the angle–time series [x(t)] was performed. The Hilbert transformation transforms all the components of a signal by 90 °, turning a real-valued signal into a complex signal. This transformation allows the assessment of the change in the phase of a signal. On overlaying $x^{(}t)$ of the two barrels, it was observed that the rates of their phase change varied in the absence of PLs (Fig. 5.5 D). In contrast, the two barrels exhibited synchronous changes in phase in the presence of PL plugs (Fig. 5.5 F). The number of frames exhibiting such concerted movement was much higher (> 70% of simulation time) in the presence of PLs than in their absence (Fig. 5.5 I).

### 5.3.2 PL ensure successful traversal through CETP

Previous studies on CETP and lipoproteins have indicated that the N-barrel terminal may selectively penetrate HDL, while the C-barrel may have a preference for LDL or VLDL (Zhang *et al.*, 2012, 2015). Considering the polar nature of the barrels, a TG molecule was placed at the mouth of the C-barrel. Following this, steered molecular dynamics

(SMD) was performed on these two systems, both in the absence and presence of PLs, to elucidate the path of TG transfer. A pull was considered successful if TG entered from the C-barrel and exited through the N-barrel end after traversing the entire tunnel. The force profiles of these pulls showed no major conformational transitions. In fact, PLs in the neck region did not cause any obstruction, as the profiles showed no drastic increment in force during the pulls.

| System | Phase 1 | Phase 2 (success) | Phase 3 (failure) |
|---|---|---|---|
| CETP – without PLs | 7 | 21/43 = 0.48 | 22/43 = 0.5 |
| CETP – with PLs | 7 | 43/43 = 1 | 0/43 = 0 |

**Table 5.4.** Simulation statistics highlighting the number of successful runs for each phase of the SMD pull.

In the absence of PLs, even though the TG entered the tunnel via the C-barrel end (in seven instances), it exited downwards through the space between the C-barrel and neck in two of the replicates. These failed simulations were revisited to elucidate the cause of this premature exit. Starting with a simulation frame such that TG had reached the end of C-barrel, a random seed was set to generate random velocities, and TG was pulled towards the N-barrel. These pulls were performed in multiple replicates (of 43) and observed that in 51% of these simulations, TG dropped out through the space between the barrel and neck when PL plugs were absent (Table 5.4). However, no falls were observed in the presence of PLs.

To explain this phenomenon, the residues that made contact with TG (within a 0.3–0.6-nm range) for > 1 ns of simulation in the successful (21/43) and failed (22/43) replicates were calculated. This led to the identification of two hydrophobic patches towards which TG gravitated. The first patch was located towards the roof of the neck region and constituted residues F236, L261, F263, F265, L273, V276, A277, A373, Y375, S439, I443, I448, L457, M459, and L485 (Fig. 5.6 C). The second patch was located towards the floor of the neck region and was made by residues S204, I205, G209, L432, and M433 (Fig. 5.6 D). These results suggested that TG navigated through the tunnel by moving from one hydrophobic cluster to another. Since TG encounters hydrophobic interactive forces only, its contacts with tunnel-lining residues are transient and, therefore, easy to break and form as TG shifts its position. A bifurcating inflection point lies at the interface of the C-barrel and neck, beyond which two fates for TG exist: TG

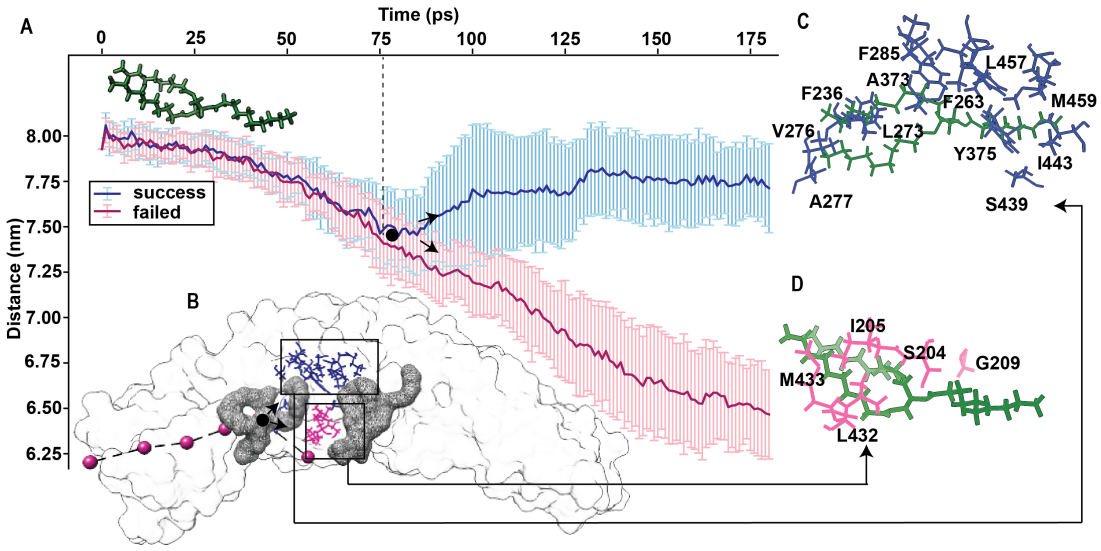



**Figure 5.6. The dichotomy of fates of TG. A.** The average path taken by TG (across 43 replicate simulations) either directs it through the neck (blue) or leads it out into the milieu through the space in the floor of CETP (pink). Error bars depict standard deviations. **B.** The inflection point is depicted as a black circle. The figure shows the path taken by TG, which leads to its fallout, overlaid with the location of PL plugs on CETP. Hydrophobic cluster located at either **C.** the roof of the neck region or **D.** the floor of the neck region adjacent to the central opening.

either reaches the center of the neck or falls out through the opening in the floor of the protein (Fig. 5.6 A). Interestingly, the first PL plug was located at this inflection point. This suggests that the PLs physically seal CETP such that TG does not slip out during its transit (Fig. 5.6 B). Moreover, favourable interactions with the acyl tails reroute TG such that it always moves towards the hydrophobic patch at the roof of the protein. This also reduces the likelihood of incidents and ensures a successful transit through the tunnel.

### 5.3.3 TG moves from one hydrophobic cluster to the next

Subsequently, all the hydrophobic clusters lining the tunnel that a TG encounters during its transit in the presence and absence of PL plugs were identified (Fig. 5.7 A, B). A complete list of all the interacting residues is summarized in Table 5.5. It was observed that during its traversal through the tunnel in the presence of PL, the path taken by TG clustered together, further implying the role of PL in guiding the movement of TG through the tunnel. Further, on comparing the average path taken by TG in this study with the previously identified path for CE (Lei *et al.*, 2016; Dixit *et al.*, 2019), it was observed that TG interacted with 18 common residues (I15, L23, F35, F93, V198, I205, L206, F263,

F265, F270, Y361, F363, M412, F408, F429, L425, F429, and M433; (Fig. 5.7 C). Moreover, torcetrapib (CETP inhibitor) interacting residues C13, H232, F263, A202, and V198 (Liu *et al.*, 2012) also lie in the path of TG, implying that the binding of the inhibitor may affect the movement of TG, just as it affects the movement of CE (Liu *et al.*, 2012; Dixit *et al.*, 2019). On comparing the common residues between the path of TG with that of CE identified in previous studies, it was evident that TG formed more contacts with CETP than CE (Fig 5.7 D. Further, both lipids contacted a similar set of residues in the neck, while the path slightly varied in the C-barrel and N-barrel. This may be due to a higher pulling velocity used in the CE studies compared to our study.

| System | Domain | Residue |
|---|---|---|
| CETP – without PLs | C-barrel | 270, 273, 274, 277, 283, 285, 307, 308, 311, 321, 323, 325, 329, 331, 338, 340, 342, 344, 346, 348, 350, 361, 363, 367, 371, 373, 375, 380, 382, 394, 397, 398, 405, 408, 409, 412, 413, 416, 417, 418, 421, 422, 425, 429, 433 |
| | Neck | 15, 195, 198, 202, 205, 206, 228, 236, 261, 263, 265, 438, 439, 441, 448, 455, 457, 459, 467, 476 |
| | N-barrel | 20, 23, 24, 27, 28, 31, 35, 64, 67, 69, 74, 82, 84, 86, 89, 91, 93, 117, 119, 121, 123, 125, 127, 145, 147, 150, 152, 167, 172, 175, 176, 179, 180, 183, 184, 187, 190, 194 |
| CETP – with PLs | C-barrel | 269, 270, 273, 283, 285, 287, 292, 295, 296, 310, 311, 321, 323, 331, 338, 340, 344, 346, 348, 350, 352, 359, 361, 363, 367, 369, 371, 373, 380, 382, 405, 406, 408, 409, 412, 413, 416, 417, 418, 420, 421, 422, 425, 429, 432 |
| | Neck | 11, 13, 15, 195, 198, 199, 201, 202, 205, 206, 215, 228, 230, 232, 236, 263, 265, 432, 448, 455, 457 |
| | N-barrel | 20, 23, 24, 27, 28, 31, 35, 40, 64, 67, 69, 72, 74, 82, 84, 86, 89, 91, 93, 117, 119, 121, 123, 125, 127, 129, 136, 138, 145, 147, 150, 152, 167, 172, 175, 176, 179, 183, 187, 191, 194 |

**Table 5.5.** Residues that make contact with TG > 1ns of simulation time in 66% of replicates of CETP without PL plugs.

### 5.3.4 Residues lining the neck allow passage of TG into the tunnel

Of the residues that made contact with TG for an extended duration, three Phe residues were uniquely oriented. These residues (F270, F265, and F263) were located on the roof of CETP and had their side chains buried inside the cavity. During the simulation, the phenylalanine side chains extensively interacted with the acyl tails of PL. To ascertain if these residues altered the route of TG, the torsion angles of $\chi 2$ side chains of these residues throughout the pulling simulations were calculated. It was observed that these residues underwent frequent rotamer switching irrespective of the PL status of CETP.

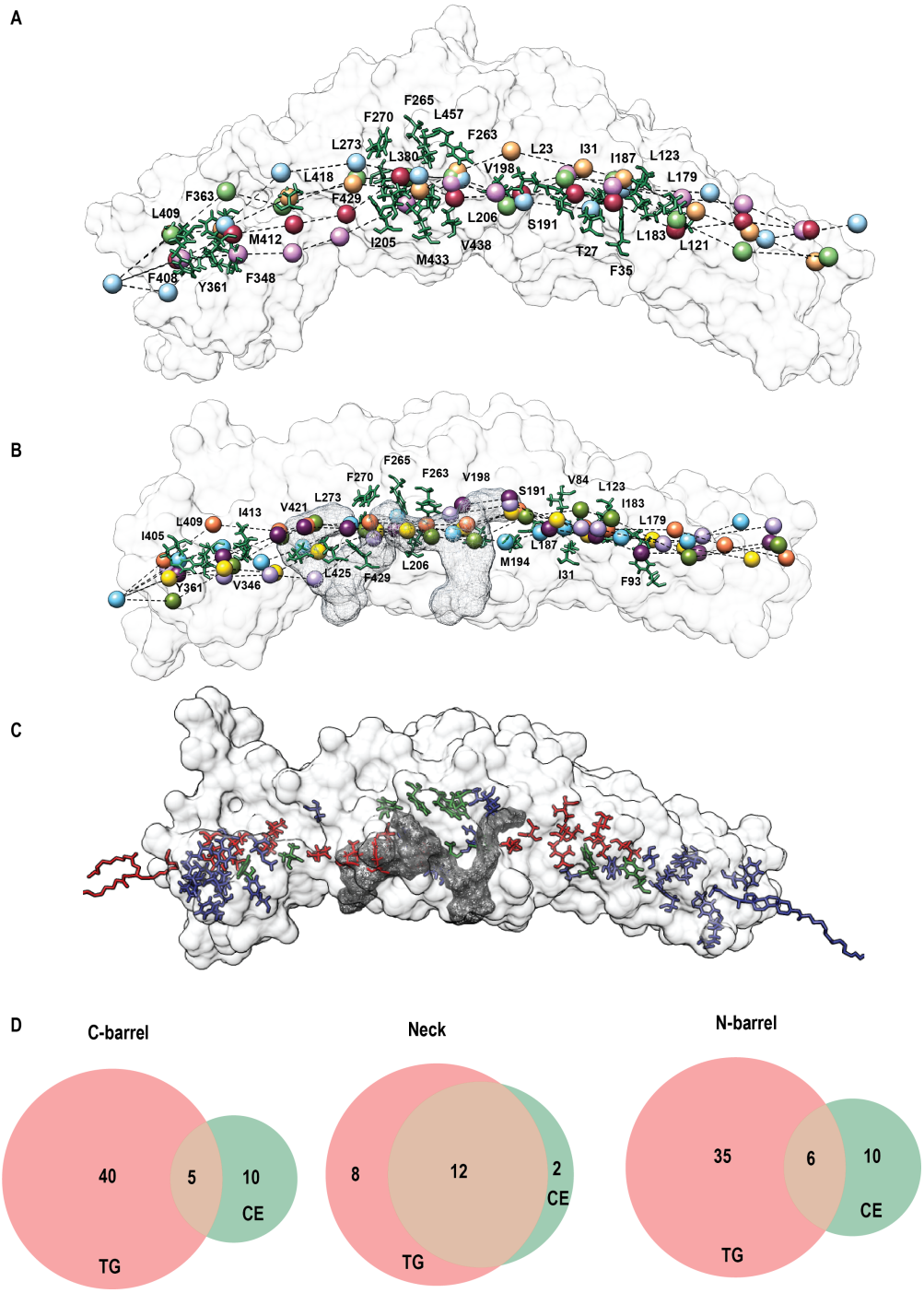



**Figure 5.7. TG transfer through CETP.** Tracing the center of mass of the glycerol moiety of TG as it moves through the tunnel, in each replicate simulation of CETP for determining the path taken by TG through the CETP tunnel in the **A.** absence and **B.** presence of phospholipids (PL) plugs. The residues within a radius of 0.3–0.6 nm staying in contact with TG for > 1 ns of simulation time in all six replicates are highlighted. Multiple paths derived from six independent replicate simulations are depicted. **C.** Comparison of residues that TG and CE interact with through their journey through the CETP tunnel. The average path taken by TG, determined by residues that made contact with TG for > 1ns of simulation time in all the replicates, is shown in red. Residues that CE makes contact with are identified in previous simulation studies and are shown in blue. Residues common to both lipid pathways are shown in green. **D.** Venn diagram showing the number of residues that TG and CE interact with through their journey through the CETP tunnel.

However, in the presence of PL plugs, as TG slid past these residues, they rotated by approximately 50˚ to pave the way for TG movement (Fig. 5.7 B-D). Once TG had passed, the side chains went back to their original orientation. This motion of the three Phe side chains appeared to be coordinated, moving sequentially as TG approached them.

Residue gates in proteins serve several functions, including ensuring substrate selectivity, preventing solvent access to specific regions, and coordinating processes that occur in different compartments of proteins. Given the location and orientation of these residues in the bottleneck of the protein and the distinct closed and open states exhibited by these gates, it was hypothesized that they may play a role in controlling the rate of transfer and direction of lipid cargo.

Since the current convention assumes that TG transfer occurs from the C-barrel to the N-barrel, it was further hypothesized that these side chain residues swing in a single direction, allowing for lipid movement in a single direction. To test this hypothesis, TG was pulled in the reverse direction (from the N-barrel to the C-barrel). It was observed that the phenylalanine side chains swung open again, allowing TG to pass through (Fig. 5.9). This led to the observation that these residues may not act as hinged gates that open in a single direction, but instead swing about their vertical axes. This is analogous to a flap-like movement that can allow both influx and efflux. It is further proposed that these flaps may prevent regurgitation, such that a substrate entering the tunnel does not reflux out of it. Moreover, since the rates of transfer of CE and TG through CETP are significantly different (Morton and Liu, 2020*a*), the flap-like movement of these residues may be coordinated with two molecules moving with varying speeds through the tunnel in the opposite direction.

Interestingly, these Phe residues are conserved in other members of the BPI family; F265 is found in PLTP; F263 is found in lipopolysaccharide (LPS)-binding protein; and F270 is present in PLTP, LPS-binding protein, and BPI in addition to CETP (Fig. 5.8 A). This implies an evolutionary conserved role of these flaps in regulating lipid movement through the BPI/LBP family proteins.

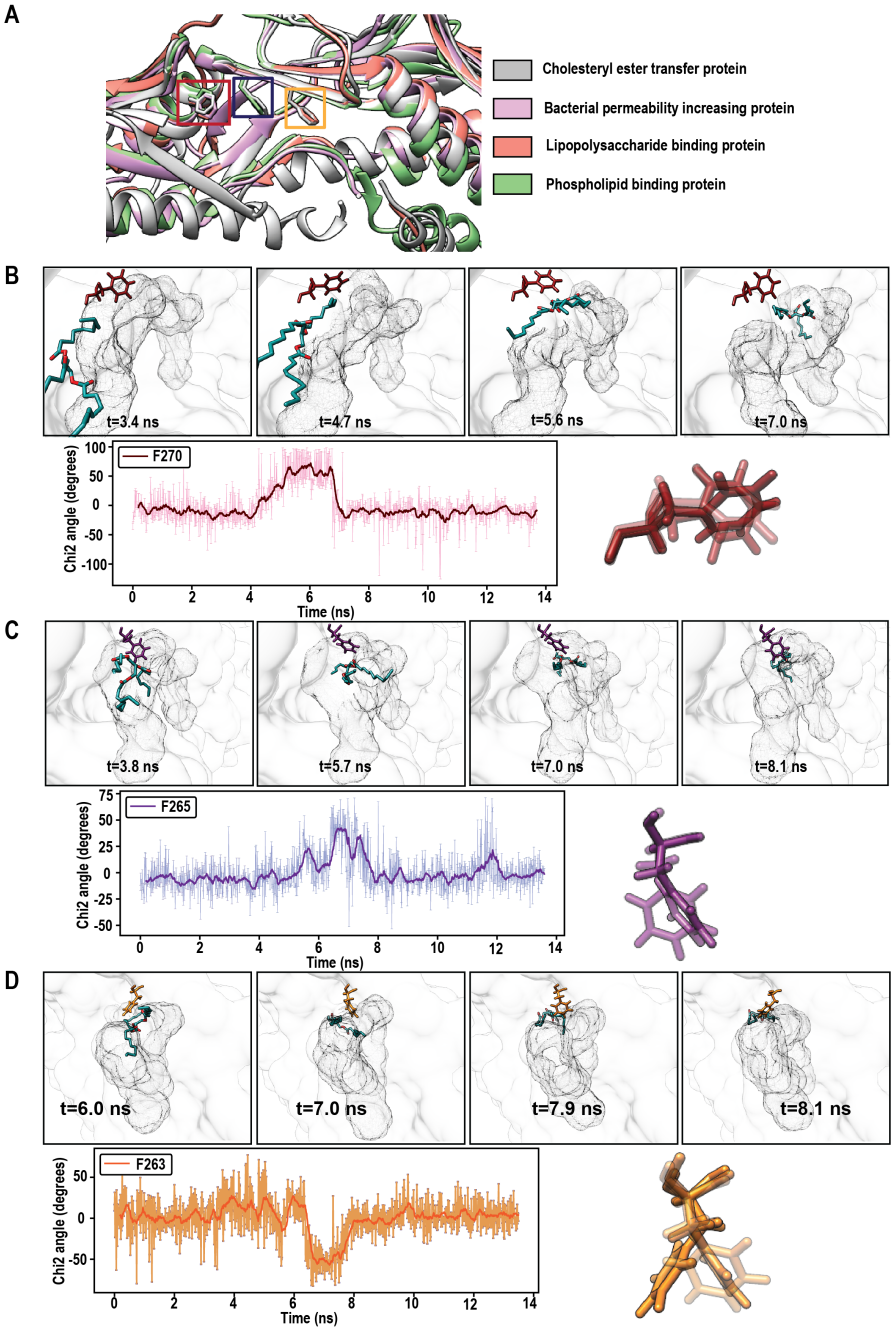



**Figure 5.8. Movement of the flap residues lining the neck region. A.** Location of flaps in other members of the BPI/LBP family. PDB structures of CETP (PDB ID: 2OBD), bacterial permeability increasing protein (BPI) (PDB ID: 1EWF), lipopolysaccharide binding protein (AF-P18428-F1), and phospholipid transfer protein (AF-P55058-F1) overlaid using Chimera. F270 on all structures is highlighted in red, F265 in blue, and F263 in yellow, respectively. Changes in normalized $\chi 2$ angle of **B.** Phe270 **C.** Phe265 and **D.** Phe263 throughout simulation. Snapshots from the simulation depict the movement of Phe side chains as TG approaches it. The solid line depicts the moving average across 20 ps in a simulation. The line plot represents the changes in the $chi^2$ angle in real-time.

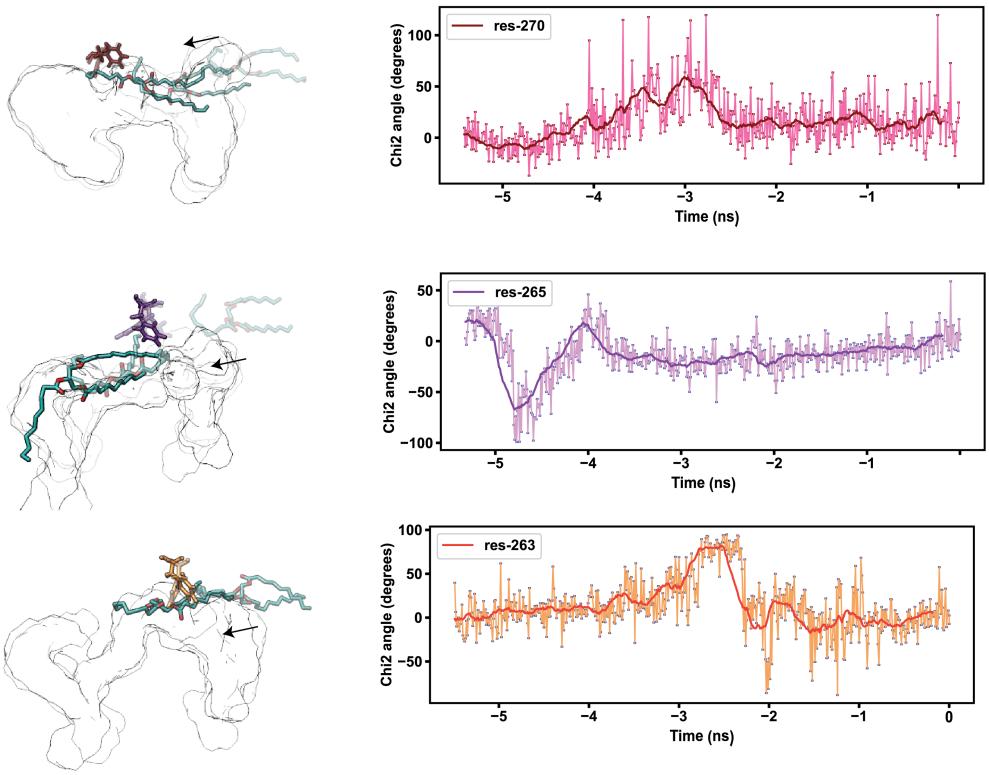



**Figure 5.9. Movement of residues lining the neck region when TG is pulled in reverse.** Changes in the normalized $\chi 2$ angle of Phe270, Phe265, and Phe263 throughout Phase 2 of the simulation when TG was pulled from the N-terminus towards the C-terminus. The solid line represents the moving average over 20 ps in a simulation, while the line plot shows the changes in $\chi 2$ angle in real-time.

### 5.3.5 PL plugs facilitate a power-stroke-like motion

As TG moved through the neck region, past the PL tails, it exhibited a forward thrusting movement. This movement was only apparent in the neck region and not during TG entry or exit through the barrels (Fig. 5.10 A). To quantify this thrust, the instantaneous velocity of TG was determined, which increased as TG moved past PL acyl tails (Fig. 5.10 B). The occurrence of spikes in the instantaneous velocity coincided with an "assisted gliding" motion. This gliding was reminiscent of the power-stroke motion of actin–myosin, leading to the sliding of thin actin filaments over thick myosin filaments.

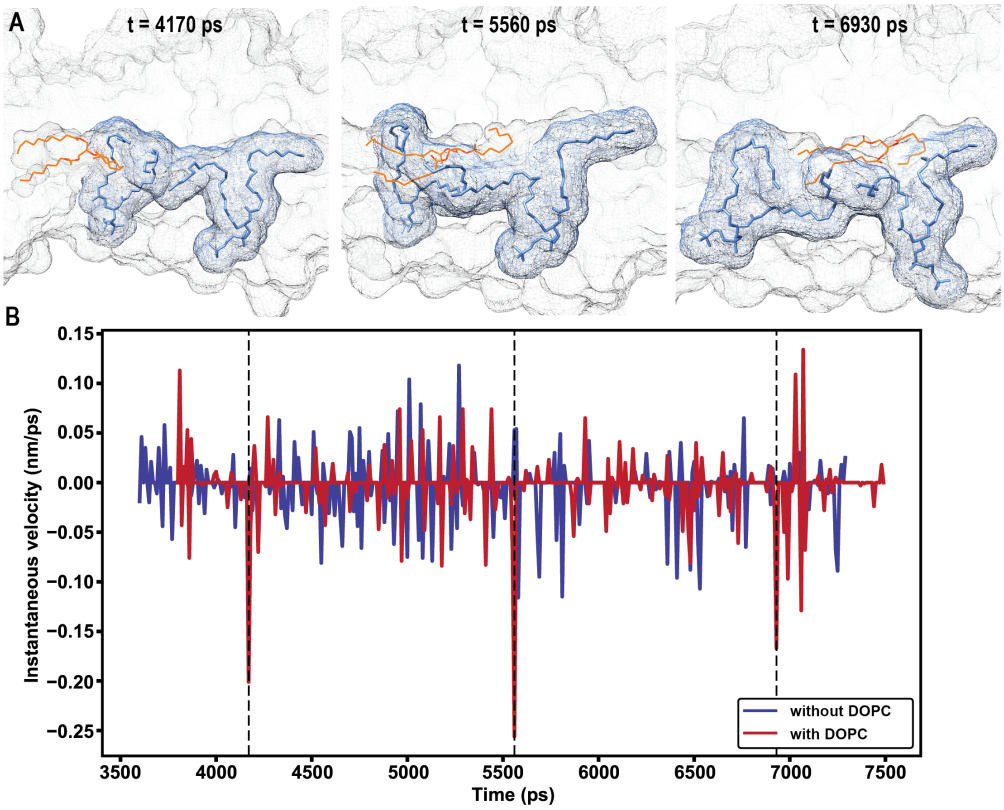



**Figure 5.10. Power-stroke movement of the acyl tails of PL. A.** The time points at which a forward gliding movement of TG is observed are depicted through simulation snapshots **B.** Instantaneous velocity of TG through the neck region of CETP.

Next, the residues lining the central pores of CETP, which surround the PL headgroups, were identified to understand their influence on PL-mediated lipid movement. These pockets are primarily comprised of charged residues (Arg, Lys, Asp, and Glu) that formed salt bridges with the PL headgroups (Fig. 5.11 A, B). R201 and R424 were mutated to Ala to further assess their effect on PL movement and lipid transfer.

In silico characterization of the two mutants revealed an increase in tunnel hydrophobicity, specifically in the neck region (Fig. 5.12 A). Moreover, the overall fluctuations

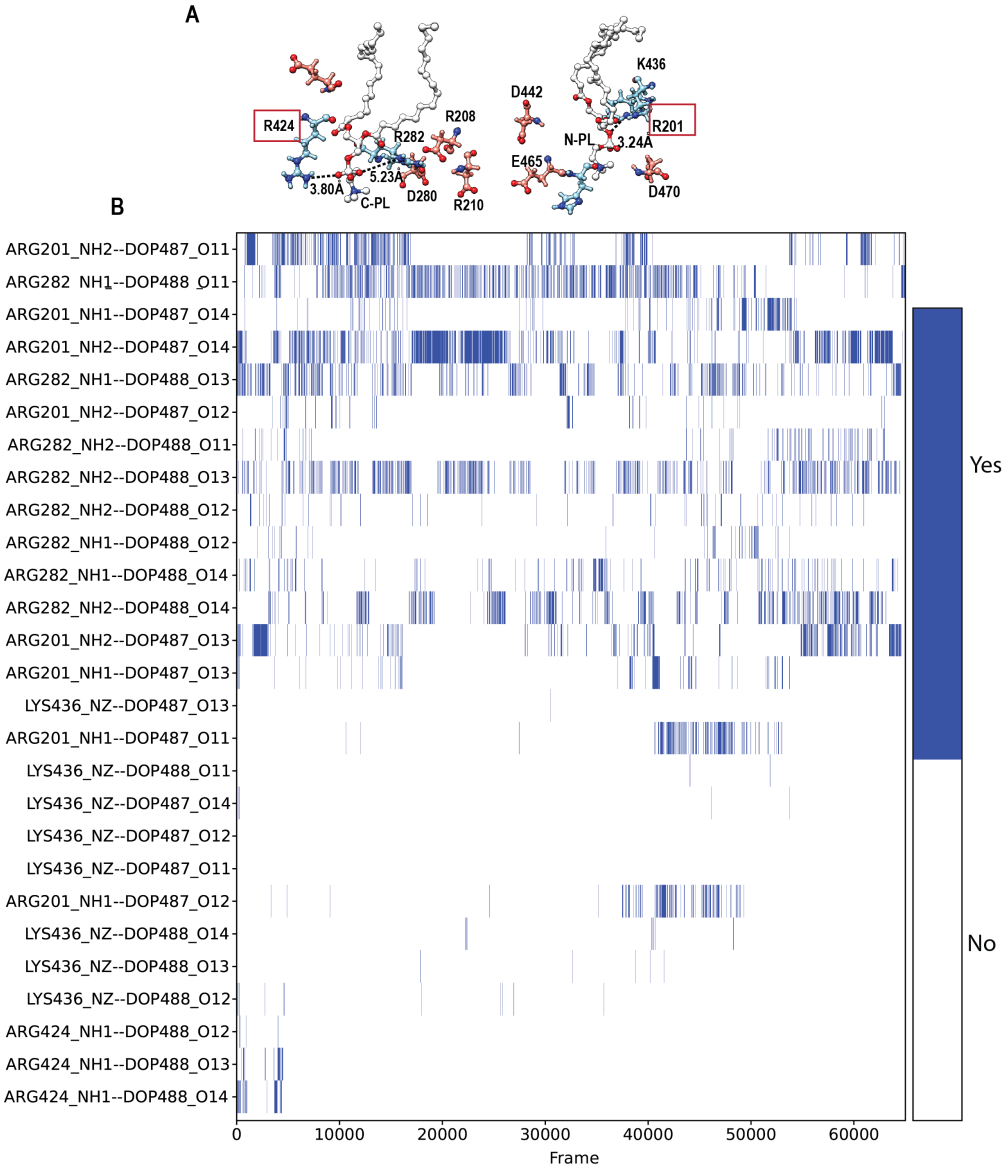



**Figure 5.11. Residues around PL headgroups.** **A.** Salt-bridge forming residues around C and N-PL plugs (PL between the C-barrel and neck, and that between the N-barrel and neck, respectively). **B.**Salt bridges between the PL headgroup and charged residues lining the PL pockets. Salt bridge distances were computed between the basic nitrogens and acidic oxygens in the PL headgroup or amino acids surrounding the PL within 6 Å

of PL plugs were reduced in the mutants compared to those in WT-CETP. Further, the extremities of acyl tails exhibited a unique orientation. To assess this, the angle formed by the ends of the acyl tails (Fig. 5.12 B) was calculated, and it was observed that they lay parallel to each other (Fig. 5.12 C,D). This may form a rigid but stable hydrophobic plane. Further, another hydrophobic plane may be formed by a TG as it approaches the unsaturated acyl tails of PL. The gliding of the two hydrophobic planes on top of each other may result in an accelerated movement of TG in the forward direction, causing a thrust (Fig. 5.12 E).

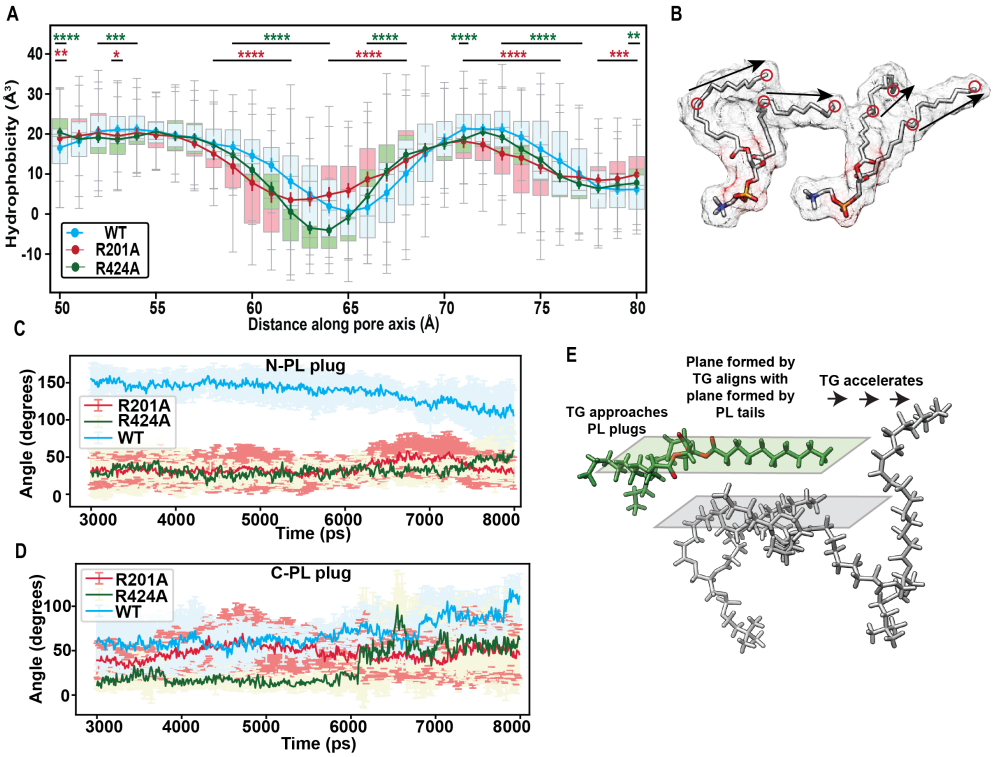



**Figure 5.12. Mechanism of powerstroke. A.** Tunnel hydrophobicity of the neck region of wild-type (WT) and mutant CETP. The box plot depicts the distribution of hydrophobicity values calculated over 100 frames (post-stabilization). Solid lines depict the median value of the distribution for the distance along the tunnel axis. The Mann–Whitney U test was used to determine the statistical significance of the differences between the profiles of WT and R201A (red) and WT and R424A (green). **B.** Schematic of vectors defining the movement of PL acyl tails. Angle between the acyl tails of **C.** N and **D.** C-PL in steered molecular dynamics (SMD) simulations. Bar plot depicting fold changes in CETP expression of the mutants with respect to that of the WT. **E.** Schematic depicting the gliding of two hydrophobic planes formed by TG and PL acyl tails.

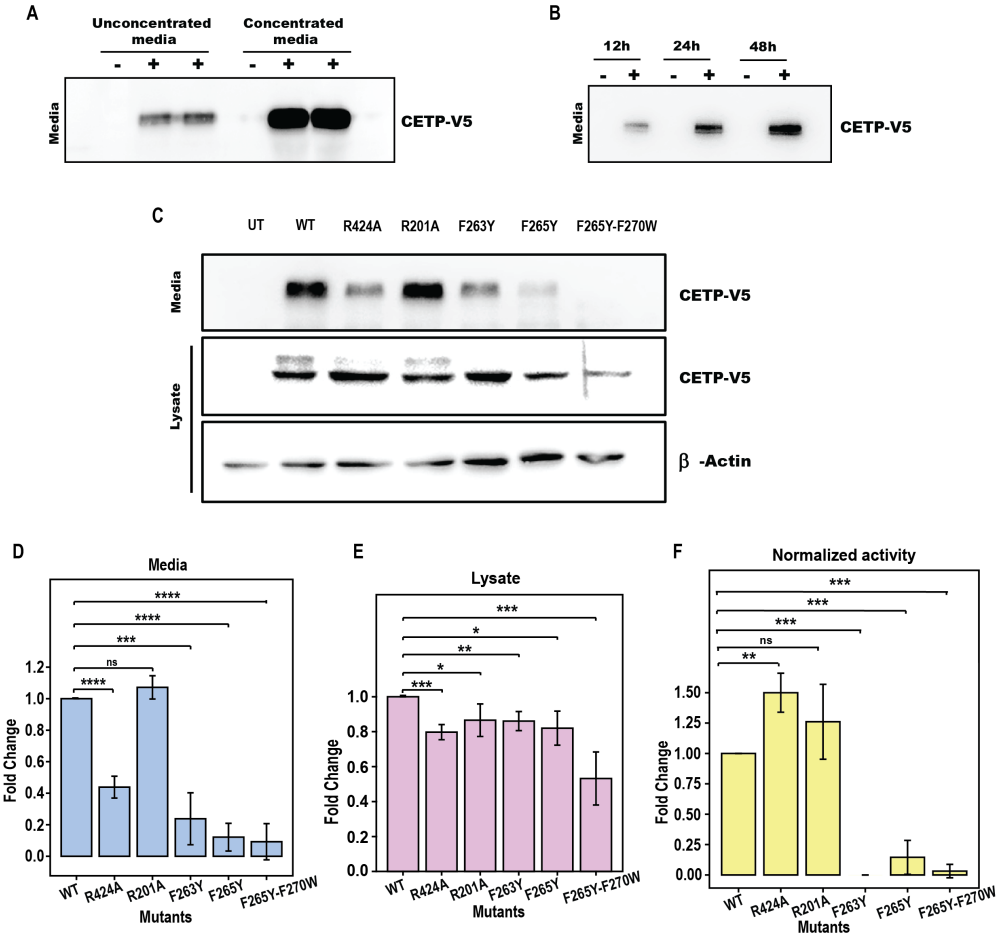



**Figure 5.13. Mutagenesis of critical CETP residues and their effect on CETP activity. A. A.** WT-CETP expression in the media of transiently transfected HEK293 cells. analyzed by Western blotting using anti-V5 antibody. **B.** WT-CETP expression in media 12h, 24h, and 48h post-transfection. **C.** Protein expression of wT-CETP & mutants in media or cell lysate. One representative blot is shown. Bar plot depicting fold changes in CETP expression of the mutants with respect to that of the WT in **D.** media & **E.** lysate **F.**Measurement of CETP activity. Values from three biological replicates are shown. The data are presented as mean ± standard deviation. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001; ns, not significant (t-test).

### 5.3.6 Mutagenesis of critical residues declines CETP activity

CETP associates with PLs primarily through hydrophobic interactions. However, the headgroup of PLs in the central pores of CETP is surrounded by charged residues (Arg and Glu) that can form salt bridges. Moreover, the acyl tails of PLs in the neck of CETP extensively interact with three Phe gates that were hypothesized to regulate the flow of lipids into and out of the tunnel. To validate this hypothesis, these Phe flaps were mutated to Tyr (F263Y, F265Y, and F265Y-F270Y), and R424 and R201 were mutated to Ala to break their favorable electrostatic interaction with the PL headgroups. HEK293 cells were transfected with plasmids expressing wild-type (WT) CETP, and the secreted protein was analyzed in concentrated or unconcentrated media (Fig. 5.13 A). Further, the expression of WT-protein post 12h, 24h & 48h of transfection was analyzed using western blotting (Fig. 5.13 B).

Subsequently, WT-CETP or mutant plasmids were transiently transfected into HEK293 cells, and the secreted protein in the media and cell lysate was analyzed using Western blotting after 48 hours. R201A showed an expression similar to that of the WT in the media (Fig. 5.13 C & D). However, R424A expression was significantly reduced (0.4-fold compared to that of the WT in the media) (Fig. 5.13 C & D). Similarly, all three Phe mutants showed poor expression in the media, with F263Y. F265Y, and F265Y-F270Y expression being 0.25-, 0.20-, and 0.1-fold, respectively, compared to that of the WT. For the PL mutants and the two gate mutants (F263Y and F265Y), the protein expression in the lysate was 0.8-times the expression of the WT, while F265Y-F270Y showed a 50% decline in protein expression in the cell lysate compared to that of the WT (Fig. 5.13 C & E). These results clearly indicated that the gate (F263Y, F265Y, and F265Y-F270Y) and PL (R424A) mutants altered CETP structure, affecting either its folding or secretion, thereby leading to decreased expression.

Furthermore, we characterized the activity of these CETP mutants by incubating equal amounts of secreted CETP (WT or mutant) with fluorescently labelled donor and acceptor particles. R201A activity was significantly higher than that of WT-CETP (Fig. 5.13 F). However, the PL mutant (R424A) showed a 30% decline in activity compared to that of the WT. Moreover, the gate mutants exhibited highly negative effects on CETP activity, with F263Y showing negligible activity, and F265Y and F265Y-F270Y showing a decline of more than 70% in their activities.

### 5.3.7 Free energy of triglyceride movement through CETP

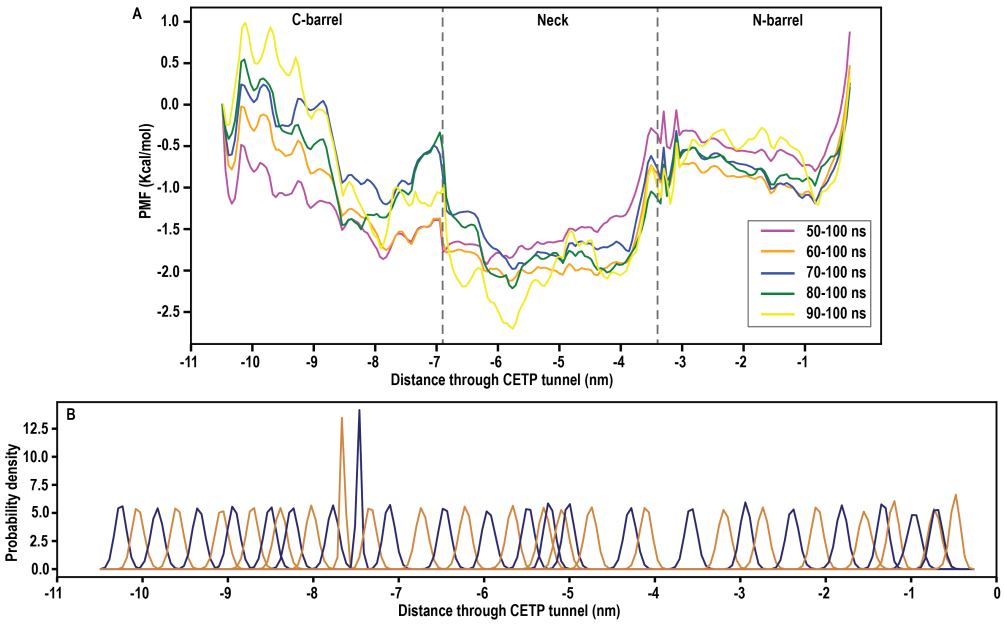



**Figure 5.14. Convergence of potential of mean force profile in the absence of PLs.** **A.** Each umbrella window was simulated for 100 ns, and PMF was calculated in blocks to ensure convergence. **B.** Overlap between adjacent umbrella histograms. In regions with low density, additional umbrella windows with stronger harmonic restraints were added.

Next, we computed the free energy of TG transfer through the CETP tunnel in the absence of presence of PLs. Using one steered MD pull as input, windows (spaced 0.1 nm apart) were chosen for umbrella sampling simulations. Umbrella sampling simulations are highly sensitive to the starting conformations for each window, as the sampling is restricted along a simple path (Lichtinger and Biggin, 2023). Further, each point along the reaction coordinate must be sampled well to obtain a meaningful free energy profile. To ensure convergence of the free energy profile, the PMF was calculated in blocks (Fig. 5.14 A & Fig. 5.15 A ). A decrease in deviation across a range of simulations indicates convergence. Moreover, a sufficient overlap in the umbrella histogram was ensured (Fig. 5.15 B).

The PMF profile suggested that the entry of TG into the CETP tunnel is thermodynamically favorable (with a downhill slope of ~ - 3.5Kcal/mol-nm) as it moves from an aqueous to a hydrophobic environment. The slope in the presence of PL is much steeper than in the absence of PL plugs (Fig. 5.16). Upon entering the tunnel, it encounters a cluster of mostly hydrophobic residues (L287, L296, E306, I307, V311, F348, F408, M412) that forms a shallow barrier of approximately +1 kcal/mol (Fig. 5.16 A

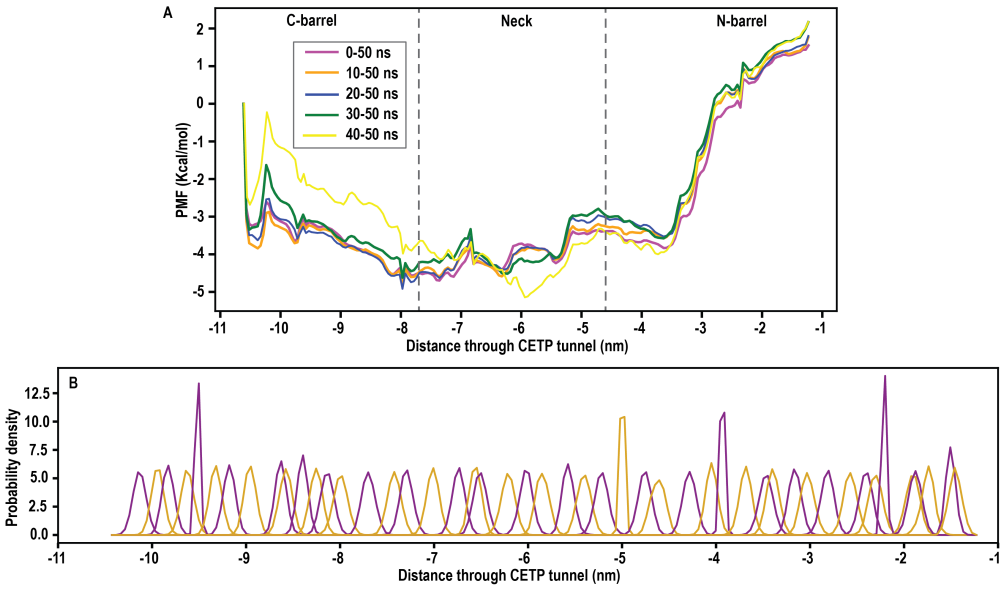



**Figure 5.15. Convergence of potential of mean force profile in the presence of PLs.** **A.** Each umbrella window was simulated for 50 ns, and PMF was calculated in blocks to ensure convergence. **B.** Overlap between adjacent umbrella histograms. In regions with low density, additional umbrella windows with stronger harmonic restraints were added.

& Fig. 5.17). Once past this barrier, the movement deeper into C-barrel is stabilizing (gradual descent of -1Kcal/mol). At the interface of the C-barrel and neck, a shallow well of depth 1Kcal/mol is observed. The movement across this minimum is guided by interaction with the acyl tail of C-PL (reaction coordinate: -8.1 nm). As TG enters the neck, it undergoes a gradual ascent (slope = 1.2Kcal/mol-nm) in three steps. The rise from the well to the first saddle presents a small energy barrier of 1Kcal/mol (Fig. 5.16 B & Fig. 5.17). This represents the opening of the F265 gate. The rise to the next saddle represents the opening of F263 gate. Finally, the movement past the third saddle into the N-barrel (the narrowest region of the tunnel) is achieved through the gliding hydrophobic interaction between N-PL acyl tail extremities and TG, resulting in a power stroke. It is essential to note that the free energy within the neck remains negative, and the local fluctuations of PL tails pull the TG forward, thereby inhibiting kinetic trapping within this region. In contrast, the PMF for the movement of TG through CETP in the absence of PL plugs shows a well of depth ~2.2 Kcal/mol. This well represents the local minima and may cause kinetic trapping of the lipid (Fig. 5.16 C ). As TG moves further into the N-barrel, closer to exit, it interacts with another cluster of mostly hydrophobic residues (Y40, Y57, L59, I117, V178, W162, F167). Furthermore, the energy barrier increases sharply (a steep ascent of 4 kcal/mol) in this region. On comparing it with the

deepest point in the PMF (interface of C-barrel & neck), this energy barrier amounts to 5Kcal/mol, which is high enough to strongly resist spontaneous passage. The steepness of this barrier is further exacerbated due to the movement of TG from a favorable hydrophobic milieu to an aqueous one. In the absence of PL plugs, a similar ascent of + ~ 1.4 Kcal/mol is observed, where the free energy becomes positive, suggesting a resistance for TG to move out of the tunnel (Fig. 5.16 D). Of note, the exit barrier in the system without PLs is lower than the system with PLs as TG exits from a less hydrophobic environment.

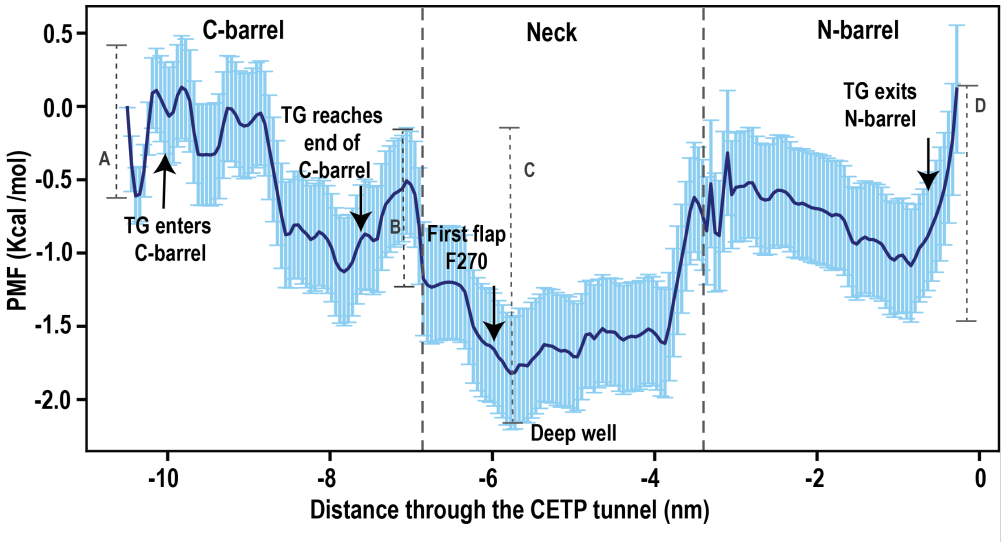



**Figure 5.16. Potential of mean force profile for the movement of TG through the CETP tunnel in the absence of PL plugs.** The major events occurring in the CETP tunnel are depicted with arrows, respectively. **A.** Energy barrier of ~1 Kcal/mol upon entry of lipid into the tunnel through C-barrel **B.** Energy barrier of ~1 Kcal/mol upon entry into the neck **C.** Energy minima within a deep well of depth ~2.2 Kcal/mol **D.** Energy barrier for exit out of the tunnel through N-barrel ~ 1.4 Kcal/mol.

## 5.4 Discussion

This chapter establishes a critical role of PL plugs in modulating the structural and functional dynamics of CETP. PL plugs enhanced the hydrophobicity of the CETP tunnel by sealing its two central orifices and rerouting TG transport to prevent loss during transit. Furthermore, they promoted forward lipid movement through favorable hydrophobic interactions between PL acyl tails and TG. Interestingly, PL acyl tails assist in the conversion of CE from a bent to linear form during its traversal through the neck region of the tunnel(Revanasiddappa, 2023). Linear CE may similarly glide over a hydrophobic

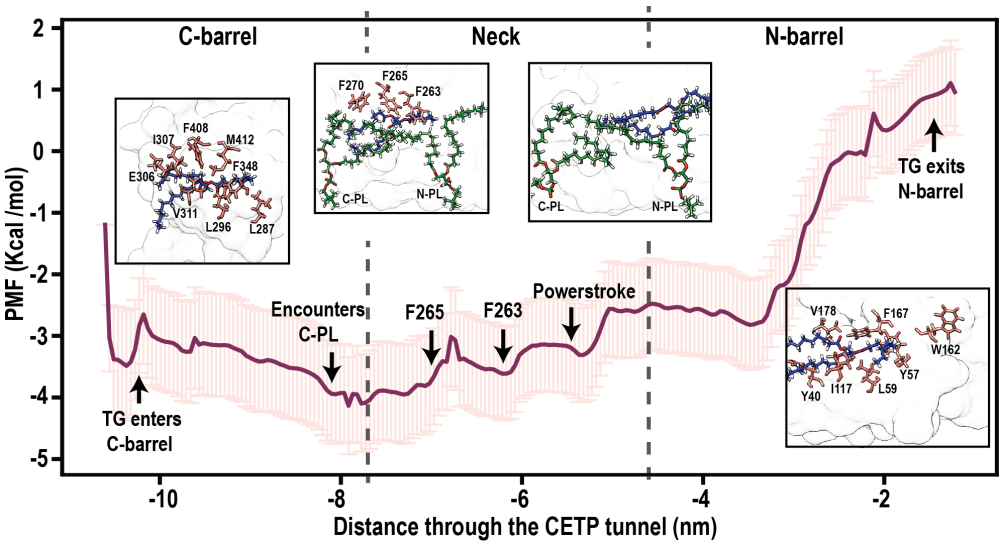



**Figure 5.17. Potential of mean force profile for the movement of TG through the CETP tunnel.** Error bars depict standard deviations from the mean and are calculated using bootstrapping. The major events that occur in the CETP tunnel are depicted with arrows and snapshots, respectively.

plane formed by PL acyl tails, as we observed for TG, for producing a forward translational motion. While the role of PLs in shielding lipid cargo from aqueous exposure has been proposed (Koivuniemi *et al.*, 2012; Cilpa-Karhu *et al.*, 2015), their impacts on tunnel hydrophobicity and lipid mobility have not been demonstrated prior to this work.

In this chapter, key structural elements within CETP were also identified, which regulated lipid traffic. Three Phe residues (F263, F265, and F270) formed flaps at the neck region, which stabilized CETP and modulated lipid passage. Given the location and orientation of these residues in the bottleneck of CETP and the distinct closed and open states exhibited by these gates, they may play a role in controlling the rate of transfer and direction of movement of the lipid cargo. Mutation of F265 or F270 to Arg abolished CETP expression, probably owing to misfolding (Qiu *et al.*, 2007), and even substitution with Tyr significantly reduced CETP expression and activity. These residues might function as directional valves as opposed to hinged gates, allowing both influx and efflux of CE and TG at differing rates (Morton and Steinbrunner, 1990). Residue gates in proteins ensure substrate selectivity, prevent solvent access to certain regions, and synchronize processes that occur in different protein compartments (Gora *et al.*, 2013). Similarly, these flaps may prevent regurgitation, such that a substrate, upon entering the tunnel, does not reflux out. Considering that lipid movement through CETP is entirely driven by diffusion, the opening and closing of these flaps are extremely important for preventing the backflow of lipid in the absence of any active, motor-driven transport.

Furthermore, the conserved nature of these residues across the BPI/LBP family members may suggest a shared mechanistic role.

Electrostatic interactions around the PL headgroup affected CETP stability and activity. Missense variants in this region (R282C and D442G) are associated with hyperalphalipoproteinemia characteristic of CETP deficiency (Nagano *et al.*, 2004). Salt bridge-forming residues R201 and R424 modulated CETP activity. Interestingly, R201 is also involved in interaction with torcetrapib (a CETP inhibitor), and its substitution for Ala increases CETP activity and half-maximal inhibitory concentration of the inhibitor (Liu *et al.*, 2012). These residues facilitated favorable hydrophobic interactions with TG, thereby increasing power-stroke-like events that propelled the lipid forward into the tunnel. These residues also increased tunnel hydrophobicity, presenting themselves as additional regulatory points that can influence CETP tunnel architecture and lipid transfer dynamics.

In addition to sealing the tunnel, PLs influenced the conformational dynamics of CETP. Their presence induced synchronized motion between the N- and C-terminal $\beta$-barrels. These concerted movements, characterized by squeezing (barrels moving toward each other) and stretching (barrels moving apart), may generate peristaltic forces that propel lipids through the tunnel. Such rhythmic dynamics may also facilitate the binding of lipoproteins. Protein domain motions influence ligand interactions (Niessen *et al.*, 2017; Turton *et al.*, 2014). Moreover, the ability of CETP to bind a second lipoprotein increases after engaging the first, suggesting allosteric communication between the termini. This effect is markedly reduced by inhibitors that prevent ternary complex formation (Zhang *et al.*, 2017). Notably, inhibitors, such as anacetrapib and torcetrapib, displace PL located at the N-barrel–neck interface (Liu *et al.*, 2012). Taken together, the displacement of PLs by inhibitors may potentially disrupt barrel synchrony and impede lipid transfer. These findings underscore the importance of PLs in regulating the mechanical and functional behaviors of CETP.

Given the indispensable role of PLs, CETP likely circulates in plasma in a PL-bound form. However, PL incorporation post-synthesis via lipoprotein binding appears unlikely: electron microscopy shows that the PL-binding pocket lies > 15 Å far from the HDL surface (Zhang *et al.*, 2012). Similarly, PL entry via the hydrophobic tunnel, followed by their rotation to seal the orifices, is geometrically implausible. Alternatively, a

scenario in which CETP acquires PLs through its central concave groove by binding to the lipoprotein (Qiu *et al.*, 2007) does not adequately account for the precise orientation required for tunnel plugging. Therefore, it is proposed that PLs are integrated during CETP folding or secretion. Furthermore, PL plugs are also present in bacterial BPI, another member of the BPI/LBP family (Beamer *et al.*, 1997), suggesting that PL-mediated lipid transport may be a conserved mechanism across the family.

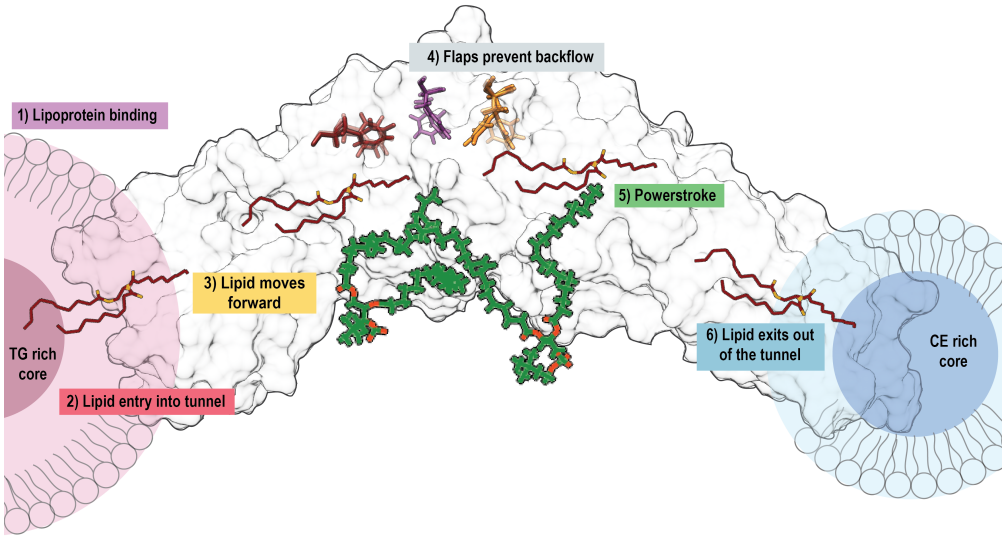



**Figure 5.18. Mechanism of lipid transfer through CETP.** PL-induced domain synchrony results in binding of CETP to lipoproteins (Step 1). Upon CETP attachment and penetration into the lipoprotein surface, neutral lipids present near the surface enter the CETP tunnel owing to a difference in concentration gradients (Step 2). Once inside the tunnel, peristaltic forces generated by domain movements of CETP guide the lipid further into the tunnel. Additionally, transient contacts with the hydrophobic clusters located inside the tunnel steer the lipid forward (Step 3). As an internal check in the neck region, the concerted movement of the flaps ensures no lipid reflux (Step 4). Subsequently, the hydrophobic interactions between PL acyl tails and lipid lead to the acceleration of lipid forward (Step 5). On reaching the end of the tunnel, lipid exits owing to a difference in concentration gradients (Step 6).

A comparison to the previously identified path for CE (Lei *et al.*, 2016; Dixit *et al.*, 2019) with the path of TG was also made, and it was observed that both lipids travel through the same route. Since CETP inhibitors (torcetrapib and anacetrapib) occlude the tunnel by interacting with residues C13, H232, F263, A202, and V198, which lie in the path of TG, the binding of inhibitors may affect TG movement, similar to the effect on CE movement (Liu *et al.*, 2012; Dixit *et al.*, 2019; Chirasani and Sanjib, 2017).

Since both the lipids travel through the same route, and their point and direction of entry or exit have no influence on the path that they take, the difference in the rates of CE and TG transfer (Morton and Izem, 2014) must be owing to additional factors such as surface availability of these neutral lipids in lipoproteins (Morton and Greene, 2003*b*) or auxiliary proteins that influence binding and stability of CETP with lipoproteins (Cilpa-Karhu *et al.*, 2015). Furthermore, missense single-nucleotide variations of several of these TG-interacting residues (I11V, V136G, I413V, M422I, R352H, V295E, Y361C, V416G, I64N, I205F, V340G, V311I, V409M, L273P, V323I, V344A, L432I, C13Y, R292C, R352C, L206R, D382H, A373P, V89M, and A195D) have been predicted as damaging and potentially affecting CETP function (Table. 5.6). These mutations may alter tunnel architecture, thereby negatively influencing CETP-mediated lipid transfer.

Based on these findings, a multi-step mechanism of lipid transfer is proposed (Fig. 5.18): (1) PL-induced domain synchrony enables lipoprotein docking and penetration; (2) neutral lipids enter the tunnel via CETP termini, guided by concentration gradients; (3) lipids traverse the tunnel through sequential hydrophobic interactions; (4) Phe residues at the neck function as a gating systems for preventing backflow; (5) hydrophobic interactions with PL acyl chains generate a power-stroke effect, propelling lipids forward; and (6) lipids exit through the distal end, driven by differences in lipid concentrations. Further, the presence of PL plugs alters the overall hydrophobicity of the tunnel, resulting in a more negative free energy of TG entry into the CETP tunnel. Further kinetic fluctuations within the neck region ensure lipid isn't trapped inside the neck. The PMF showed a very high exit barrier into the aqueous environment, which is strong enough to resist the spontaneous diffusive movement of TG. However, the presence of lipoproteins at the ends of CETP may significantly reduce this barrier, further supporting the formation of the ternary CETP complex during lipid transfer.

In summary, this chapter demonstrates that PL plugs act as structural cofactors, which enhance the lipid-transport efficiency of CETP by modulating tunnel hydrophobicity and conformational dynamics. This PL-mediated mechanism may extend to other BPI/LBP family proteins, such as PLTP, LBP, and BPI, offering a broad framework for understanding lipid-transfer processes. F265, F263, and F270 flap and salt bridge-forming residues are important regulatory points in the CETP structure, which influence its stability and activity. CETP inhibition is focused on tunnel occlusion, which requires hydrophobic drug molecules that can enter the CETP tunnel. This work suggests an al-

ternative strategy, in which the design of therapeutics that target the solvent accessible PL-binding pockets may effectively control CETP function in patients with atherosclerosis.

| Sno. | Accession | dbSNP ID | Protein change | Location | Mapped on protein | SIFT | PolyPhen2 | Panther |
|---|---|---|---|---|---|---|---|---|
| 1 | VCV003491136 | | I28V | Neck | I11 | - | possibly damaging | - |
| 2 | VCV003143283 | rs2543645399 | V153G | N-barrel | V136 | - | probably damaging | probably benign |
| 3 | VCV003143280 | rs762811493 | I430V, I370V | C-barrel | I413,I353 | - | benign | probably benign |
| 4 | VCV003019020 | rs760836166 | M439I, M379I | C-barrel | M422, M362 | - | benign | probably bengin |
| 5 | VCV002980442 | rs144949752 | R369H, R309H | C-barrel | R352, R292 | Tolerated | benign | probably bengin |
| 6 | VCV002864541 | rs1323326418 | V252E, V312E | C-barrel | V235, V295 | - | possibly damaging | - |
| 7 | VCV002722544 | rs36122917 | Y318C, Y378C | C-barrel | Y301, Y361 | Tolerated | probably damaging | |
| 8 | VCV002623272 | rs1567476542 | V373G, V433G | C-barrel | V356, V416 | - | probably damaging | probably benign |
| 9 | VCV002539456 | rs142459781 | R299C | Salt-bridge | R282 | Deleterious | - | - |
| 10 | VCV002302010 | rs2543643841 | I81N | N-barrel | I64 | - | probably damaging | probably benign |
| 11 | VCV002301233 | rs933760754 | I222F | Neck | I205 | - | benign | |
| 12 | VCV002238319 | rs781348259 | V357G, V297G | C-barrel | V340, V280 | - | probably damaging | probably benign |
| 13 | VCV002173132 | rs147412224 | V388I, V328I | C-barrel | V311, V371 | Tolerated | benign | probably benign |
| 14 | VCV002078436 | rs5887 | V426M, V486M | C-barrel | V409, V469 | Deleterious | | probably |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 15 | VCV001804584 | rs369941536 | L290P | C-barrel | L273 | Deleterious | probably damaging | - |
| 16 | VCV001405295 | rs141310739 | V280I, V340I | C-barrel | V263, V323 | Tolerated | benign | probably benign |
| 17 | VCV001366232 | rs200134880 | V361A, V301A | C-barrel | V344, V284 | Tolerated | probably damaging | probably benign |
| 18 | VCV001357558 | rs1164754201 | L449I, L389I | C-barrel | L432, L372 | - | possibly damaging | probably benign |
| 19 | VCV001314022 | rs771585518 | C30Y | Neck | C13 | - | probably damaging | probably benign |
| 20 | VCV001305453 | rs774433762 | R309C, R369C | C-barrel | R292, R352 | - | - | probably benign |
| 21 | VCV000886619 | rs747056056 | L223R | Neck | L206 | - | possibly damaging | - |
| 22 | VCV000885657 | rs1326253991 | D399H, D459H | C-barrel | D382, D442 | - | - | probably benign |
| 23 | VCV000319993 | rs5880 | A390P, A330P | C-barrel | A373, A313 | Deleterious | probably damaging | probably benign |
| 24 | VCV000319976 | rs371258270 | V106M | N-barrel | V89 | Deleterious | probably damaging | probably benign |
| 25 | VCV003716673 | | | | | | possibly damaging | |

**Table 5.6. Missense single nucleotide variations of TG interacting residues along with their pathogenicity prediction**

# CHAPTER 6

# Factors Affecting Lipid Movement within the CETP Tunnel

## 6.1 Introduction

Lipid transfer through CETP is dependent on its ability to form stable lipoprotein complexes, which enable access to the lipid core and the acquisition of neutral lipids. This is further supported by the fact that free circulating CETP is inactive and that lipoprotein concentration determines CETP activity (Charles and Kane, 2012). Therefore, lipid transfer depends on lipoprotein abundance, substrate availability within lipoproteins and formation of a stable lipoprotein-CETP complex in addition to protein-specific factors discussed in the previous chapter. This chapter highlights some of these additional factors that govern CETP-mediated lipoprotein remodelling.

**1. Lipoprotein binding selectivity**

CETP has asymmetric $\beta$-barrels of differential length; the N-barrel is tapered while the C-barrel is globular. Using electron microscopy, it was observed that the N-barrel selectively penetrates HDL, while the C-barrel associates with LDL or VLDL (Zhang *et al.*, 2012). Furthermore, incubating CETP with H300 (a C-terminus-specific antibody) blocked the formation of the LDL-CETP complex but not the HDL-CETP complex (Zhang *et al.*, 2017). Alternatively, on incubating CETP with antibody fragments Fab 6/2, Fab JHC1, and Fab 6/17 that were selective for the N-terminal barrel, concave groove of the N-terminal barrel, and the C-terminal barrel, respectively, HDL-CETP complexes were observed in all three cases, implying that CETP can bind HDL through its N- or C-terminal tip (Lauer *et al.*, 2016). Given this contradictory evidence, it is difficult to associate CETP barrels with lipoprotein binding selectivity. It is possible that a specific interaction between the terminal barrel residues of CETP initiates lipoprotein sensing and penetration.

**2. Formation of a stable lipoprotein complex**

The circulating molar ratios of lipoproteins are the most important physiological determinant of CETP activity. Since HDL ($\sim 13\mu$M) far exceeds the plasma levels of LDL ($\sim 2\mu$M) and VLDL ($\sim 0.1\mu$M), CETP-HDL complexes are more prevalent than CETP-LDL or CETP-VLDL complexes (Charles and Kane, 2012). Additionally, CETP binding to the lipoprotein surface is enhanced by the negative charge of the lipoprotein, either due to apolipoprotein acylation (Pattnaik and Zilversmit, 1979) or accumulation of products of lipolysis, especially fatty acids (Sammett and Tall, 1985). The latter may be an important metabolic step through which lipoprotein lipase and CETP accomplish lipoprotein remodeling. However, the relationship between CETP's affinity for lipoprotein and its activity is not linear (Morton and Greene, 2003*a*). This suggests that CETP activity may be dependent on the presence of an optimum surface charge (Nishida and Nishida, 1993; Masson *et al.*, 1996; Morton and Greene, 2003*a*). A large increase in CETP binding affinity may lead to an inability to detach from lipoproteins, thereby decreasing activity.

Another important physiological regulator of CETP, apoC1, regulates CETP activity by altering HDL surface charge (Rouland *et al.*, 2022). Upon binding to HDL apoC1, it activates the LCAT reaction, leading to the maturation of HDL and the inhibition of phospholipase A2, which decreases HDL catabolism. The presence of a positively charged, lysine-rich cluster in the C-terminal alpha helix of apoC1 further alters the negative surface charge of HDL, leading to the disassociation of HDL-CETP complexes (Dumont *et al.*, 2005). Interestingly, apoC1 does not inhibit CETP when bound to VLDL (Dautin *et al.*, 2007), suggesting that the binding orientation of apoC1 on VLDL may differ from that of HDL. Alternatively, it is suggested that the accumulation of TG-rich lipoproteins may result in the loss of the inhibitory capability of apoC1 (Pillois *et al.*, 2012). This non-functional apoC1 associates with VLDL, further inhibiting their clearance and exacerbating plasma TG accumulation. The structural differences, if any, between the functional and non-functional apoC1 have not been studied.

**3. Surface availability of neutral lipids**

Several studies have suggested that CETP obtains neutral lipids that are solubilized within the phospholipid bilayer of the lipoprotein, rather than through the lipoprotein core, as these are more accessible (Miller and Small, 1983; Morton and Steinbrunner, 1990). In the absence of a mechanism for selective uptake, both surface available CE

and TG above an effective concentration of ~ 0.1 mole% may compete for transfer by CETP (Morton and Steinbrunner, 1990). Moreover, since CETP penetrates into lipoproteins at differential depths (HDL: ~55 Å; VLDL: ~ 20-25 Å) (Charles and Kane, 2012), therefore, the surface availability of TG more than CE may be of importance. The acyl chain composition of neutral lipids, such as chain length and unsaturation, may further introduce disorder within the surface PL monolayer of lipoproteins, possibly affecting substrate accessibility to CETP (Foucher *et al.*, 1996).

**4. Lipid entry**

Since the evidence suggests that both N- and C-terminal barrels may bind either HDL/LDL or VLDL, it is likely that the termini are non-specific for either TG or CE. Furthermore, the surface availability of both lipids may be a crucial factor influencing lipid entry. However, there may be a possibility that the lipid entry through the N- or C-barrel is hindered by selectivity "filters" within the barrel opening, such that the N-barrel allows only CE while the C-barrel allows only TG to enter. MD simulations studying the CETP-HDL complex have proposed that F35, F93, and F147 play a role in CE entry through the N-barrel (Cilpa-Karhu *et al.*, 2015). However, whether these residues are selective towards CE and inhibit TG entry remains unexplored.

**Simultaneous exchange of lipids**

CETP-mediated neutral lipid transfer is hypothesized to be a "bi-directional" exchange, as it is dependent on the concentration of neutral lipids at the surface of both the donor and acceptor particles (Morton and Steinbrunner, 1990). During this exchange, the unidirectional movement of one lipid is tightly coupled to the successful on-loading of another exchangeable lipid substrate, such that this "forward" and "reverse" reaction is kinetically coupled (Morton and Greene, 2003*a*). Further, CETP is known to facilitate both a homoexchange (exchange of the same substrate) or heteroexchange (exchange of CE for TG or vice-versa); however, only the heteroexchange activity is associated with lipoprotein remodeling (Serdyuk and Morton, 1999). Therefore, characterizing the movement of two different neutral substrates across the CETP tunnel is of paramount importance.

This chapter addresses these additional factors that influence lipid movement through CETP and, in turn, CETP activity. Selectivity for the entry of lipid through the N- or

C-barrel is addressed by steering the movement of lipid in both directions, followed by free energy calculations through umbrella sampling. While this does not directly address the binding specificity of CETP barrels to lipoproteins, it does help identify important residues that enable or hinder lipid entry into the tunnel. Similarly, a bi-directional heteroexchange is simulated to identify and understand the structural dynamics of CETP that allow the simultaneous movement of two lipids through the tunnel. However, lipoprotein-specific factors, such as their composition, surface availability of substrate, presence of apoC, and negative surface charge that influence CETP binding affinity and activity, aren't explicitly discussed.

## 6.2 Methodology

### 6.2.1 Steered MD Simulations set-up

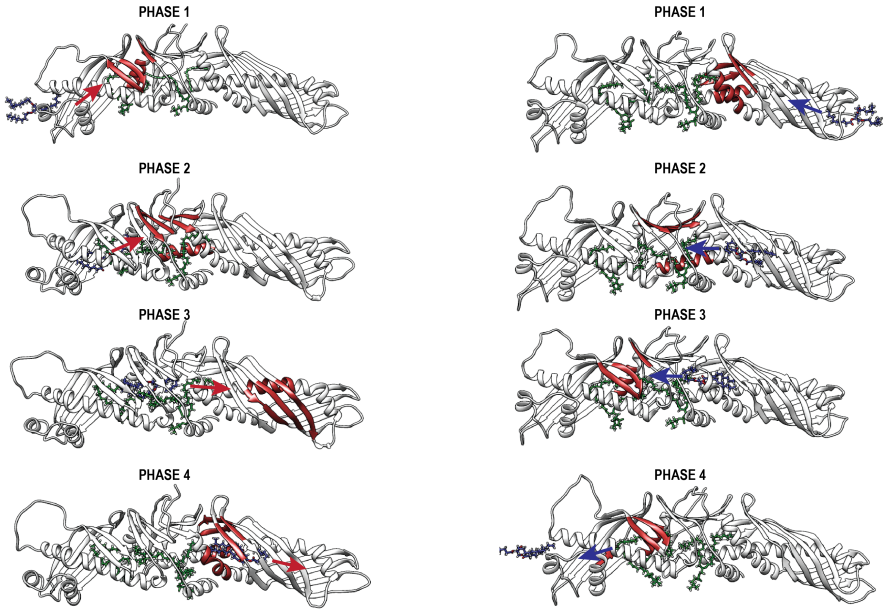



**Figure 6.1. Schematic depicting the four phases in which TG was steered through the CETP tunnel in which it entered through the C-barrel (forward) or N-barrel (reverse).**TG was placed at the entrance of the C-barrel or N-barrel. A pull force was applied on the center of mass of the termini atoms of TG to pull it towards the center of mass of CETP tunnel residues in the direction of the N- or C-terminal barrel, respectively (red) (Phase 1-3). Subsequently, TG is pushed out of CETP through the N-barrel or C-barrel (Phase 4) by applying a push force instead of a pull force.

**Simulations for lipid entry through C-barrel or N-barrel** Following the conventional MD run, the last frame of the trajectory was used as the starting point for SMD

simulations. One TG (10:0/10:0/10:0) was manually placed at the opening of the C-barrel or N-barrel of CETP. TG was pulled through the CETP tunnel using constant-velocity SMD simulations with a force constant of 1000 kJ-mol-1-nm-2. The pulling vector was defined on the basis of the center of mass of the steered group, i.e., the tail region of TG and the center of mass of CETP tunnel residues in the direction of the N-barrel or C-barrel. Considering the asymmetric shape of CETP, its center of mass was dynamically switched from the C-barrel to the neck and finally N-barrel or vice-versa, depending on the location of the steered group. Therefore, the entire pull simulation through the CETP tunnel was executed in four phases. For each system, TG was pulled with a velocity of 1 nm/ns. This rate was chosen to obtain an optimal balance of accuracy and computational speed. Simulations were performed in replicates of three (Table 6.1). Fig. 6.1 depicts the four phases in which SMD was performed.

| System | Number of protein atoms | Number of lipid atoms | Number of water atoms | System size | Number of replicates |
|---|---|---|---|---|---|
| Forward (C- to N-barrel) | 7,440 | 276 + 101 | 207,009 | 214,826 | 6 |
| Reverse (N- to C-barrel) | 7,440 | 276 + 101 | 207,051 | 214,868 | 3 |

**Table 6.1.** Table summarizing the steered MD simulation setup to characterize CETP's barrel specificity towards lipid entry or exit.

**Potential of Mean Force Calculation** To calculate the PMF, we performed umbrella sampling simulations on the path taken by TG derived from our SMD simulations. Starting from the initial position of TG at the mouth of the C-barrel or N-barrel, the path traversed through the tunnel was sampled at every 0.1 nm, resulting in ~ 52 windows. A force constant of 1000 kJ-mol-1-nm-2 was used. To ensure sufficient overlap of histograms along the reaction coordinate, a force constant of 5000 kJ mol-1 nm-2 was used for certain windows to ensure sufficient sampling in specific regions of the reaction coordinate. These restraining umbrella potentials ensured the sampling of every region of the conformational space along the reaction coordinate, which would not otherwise be accessible during direct sampling. Each window was thoroughly equilibrated for 1 ns and subjected to a production run for 50 ns or 100 ns to ensure convergence of PMF. Hence, a series of histograms containing a biased distribution of the reaction coordinate from each window was obtained. These histograms were unbiased and combined to produce the PMF using the weighted histogram analysis method (WHAM) (Kumar *et al.*, 1992). The last 30 ns of the trajectory were used to calculate the PMF for both systems

(forward and reverse). The first and last bins of the reaction coordinate were assumed to be neighbors to generate a periodic PMF. The convergence of PMF was ensured by checking the overlap between individual histograms representing the distributions of sampled configurations and by calculating the deviations between individual PMFs for different blocks of the simulation. An average PMF profile was generated by considering the mean potential at each point along the reaction coordinate, and errors were estimated using 100 bootstrap runs. Table 6.2 summarizes the US simulation setup.

| System | System size | Number of windows | Duration of each window | Total time |
|---|---|---|---|---|
| Forward (C- to N-barrel) | 214,826 | 48 | 50 ns | 2.4 µs |
| Reverse (N- to C-barrel) | 214,868 | 50 | 100 ns | 5.0 µs |

**Table 6.2.** Table summarizing the umbrella sampling simulation setup to characterize the free energy change for lipid entry through N or C-barrel.

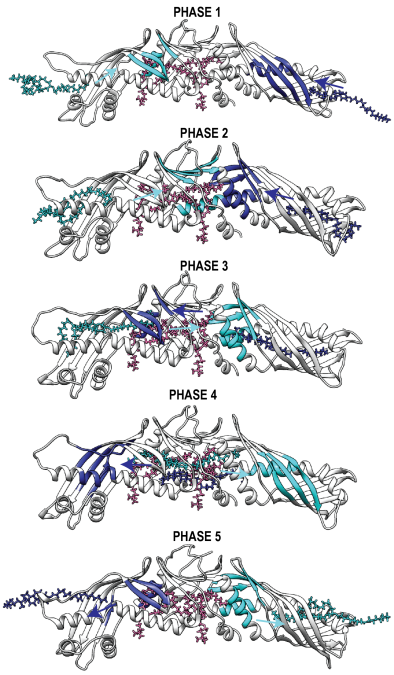



**Figure 6.2. Schematic depicting the five phases in which TG & CE were steered through the CETP tunnel simultaneously.** TG & CE were placed at the entrance of the C- or N-barrel domain, respectively, and a pull force was applied on the center of mass of the termini atoms of TG & CE simultaneously pulling them towards the center of mass of CETP tunnel residues in the direction of the N-terminal barrel (cyan) for TG and C-barrel for CE (blue). Subsequently, both lipids are pushed out of CETP through the N- & C- barrel end, respectively by applying a push force instead of a pull force.

**Simulations for simultaneous transfer of lipids through CETP** To understand heteroexchange, a TG (18:1/18:1/18:1) was placed at the C-barrel mouth, while a CE (18:1) was placed at the N-barrel mouth. The acyl chain lengths of both the lipids were intentionally kept the same so they may traverse the entire tunnel in a similar time. Both lipids were then pulled into the tunnel using a constant velocity of 1 nm/ns by implementing a force constant of 1000 kJ-mol-1-nm-2. Two pulling vectors were defined. First, on the basis of the center of mass of the steered group, i.e., the tail region of TG and the center of mass of CETP tunnel residues in the direction of the N-barrel. Second, on the basis of the center of mass of CE, i.e., its head region, and the center of mass of CETP tunnel residues in the direction of C-barrel. Considering the asymmetric shape of CETP, especially the differential length of the two barrels, its center of mass was dynamically switched throughout the pull, depending on the location of the steered group. The entire pull simulation through the CETP tunnel was executed in five phases (Fig. 6.2). Table 6.3 summarizes the setup.

| System | Number of protein atoms | Number of lipid atoms | Number of water atoms | System size | Number of replicates |
|---|---|---|---|---|---|
| 2 lipid SMD | 7,440 | 276 (PL) + 125 (CE) + 167 (TG) | 206,898 | 214,921 | 6 |

**Table 6.3.** Table summarizing the steered MD simulation setup for simultaneous movement of two lipids within CETP tunnel.

### 6.2.2 Analysis

**Change in CETP tunnel volume as lipid traverses through the protein** For every window in the umbrella sampling simulation, a frame was extracted every 10 ns post-stabilization using gmx trjconv and aligned with respect to the first frame. The volume of the tunnel was calculated using CICLOP (Garg *et al.*, 2022) for each frame.

Changes in tunnel volume during simultaneous exchange of lipids were calculated using replicate steered MD trajectories, similarly as described above.

**Analysis of Path traversed by TG through CETP tunnel and calculation of contacts** For each SMD run where TG entered either through the C-barrel or N-barrel, the position of the C1 carbon of TG was traced throughout the simulation trajectory. The structures were visualized and rendered using UCSF Chimera (Pettersen *et al.*, 2004) to trace the

path taken by TG. Contact analysis was performed using CONAN (Mercadante *et al.*, 2018) to identify CETP residues within the radius of 0.3–0.6 nm of TG as it traversed the tunnel. Residues within this contact radius for > 1 ns were extracted for each SMD run. Residues that were common in all replicates are reported.

**Calculation of residence time of TG in the CETP tunnel** The distance between the group of atoms of TG and the group of atoms constituting every residue of protein was calculated within a radius of 6Å for every frame in the simulation to generate a contact matrix. The lifetime of contact was calculated by dividing the total number of contacts made between a residue and TG or CE by the total simulation time.

A time auto-correlation function (ACF) $C_x(t)$ was used as a measure for the presence of a contact between TG and protein residue, and hence a measure of its residence time. Using the contact matrix generated above, all residues having a contact lifetime 0.05 were considered as the population of TG interacting residues at the start point $t_0$. The lifetime of these bonds is monitored over time to build an average behavior.

$$C_x(t) = \left\langle \frac{h_{ij}(t_0)h_{ij}(t_0+t)}{h_{ij}(t_0)^2} \right\rangle \tag{6.1}$$

The continuous definition measures the time that a particular hydrogen bond remains continuously attached, whereas the intermittent definition allows a bond to break and then subsequently reform, allowing it to be counted again. The relevant lifetime, $\tau_x$ can then be found via integration of this function

$$\tau_x = \int_0^\infty C_x(t)dt \tag{6.2}$$

The observed behaviour was then fit to a multi-exponential function using 2 exponents.

**Stability of hydrophobic plane formed by PLs** The plane formed by the acyl tails of PL was measured by calculating the vectors formed between C29 and C218 (v1) and C39 and C318 (v2) for each PL. The angle between the two vectors depicts the overall conformation of PL tails in the tunnel, such that an angle close to 0° represents that the

two tails of PL are parallel to each other.

## 6.3 Results

### 6.3.1 Lipid entry directionality is promiscuous in nature

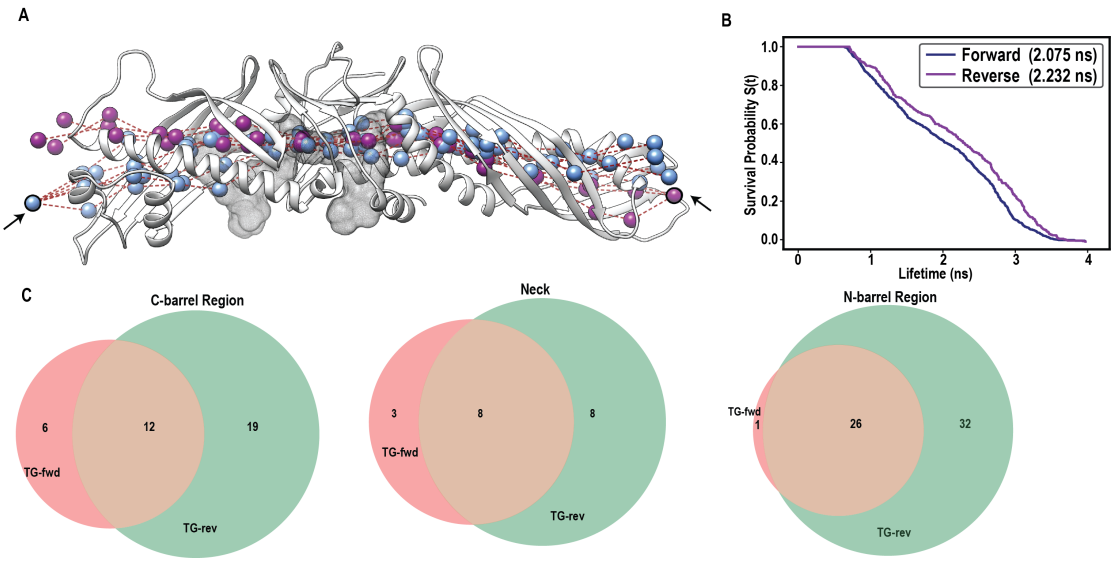



**Figure 6.3. Path taken by TG as it enters CETP through C- (forward) or N-barrel (reverse). A.** Tracing the center of mass of the glycerol moiety of TG as it moves through the tunnel, in each replicate simulation of CETP, where TG moves from C-barrel to N-barrel (blue circle) or from N-barrel to C-barrel (purple circle). The starting point for both simulations is highlighted with a black arrow. Multiple paths derived from three independent replicate simulations are depicted. **B.** Survival probability function of contacts between TG and CETP tunnel residues when TG enters through C- or N-barrel. **C.** Venn diagram showing the common tunnel residues within a radius of 0.3–0.6 nm staying in contact with TG for > 1 ns of simulation time in all three replicates of forward or reverse simulations.

Initially, it was hypothesized that the C-barrel of CETP selectively associated with TG-rich VLDL, which suggested that lipid transfer would proceed from the C-barrel to the N-barrel. However, experimental observations revealed that the CETP barrels bound non-specifically to both CE-rich and TG-rich lipoproteins (Lauer *et al.*, 2016). For successful lipid exchange to occur, the lipid must fully enter and traverse the CETP tunnel. To investigate barrel-specific lipid entry, we hypothesized that the N-barrel selectively admitted cholesteryl ester (CE) and the C-barrel selectively admitted triglyceride (TG). To test this, we set up two systems: one with a single TG placed at the C-barrel opening

(forward) and one with it placed at the N-barrel opening (reverse) Fig. 6.1. We then used the Steered Molecular Dynamics (SMD) methodology, as described in the Methods, to pull the TG through the CETP tunnel in four distinct phases (Chapter 5, Fig. 5.1).

| Domain | Forward (C-barrel to N-barrel) | Reverse (N-barrel to C-barrel) | Common residues |
|---|---|---|---|
| N-barrel | S119 | L23, A28, Q32, Y40, I43, L52, V55, K56, Y57, L59, I62, I64, I69, S72, V74, L76, L97, Y99, G100, Y101, W105, W106, I113, F115, L152, L154, W162, I163, L166, F167, I171, L175 | L20, N24, T27, I31, F35, L67, I82, V84, I86, V89, V91, F93, I117, I121, L123, I125, T127, L129, T138, L145, F147, L179, I183, I187, S191, M194 |
| Neck | I11, I215, H232 | I15, I205, P221, S230, L261, L455, L457, M459 | C13, A195, V198, A202, L206, L228, F263, F265 |
| C-barrel | V311, F348, F350, V359, Y361, L409 | I331, V338, S342, V344, F363, E365, I367, T369, A373, L380, L382, L384, F387, I389, Q406, Q410, I418, M422, M433 | F270, L273, V323, V340, V346, V371, L409, I413, G417, V421, L425, F429 |

**Table 6.4.** Residues that make contact with TG > 1ns of simulation time in all three replicates of CETP within 0.3-0.6 nm.

The TG molecule successfully entered the CETP tunnel and exited through the opposite end without any hindrance, regardless of whether it entered through the N- or C-barrel. Tracing the center of mass revealed that the path taken by the TG was identical, irrespective of its entry point (Fig. 6.3 A). Interestingly, when TG was pulled from the N-barrel to the C-barrel (the reverse pull), it contacted a greater number of residues, leading to a longer residence time in the tunnel. This residence time was ~ 1.07 times longer than the forward pull (C-barrel to N-barrel) (Fig. 6.3 B).

Furthermore, the set of residues (within 0.6nm) that contacted the TG was highly similar in both pull directions. These contact residues included:

N-barrel: L20, N24, T27, F35, L67, I82, V84, I86, V89, V91, F93, I117, I121, L123, I125, T127, L129, T138, L145, F147, L179, I183, I187, S191, M194.

Neck region: C13, A195, V198, A202, L206, L228, F263, F265.

C-barrel: F270, F273, V323, V340, V346, V371, L409, I413, G417, V421, L425, F429 (Table. 6.4 & Fig. 6.3 C).

### 6.3.2 Lipids can traverse through the tunnel irrespective of their entry point.

To assess if the lipid's movement—in either the forward (C- to N-barrel) or reverse (N- to C-barrel) direction—induced conformational changes in CETP, the Root-Mean-Square Deviation (RMSD) was calculated during steered MD. CETP exhibited a greater deviation when the lipid exited the tunnel (~20 Å) while moving from the C- to the N-

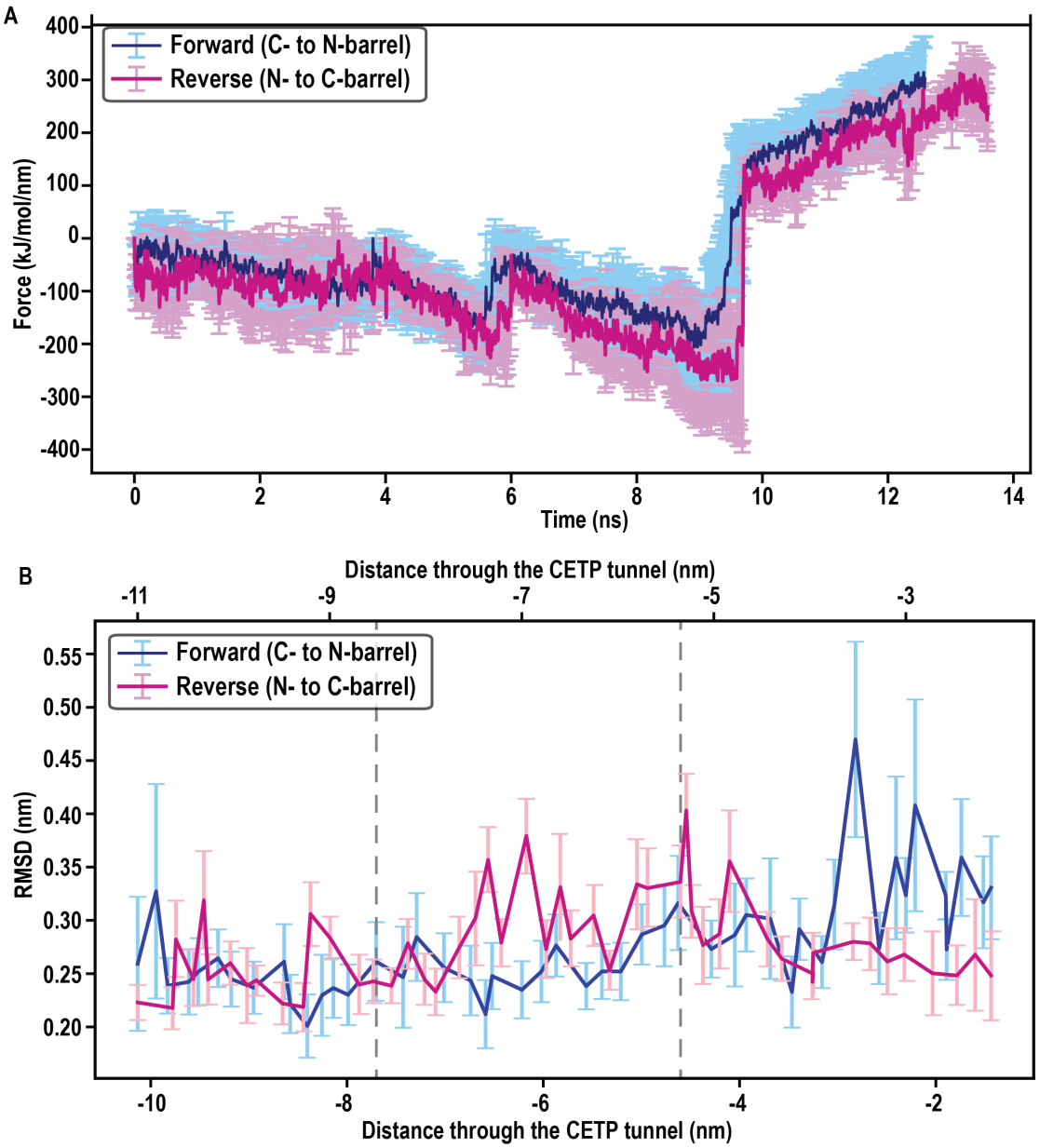



**Figure 6.4. Conformational changes when TG is pulled in forward or reverse direction. A.** Force profile of steered MD run in forward and reverse direction. The average force is shown, and the error bars depict the standard deviation from the mean. **B.** Root mean square deviation of backbone atoms of CETP in umbrella windows as TG moves through the tunnel.

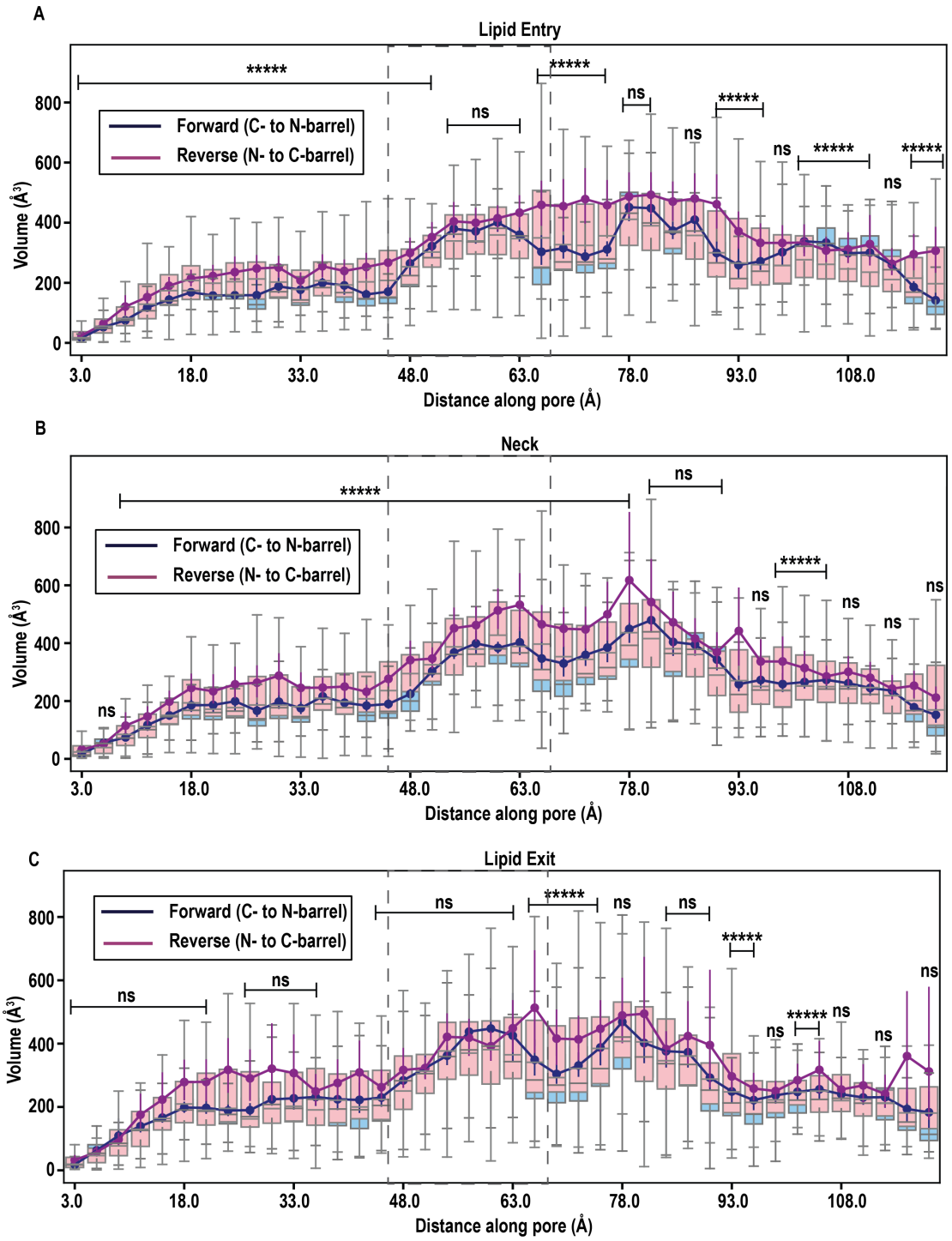



**Figure 6.5. CETP tunnel volume dynamics as lipid moves in forward or reverse direction.** Volume profile depicting changes in volume during **A.** lipid entry **B.** neck traversal and **C.** lipid exit are shown. The profiles from left to right depict the volume from C- to N-barrel. Simulation frames spaced 10ns from the umbrella simulation trajectories for each window were aligned such that the CETP tunnel aligned along the z-axis. These structures were used to calculate volume using CICLOP (Garg *et al.*, 2022). The box plot illustrates the distribution of volume. Solid lines depict the median value of the distribution for the distance along the tunnel axis. Mann Whitney U test was used to calculate the statistical significance of the difference between the two profiles (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

barrel. Conversely, movement in the opposite direction showed only minor fluctuations, which occurred specifically as the lipid passed through the neck region (~10 Å) (Fig. 6.4 B).

The force profile was observed to be similar in both scenarios. A negative force indicates that the system's kinetic or thermal energy is driving the molecule's motion, causing the spring connecting the dummy group and the molecule to compress. This favorable, kinetic-energy-driven movement, likely assisted by PLs (as discussed previously), was observed for both systems between 4-10 ns as the lipid entered the neck and the adjacent barrel. Following this, the force becomes positive as a push force is applied instead of a pull force (Fig. 6.4 A).

Next, we calculated the changes in tunnel volume as the lipid traversed CETP. Since the large pulling force in steered MD can induce artifacts due to protein rotation or translation, position restraints were applied to inhibit these movements. However, these restraints were removed during umbrella sampling (US) simulations, which are better for sampling natural protein conformations as the lipid moves through the tunnel. Therefore, US trajectories, rather than the raw steered MD trajectories, were used for calculating the volume profile. The results showed that during lipid movement, the volume of the C-barrel (3-45 Å) was significantly lower compared to the volume of the empty tunnel (Fig. 4.13 A). Conversely, the volumes of the neck (45-65 Å) and the N-barrel (65-123 Å) were substantially higher when the lipid was moving through them than in the absence of a lipid ( Fig. 4.13 A & Fig. 6.5).

Upon comparing the tunnel volume changes between forward and reverse simulations, it was observed that the tunnel volume of the N-barrel region (65-123 Å) is significantly higher during reverse simulations (purple, where the lipid enters the N-barrel) than in forward simulations (blue) ($p<0.00001$) (Fig. 6.5 A). This is consistent with the lipid entering the longer and narrower N-barrel end. Conversely, no significant volume change is observed in the C-barrel region (3-45 Å) during this entry phase (Fig. 6.5 A).

Similarly, as the lipid moves through the neck region (Fig. 6.5 B), the tunnel volume remains significantly higher in the reverse direction across the C-barrel region (3-45 Å), $p<0.00001$) and the N-barrel region (65-123Å), $p<0.00001$) with no significant difference in the neck itself.

Finally, as the lipid exits (Fig. 6.5 C), the C-barrel region (3-45 Å) exhibits a significantly higher volume; however, this change is not statistically significant.

### 6.3.3 Free energy of the movement of lipids in forward or reverse direction is similar

The overall free energy for the movement of triglyceride (TG) in either the forward or reverse direction is similar. To pinpoint the energy barriers for TG traversal, a Potential of Mean Force (PMF), or free energy profile, was generated for both the forward and reverse systems ( Fig. 5.15 A & Fig. 6.6), with convergence confirmed by monitoring deviation decrease across simulation blocks and ensuring sufficient umbrella histogram overlap.

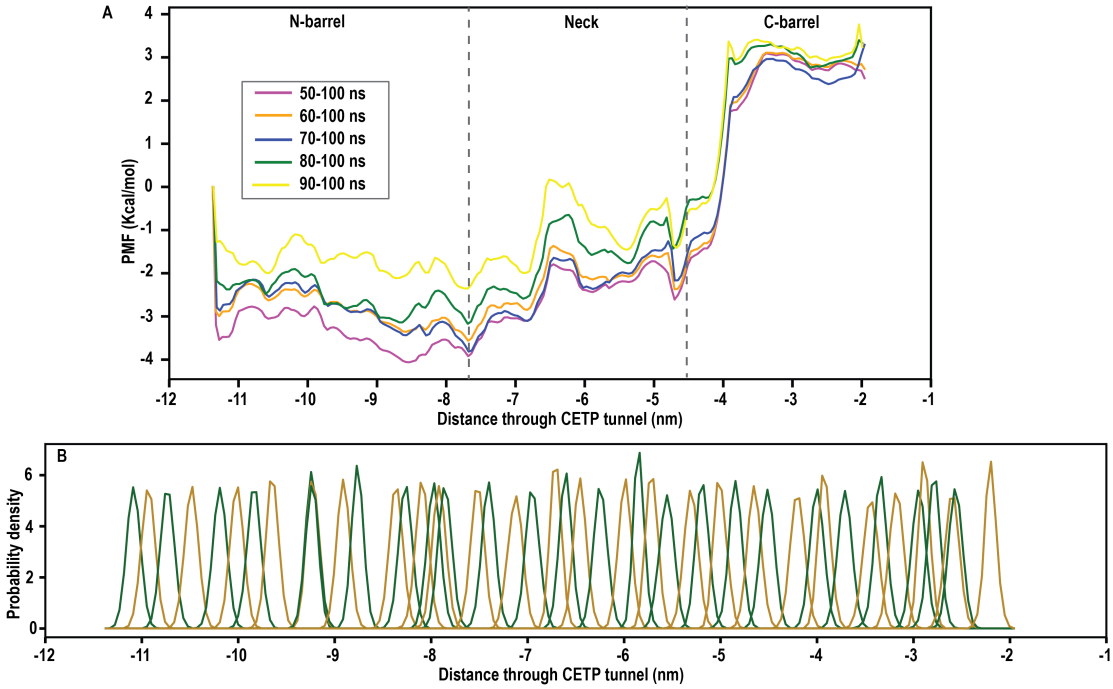



**Figure 6.6. Convergence of potential of mean force profile when TG is pulled from N- to C-barrel.** **A.** Each umbrella window was simulated for 100 ns, and PMF was calculated in blocks to ensure convergence. **B.** Overlap between adjacent umbrella histograms. In regions with low density, additional umbrella windows with stronger harmonic restraints were added.

Although distinct PMFs were obtained due to variability in the precise paths taken, they followed a similar trend: the entry into the CETP tunnel (either C- or N-barrel) and subsequent traversal to the neck region were found to be energetically favorable (-4.5kcal/mol), followed by a steady and steep increase as the TG exited through the

opposite end (+5kcal/mol) (Fig. 6.7). Specific features included a small initial barrier of +1kcal/mol upon entry into either barrel, and three small energy ”saddles” in the neck region corresponding to gate movement and the ”powerstroke” that accelerates TG into the adjacent barrel. Crucially, the average PMF for TG traversal from the N-barrel to the C-barrel was essentially an inversion of the profile obtained for traversal from the C-barrel to the N-barrel, with comparable barrier heights (Fig. 6.7). This indicates that the entry of TG into the tunnel from either side is thermodynamically favorable. The previously noted increase in residence time of TG when transitioning from the N-barrel is attributed solely to the additional hydrophobic contacts formed within the longer N-barrel, and not to higher energy barriers.

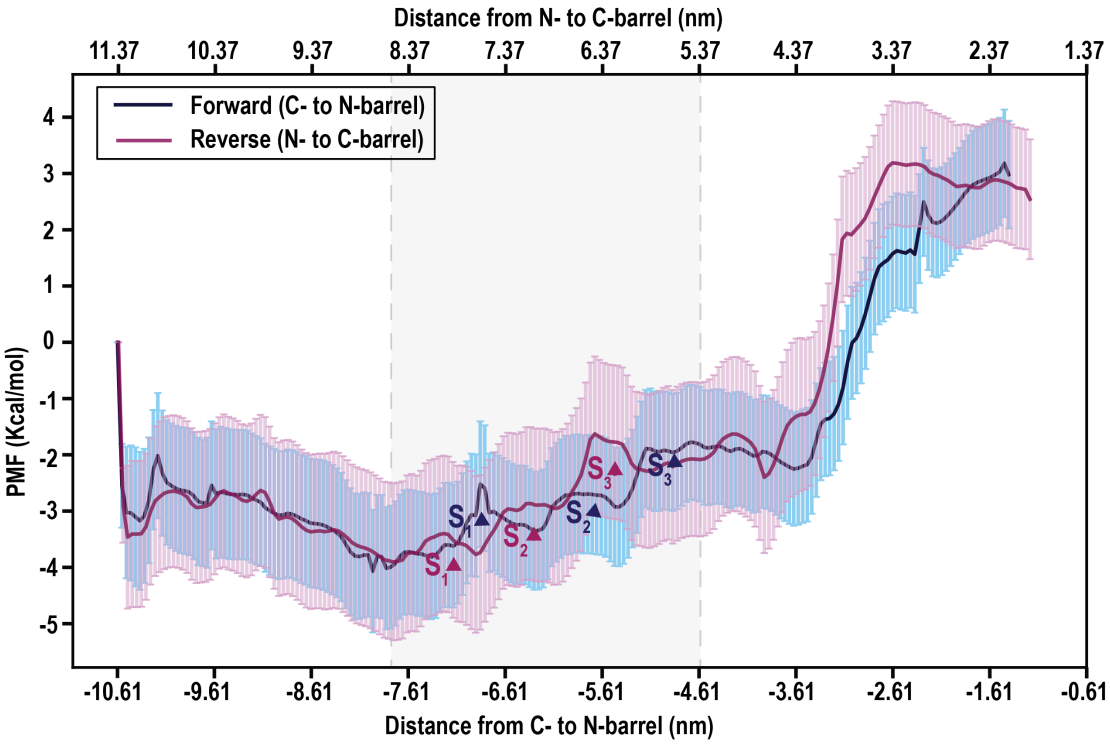



**Figure 6.7. Potential of mean force profile for the movement of TG through the CETP tunnel as it enters from C- or N-barrel**. Three saddles marked as S1, S2, & S3 for forward (blue) and reverse (red) pulls corresponding to the opening of the gates in the neck region and the powerstroke are highlighted with arrows. The average profile was calculated using 100 bootstraps, and the error bars depict the standard deviation from the mean.

### 6.3.4 CE and TG can be exchanged simultaneously through CETP

To determine if lipids move through the CETP tunnel one at a time or concurrently, we simultaneously placed a CE at the N-barrel mouth and a TG at the C-barrel mouth and

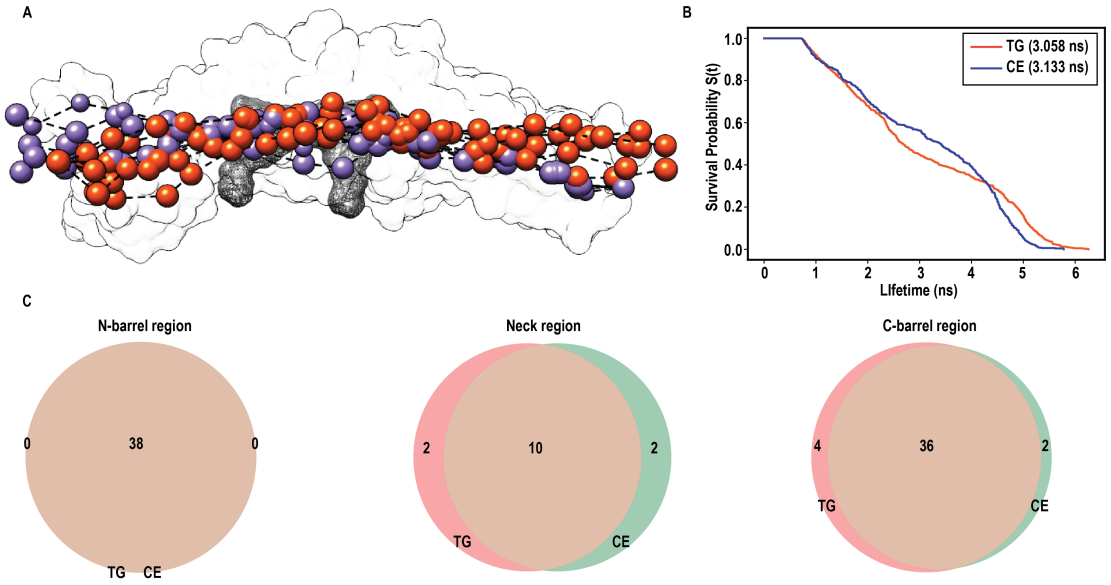



**Figure 6.8. Path taken by TG & CE as they enter through C- barrel & N-barrel respectively A.** Tracing the center of mass of the glycerol moiety of TG (orange circles) and C2 atoms of cholesteryl ester (purple circles) as they moves through the tunnel, in each replicate simulation of CETP. Multiple paths derived from three independent replicate simulations are depicted. **B.** Survival probability function of contacts between TG & CE with CETP tunnel residues. **C.** Venn diagram showing the common tunnel residues within a radius of 0.3–0.6 nm staying in contact with TG or CE for > 1 ns of simulation time in all replicate simulations.

pulled both into the tunnel using steered Molecular Dynamics (MD) (Fig. 6.2).

Replicate simulations showed that both lipids contacted the exact same residues (within 0.6 nm) as they entered the tunnel, traversed the neck, and exited the opposite barrel (Fig. 6.8 A). This suggests that CE and TG travel the same path. Although TG consists of three oleates (compared to the single oleic acid chain in CE) and was expected to form more hydrophobic contacts, the number of contacts formed by both lipids was surprisingly identical (Fig. 6.8 C & Table. 6.5).

Analysis of the residence time, based on the duration of contacts, showed that TG moving from the C- to the N-barrel was ~1.02 times faster than CE moving from the N- to the C-barrel (Fig. 6.8 B). This confirms the earlier observation that movement from the N-barrel to the C-barrel is slower, irrespective of the lipid type (TG or CE).

Furthermore, the tunnel volume profile exhibited a similar trend to single-lipid movement: the simultaneous movement of both lipids resulted in a decrease in the C-barrel volume, but an increase in the volume of the neck and N-barrel (Fig. 6.9). We must note, however, that this volume profile was calculated using raw steered MD trajecto-

ries, which may introduce simulation artifacts.

| Domain | CE or TG |
|---|---|
| N-barrel | L20, L23, N24, T27, A28, I31, F35, Y40, I64, L67, I69, I82, V84, I86, V89, V91, F93, I117, S119, I121, L123, I125, T127, L145, F147, L150, L152, K164, Q165, F167, T168, S172, L175, K176, L179, I183, I187, S191 |
| Neck | I11, I15, V198, A202, L206, I215, L228, F263, F265, L455, L457, L468 |
| C-barrel | F270, L273, A274, L283, L285, F292, I307, V311, V321, V323, P329, I331, V338, V340, S342, V344, V346, F348, F350, V359, Y361, F363, I367, V371, I389, P391, I405, F408, L409, M412, I413, V416, G417, I418, V421, M422, L425, F429 |

**Table 6.5.** Residues that make contact with TG or CE > 1ns of simulation time in all six replicates within 0.3-0.6 nm.

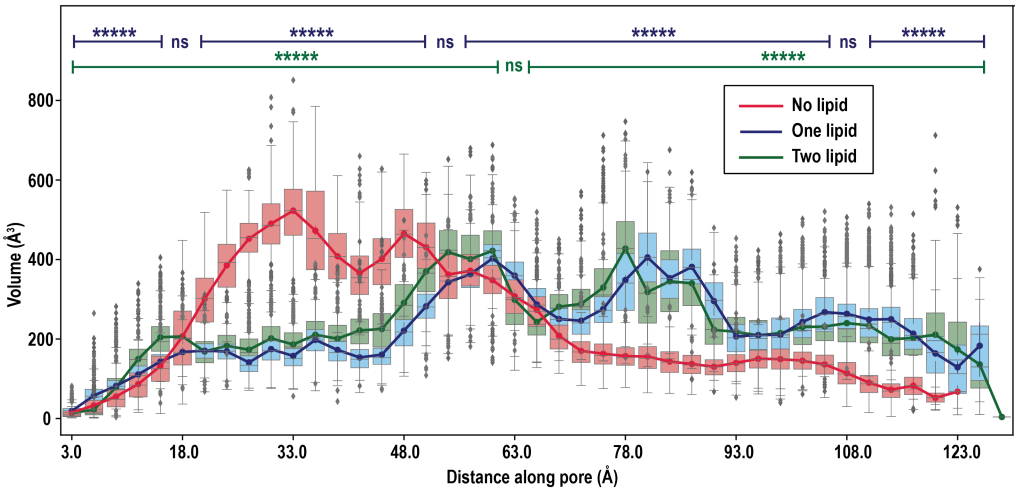



**Figure 6.9. CETP tunnel volume dynamics as both TG & CE move through the tunnel simultaneously.** The profiles from left to right depict the volume of the tunnel from C- to N-barrel. Simulation frames from independent steered MD trajectories where a single lipid or two lipids moved through the tunnel simultaneously. They were aligned such that the CETP tunnel aligned along the z-axis. These structures were used to calculate volume using CICLOP (Garg *et al.*, 2022). The box plot illustrates the distribution of volume. Solid lines depict the median value of the distribution for the distance along the tunnel axis. Mann Whitney U test was used to calculate the statistical significance of the difference between the no lipid and one lipid (blue) & no lipid and two lipids (green) (*p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).

### 6.3.5 Gliding of hydrophobic planes promotes lipid movement in the neck

By tracing the position of both lipids within the CETP tunnel (Fig. 6.10 A), we observed that the TG and CE pass through a common point, located in the neck region and occupied by phospholipids (PLs). Interestingly, both lipids reorient themselves as they approach this common PL region (indicated by the arrows in Fig. 6.10 A). The TG (red line) and CE (blue line) exhibit a reorientation, followed by an acceleration. Specifically,

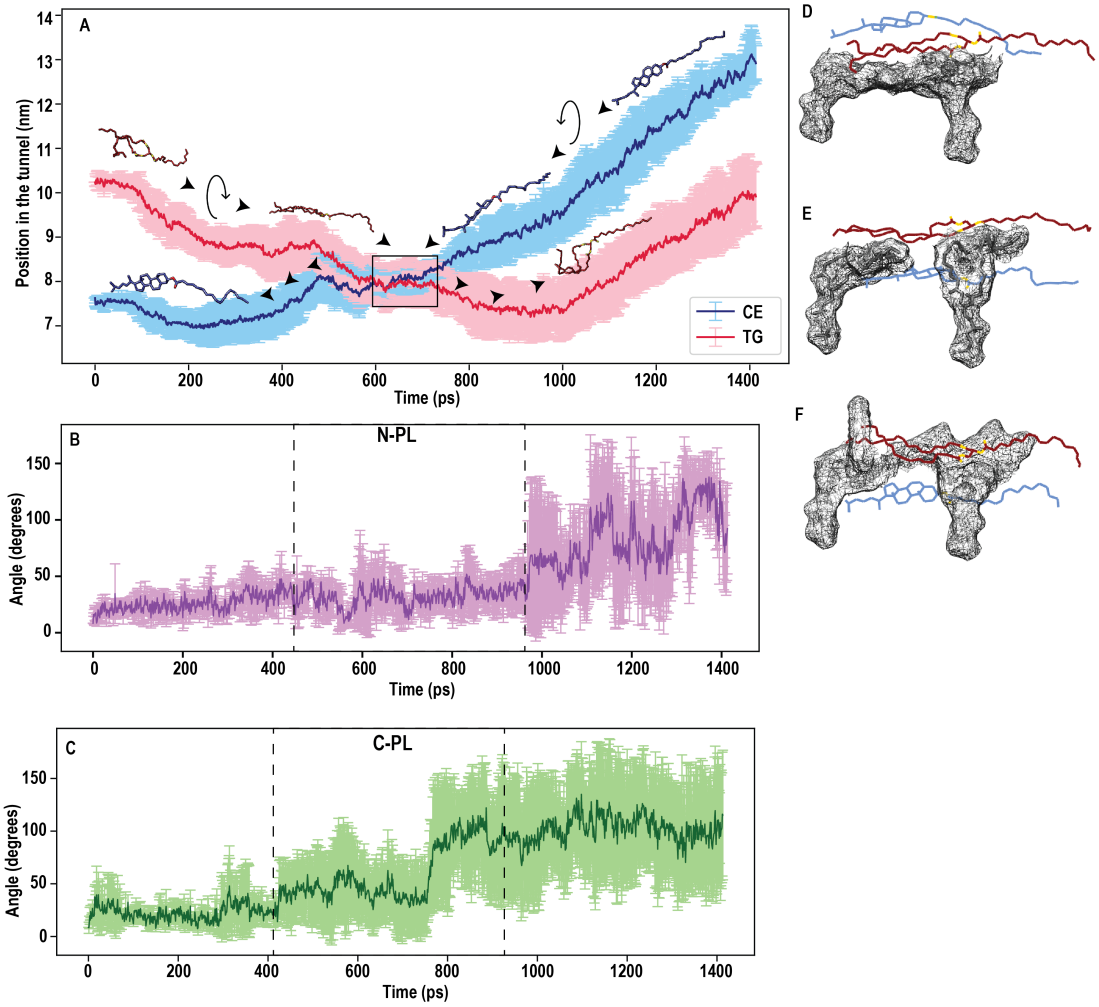



**Figure 6.10. Gliding of TG & CE past the PLs in neck of CETP. A.** Position of lipid in the tunnel throughout replicate simulations. The schematic depicts the orientation of the two lipids as they approach PLs in the neck. Angle between the acyl tails of **B.** N and **C.** C-PL in steered molecular dynamics (SMD) simulations of the movement of two lipids concurrently. Snapshots from replicate simulations **D-F.** depicting the orientation of TG & CE with respect to PLs as they glide past the neck.

both lipids spin to lie flat as they approach the PLs in the neck. Fig. 6.10 D-F illustrate this reorientation, showing the TG (red) and CE (blue) molecules lying nearly parallel to the PL surface mesh as they move through the tunnel. Moreover, the acyl tail ends of the PLs also form a stable hydrophobic plane with significantly reduced fluctuations as the lipids pass. This stability is quantitatively shown in Fig. 6.10 B & C, where the angle of the PL acyl tails (at the N-PL and C-PL positions, respectively) is maintained between 0 and 50 degrees during the lipid traversal. This stable, flat arrangement facilitates lipid movement through the gliding of these hydrophobic planes.

## 6.4 Discussion

Our findings fundamentally challenge the traditional model of CETP function by demonstrating that TG can successfully enter and traverse the entire CETP tunnel, irrespective of the barrel entry (N- or C-barrel). This refutes the initial hypothesis of strict barrel-specific recognition and confirms the non-specific lipoprotein binding observed by Lauer et al. (Lauer *et al.*, 2016). This non-specificity extends to lipid cargo, as the tunnel permits entry of any neutral lipid. Consequently, lipid entry into the tunnel is primarily governed by the surface availability of neutral lipids within the lipoprotein bilayer. This surface availability may be influenced by external factors, such as plasma lipid composition (dietary factors) or biophysical properties, including packing defects in the lipoprotein surface that increase local lipid accessibility.

Furthermore, detailed analysis of the TG's center of mass and contacting residues revealed that the lipid follows a single, conserved physical path within the tunnel regardless of its entry point. This suggests that the CETP tunnel serves as a passive guide for lipid movement, where generalized hydrophobic interactions effectively direct the cargo. Since the same set of residues contacts both TG and CE, there is no evidence for specific residues dedicated to a particular lipid type. This strongly suggests that overall tunnel integrity is the critical parameter for efficient lipid transport. Therefore, mutations to residues lining the cavity would not impair transfer by blocking a specific lipid interaction, but rather by destabilizing the protein structure as a whole, which subsequently hinders transport.

The free energy profiles further support this bidirectionality. The average PMF for

TG moving in the reverse direction (N- to C-barrel) is an approximate inversion of the forward profile (C- to N-barrel), and critically, the thermodynamic barrier heights are similar. Entry into the tunnel is thermodynamically favorable (~−4.5 kcal/mol), indicating that the thermal energy of the system can drive the initial steps for lipid entry. The observed longer residence time for TG traversing the N-to-C-barrel direction is not due to a higher energy barrier, but rather due to an increased number of hydrophobic contacts made in the longer N-barrel, suggesting a kinetic difference driven by physical interaction rather than a thermodynamic one.

However, the bi-directional movement of TG did demonstrate directional asymmetry in terms of structure, conformation, and kinetics. The N-barrel (65-123 Å) volume increases significantly upon lipid entry in the reverse direction (N-to-C-barrel entry), suggesting that this longer barrel end is more readily able to deform to accommodate the entering TG. Conversely, the C-barrel (3-45 Å) showed a negligible volume change upon entry in the forward direction. As the lipid moves deeper, the reverse direction consistently leads to a significantly higher tunnel volume across multiple domains (including the N-barrel and C-barrel), demonstrating that N-to C-barrel traversal requires a greater, more widespread structural dilation of the tunnel.

The observation that the C-barrel volume decreases (constricts) while the N-barrel volume increases (dilates) as the lipid passes strongly suggests a dynamic, directional "push-pull" mechanism that mimics peristaltic movement. It is likely that CETP uses concerted conformational changes to achieve propulsion: the region of the tunnel ahead of the lipid (N-barrel in this case) actively dilates to reduce the energetic barrier and create space, while the region behind the lipid (C-barrel) constricts or recovers its volume, effectively creating a structural "wave" that pushes the lipid forward. Therefore, CETP tunnel plasticity actively assists lipid transit by reshaping itself around the cargo.

Our simulations demonstrated that two lipids (TG and CE) can move concurrently through the CETP tunnel, following the same physical path and forming an identical number of hydrophobic contacts despite their size difference. This highlights the tunnel's geometric efficiency and non-specificity. The kinetic difference observed such as slower movement from N-barrel to C-barrel for both TG and CE, confirms that the directional bias is an inherent structural property (kinetics), independent of the lipid type.

Crucially, the ability of CETP to exchange TG and CE simultaneously directly sup-

ports the physiological relevance of the ternary complex model (HDL-CETP-VLDL/LDL). The free energy profile for single-lipid transfer showed a sharp, high exit barrier of +5 kcal/mol as the lipid leaves the tunnel. This significant energy cost for dissociation suggests that a second lipoprotein must be attached to the opposite barrel to act as a thermodynamic sink, lowering the exit barrier and making the complete transfer feasible. The simultaneous movement of two lipids, therefore, provides structural and kinetic evidence for the efficiency of the TG-CE counter-exchange. However, a limitation of this study is that it cannot definitively determine if the concurrent movement of the two lipids is thermodynamically or kinetically coupled. Mechanistically, the PL molecules residing in the neck region (45-65 Å) are critical; the lipids reorient themselves to lie flat as they approach this PL common point, while the PL acyl tails form a stable hydrophobic plane (angle close to 0°) with reduced fluctuations. This stable arrangement facilitates lipid movement through the gliding of hydrophobic planes (Chapter 5).

# CHAPTER 7

# Role of Acyl Chain Length in Lipid Traversal through CETP tunnel

## 7.1 Introduction

Plasma triglyceride (TG) levels are strongly and causally linked to atherosclerotic cardiovascular disease (ASCVD), reflecting an important component of systemic lipid metabolism (Libby, 2021; Nordestgaard and Varbo, 2014). TGs are transported primarily within triglyceride-rich lipoproteins (TRLs), namely chylomicrons (formed post-absorption of dietary fats in the small intestine) and Very Low-Density Lipoproteins (VLDL) (synthesized in the liver in between meals) (Ginsberg *et al.*, 2021). Consequently, plasma TG concentration serves as a proxy for the abundance of TRLs and, critically, the generation of their cholesterol-enriched remnants following lipolytic activity (Packard, 2023). Although TG itself is not a direct constituent of atherosclerotic plaque, the cholesterol-rich remnants of TRLs—specifically large VLDL particles and the small, dense LDL (sdLDL) generated from them—are strongly associated with atherosclerotic risk and metabolic syndrome (Lee *et al.*, 2022).

**Pivotal Role of CETP**

Within the plasma, TRLs are remodeled by CETP, which mediates the reciprocal exchange of TGs from TRLs for cholesteryl esters (CEs) from HDL. This activity is central to modulating the composition and subsequent catabolic turnover of all major lipoproteins. Modulating CETP's substrate preference can affect the plasma ratio of HDL, VLDL, and LDL, and influence hepatic expression of receptors like LDL-R (LDL-receptor), SCARB1 (scavenger receptor class B type 1), and CYP7A1 (cholesterol 7-alpha-hydroxylase) (Morton and Liu, 2020*b*).

Experimental evidence suggests that CETP activity is highly modifiable. Mutants, such as Q199A, have been shown to promote the non-reciprocal transfer of TG (TG transfer rate >> CE transfer rate), leading to increased production and rapid catabolism

of TG-rich HDL without significantly affecting VLDL levels (Morton and Izem, 2015). Conversely, certain anti-CETP antibodies can induce non-reciprocal CE transfer (CE transfer rate >> TG transfer rate), resulting in CE-rich VLDL, subsequent formation of sdLDL due to lipolytic activity, and no change in HDL (Morton and Izem, 2015). These findings highlight the potential for agents to modulate CETP's activity and, consequently, alter the atherogenic profile of plasma lipoproteins.

**Intrinsic Lipid Factors and CETP Transfer Dynamics**

Beyond extrinsic modulators, the intrinsic composition and structure of the neutral lipid cargo also critically influence CETP-mediated exchange. Previous studies have demonstrated that CE exchange activity increases with the acyl chain length of CE (Green and Pittman, 1991). Similarly, the transfer of medium-chain TGs (MCTs) from emulsion particles to LDL was found to be faster than that of long-chain TGs (LCTs), a phenomenon attributed to the greater surface solubility of MCTs (Richelle *et al.*, 1994). Furthermore, the presence of non-esterified fatty acids (NEFAs) can alter the molecular mobility of the lipoprotein phospholipid monolayer. Medium- and long-chain saturated NEFAs (C10-C16) significantly increase the fluidity of HDL and LDL, which, in turn, enhances the rate of CE exchange from HDL3 to LDL in a CETP-dependent manner (Foucher *et al.*, 1996; Lagrost and Barter, 1991*b*,*a*). This suggests that increasing lipid disorder and lipoprotein fluidity may enhance the availability of underlying neutral lipids for efficient exchange by CETP.

Given the substantial influence of dietary and metabolic processes on overall lipoprotein lipid composition, the surface availability of core lipids is likely a major determinant of exchange dynamics (Grundy and Denke, 1990; Stange *et al.*, 1975). Computational models comparing lipid droplets from high, normal, and low HDL-C groups observed that TGs were more prevalent at the surface of lipoproteins in the low HDL-C group compared to CE (Yetukuri *et al.*, 2010). Such changes in surface availability of TGs relative to CE could drastically affect the net effective exchange mediated by CETP.

Therefore, this chapter employs computational analysis to elucidate the influence of acyl chain length and TG conformation on the traversal dynamics of TG through the CETP tunnel. The study systematically compares the movement of a long-chain TG (LCTs; 18:1/18:1/18:1) with that of a medium-chain TG (MCTs; 10:0/10:0/10:0) and a short-chain TG (SCTs; 4:0/4:0/4:0). The effect of saturation is not explicitly investi-

gated, and the choice of oleic acid (18:1) over stearic acid (18:0) is based solely on the prevalence of unsaturated fatty acids in the human diet. Furthermore, the role of external lipoprotein composition on neutral lipid surface availability, thermal stability, and overall CETP exchange propensity is excluded from the scope of this investigation.

## 7.2 Methodology

**Steered MD simulations for transfer of TG varying in acyl chain length** The system was prepared as described previously, wherein one Long Chain TG (18:1/18:1/18:1), Medium Chain TG (10:0/10:0/10:0), or Short Chain TG (4:0/4:0/4:0) was manually placed at the opening of the C-barrel of CETP. Constant velocity simulations were performed as described previously in four phases, pulling TG from C-barrel to N-barrel.

**Simulations for sampling TG conformations** In order to sample the various TG conformations, a single TG, either long (18:1/18:1/18:1), medium (10:0/10:0/10:0), or short (4:0/4:0/4:0), was placed in a box full of water (TIP3P) (Price and Brooks, 2004) or cyclohexane. A polar and a non-polar solvent were deliberately chosen so as to sample more conformations, given the hydrophobic nature of TG. Simulations were performed with the GROMACS 2021.1 suite (Bekker *et al.*, 1993; Lindahl *et al.*, 2001). Each of the three starting conformations was placed in a cubic box large enough to contain the system with at least 1.0 nm of solvent on all sides. Periodic boundary conditions were used, and the long-range electrostatic interactions were treated with the particle mesh Ewald method (Darden *et al.*, 1993) using a grid spacing of 0.16 nm (for water simulations) and 0.12nm (for cyclohexane simulations) combined with a fourth-order B-spline interpolation to compute the potential and forces in-between grid points. The real-space cutoff distance was set to 1.4 nm, and the Van der Waals cutoff was also set to 1.4 nm. The bond lengths were fixed (Hess *et al.*, 1998), and a time step of 2 fs for numerical integration of the equations of motion was used. Coordinates were saved every 2 ps. The pressure coupling was performed using a Berendsen barostat (Berendsen *et al.*, 1984*a*), with a 1 bar reference pressure and a time constant of 2.0 ps, and a compressibility of 4.5e-5 bar, employing an isotropic scaling scheme. In the case of cyclohexane, a time constant of 0.5 ps with compressibility 1.21e-8 bar (Aicart *et al.*, 1981) using an isotropic

scaling scheme was used. All the starting structures were subjected to a minimization protocol of 50,000 steps using the steepest descent method, followed by equilibration with restraints on the lipid for 1 ns. Systems with cyclohexane were equilibrated for 10 ns to obtain an average density of ≈760 kg/$m^3$ (Kerimov and Apaev, 1972). Three independent trajectories, each 500 ns in duration at 310K, were carried out The simulation setup is summarized in Table 7.1.

| TG length | Medium | Number of lipid atoms | System size | Number of replicates | Duration |
|---|---|---|---|---|---|
| Long | Water | 167 | 5087 | 3 | 500 ns |
| (18:1/18:1/18:1) | Cyclohexane | 167 | 9167 | 3 | 500 ns |
| Medium | Water | 101 | 4964 | 3 | 500 ns |
| (10:0/10:0/10:0) | Cyclohexane | 101 | 9101 | 3 | 500 ns |
| Short | Water | 47 | 3803 | 3 | 500 ns |
| (4:0/4:0/4:0) | Cyclohexane | 47 | 9047 | 3 | 500 ns |

**Table 7.1.** Table summarizing the simulation setup for sampling the different TG conformations.

**Calculation of TG conformation** TGs were classified according to the classification proposed by Backle et al (Bacle *et al.*, 2017), with slight modification as follows. Briefly, three unit vectors between each glycerol carbon till the first carbon atom of the double bond in each fatty acyl chain were defined. In the case of long TG were defined between C1 to C12 atoms. Alternatively, for medium TG, the vectors were defined between C1 to C7. For short TG, vectors between the three glycerol carbons, as well as the last carbon of the fatty acyl chain, were considered [C1 to C4]. In the case of long and medium TG, the entire length of the fatty acyl chain was not considered to define the vector to account for the huge fluctuations in the terminal part of these chains (Fig. 7.1). Using the dot products between a combination of two vectors, the angle between each vector was calculated. Based on these three angles designated as $\theta 1$ (v1 . v2), $\theta 2$ (v2 . v3), $\theta 3$ (v3 . v1), each of the six conformations was classified as follows:

**Stacker** : [0°, 90°, 90°] or [90°, 0°, 90°] or [90°, 90°, 0°] with an allowed deviation of 10%.

**Trident** : [0°, 0°, 0°] with an allowed deviation of 10%.

**Fork** : [180°, 180°, 0°] with an allowed deviation of 10%.

**Chair**: [180°, 0°, 180°] or [0°, 180°, 180°] with an allowed deviation of 10%.

**T** : [180°, 90°, 90°] or [90°,180°,90°] or [90°, 90°, 180°] with an allowed deviation of 10%.

**Hand** : [90°, 90°, 90°] with an allowed deviation of 10%.

Each frame in the simulation was assigned a TG conformation based on the above scheme. Frames that did not fall in either of the six classes were classified as "Other". The sum of each conformation at the end of the simulation was calculated and used as a means to determine the prevalence of a conformation within the CETP tunnel.

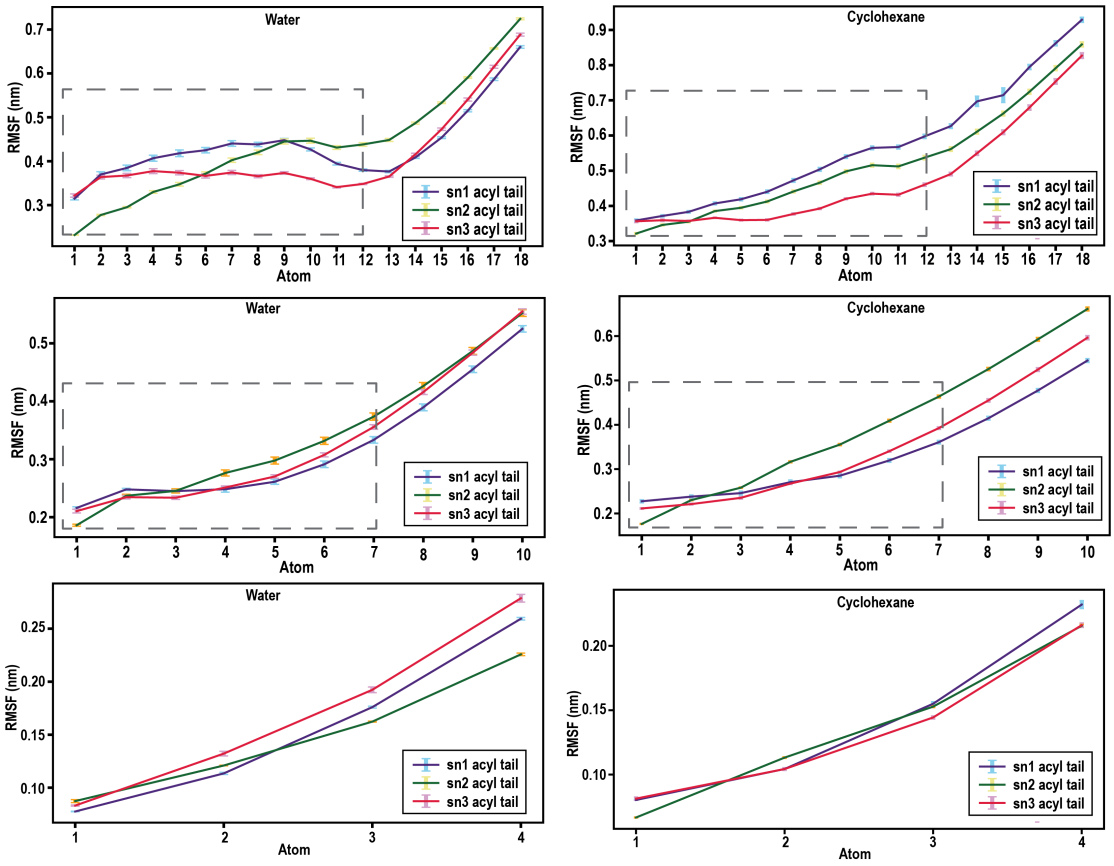



**Figure 7.1. Root mean square fluctuation (RMSF) of TG in water and cyclohexane.** The region in dotted line represents the region used for defining the vectors from the central carbon during assignment of conformations for both long and medium TG. In case of short TG, the terminal carbon of the acyl chain was considered as mentioned in Materials and Methods. The solid line depicts the mean, while the error bars represent absolute deviation from the mean.

**Steered MD simulations for traversal of various TG conformations** In case of various conformations of TG, a single TG in either Stacker, Trident, Fork, Chair, T, or Hand conformation was manually placed at the opening of the C-barrel of CETP. Constant velocity simulations were then performed as described previously. Considering the stochastic nature of MD simulation, SMD pulls were performed in several replicates. The simulation setup is summarized in Table 7.2.

| TG length | Conformation | Number of protein atoms | Number of lipid atoms | Number of water atoms | System size | Number of replicates |
|---|---|---|---|---|---|---|
| Long | Stacker | 7,440 | 276+167 | 222,480 | 230,378 | 37 |
| (18:1/18:1/18:1) | Trident | 7,440 | 276+167 | 206,967 | 214,865 | 30 |
| | Fork | 7,440 | 276+167 | 212,352 | 220,250 | 8 |
| | Chair | 7,440 | 276+167 | 212,115 | 220,013 | 20 |
| | T | 7,440 | 276+167 | 206,946 | 214,844 | 31 |
| | Hand | 7,440 | 276+167 | 222,591 | 230.489 | 100 |
| Medium | Stacker | 7,440 | 276+101 | 207,084 | 214,916 | 100 |
| (10:0/10:0/10:0) | Trident | 7,440 | 276+101 | 207,012 | 214,844 | 88 |
| | Fork | 7,440 | 276+101 | 206,937 | 214,769 | 15 |
| | Chair | 7,440 | 276+101 | 222,735 | 230,567 | 14 |
| | T | 7,440 | 276+101 | 206,994 | 214,826 | 56 |
| | Hand | 7,440 | 276+101 | 206,937 | 214,769 | 43 |
| Short | Stacker | 7,440 | 276+47 | 206,964 | 214,742 | 83 |
| (4:0/4:0/4:0) | T | 7,440 | 276+47 | 206,982 | 214,760 | 11 |
| | Hand | 7,440 | 276+47 | 222,687 | 230,465 | 39 |

**Table 7.2.** Table summarizing the steered MD simulation setup to understand the movement of TG varying in acyl chain length and conformation.

**Calculation of residence time of TG in the CETP tunnel** The distance between the group of atoms of TG and the group of atoms constituting every residue of protein was calculated within a radius of 6 Å for every frame in the simulation to generate a contact matrix. The lifetime of contact was calculated by dividing the total number of contacts made between a residue and TG or CE by the total simulation time.

A time auto-correlation function (ACF) $C_x(t)$ was used as a measure for the presence of a contact between TG and protein residue, and hence a measure of its residence time. Using the contact matrix generated above, all residues having a contact lifetime 0.05 were considered as the population of TG interacting residues at the start point $t_0$. The

lifetime of these bonds is monitored over time to build an average behavior.

$$C_x(t) = \left\langle \frac{h_{ij}(t_0)h_{ij}(t_0+t)}{h_{ij}(t_0)^2} \right\rangle \tag{7.1}$$

The continuous definition measures the time that a particular hydrogen bond remains continuously attached, whereas the intermittent definition allows a bond to break and then subsequently reform, allowing it to be counted again. The relevant lifetime, $\tau_x$ can then be found via integration of this function

$$\tau_x = \int_0^\infty C_x(t)dt \tag{7.2}$$

The observed behavior was then fit to a multi-exponential function using 2 exponents.

## 7.3 Results

### 7.3.1 Sampling and classifying the various TG conformations

To effectively sample and classify TG conformations, a single TG molecule was simulated in two different environments: a polar solvent (water) and a non-polar solvent (cyclohexane) (Fig. 7.2A). Studying an isolated TG was necessary because conformations in bulk lipid assemblies, which can form a neutral lipid core, may exhibit significantly different biophysical properties (Hall *et al.*, 2008).

The Radial Distribution Function [RDFs, *g(r)* ] was used to evaluate the organization of solvent molecules around the TG by examining the site-to-site pair correlations between the functional groups of the TG and the solvent molecules. Fig. 7.2B illustrates the $g(r)$ profile between the center of mass (COM) of TG and water molecules. The profile plateaued at 0.85 nm for long-chain TG and 1.25 nm for medium- and short-chain TG. The area under the curve, which represents the number of molecules in the first solvation layer, was found to decrease with increasing acyl chain length (Fig. 7.2B). When TG was immersed in cyclohexane, three broad peaks were consistently observed, irrespective of the acyl chain length, representing three distinct solvation layers (Fig. 7.2C).

The intensity of these peaks declined in a sequential manner. The presence of multiple solvation layers suggests hydrophobic forces of a higher magnitude that enable TG to keep cyclohexane molecules in closer proximity. Furthermore, the RDF profiles for the acyl chain and glycerol subunits were very similar to those of the COM for both water and cyclohexane (Fig. 7.2D-G).

A sharp, less intense peak was observed at $\sim 0.25$ nm in the $g(r)$ profile for all three TGs in water (Fig. 7.2B), but this peak was absent in cyclohexane (Fig. 7.2C). This peak was hypothesized to be a result of hydrogen-bond (H-bond) interactions, which are prevalent in water but not in cyclohexane. To test this hypothesis, we computed the $g(r)$ profile selectively for the H-atoms of water around the carbonyl oxygen of the TG acyl tails. A distinct peak was observed at $\sim 0.25$ nm. This peak overlapped across all three acyl chain lengths (Fig. 7.2 H & I). It was concluded that an H-bond attraction exists between the carbonyl oxygen of the acyl tails and water, pulling a small number of water molecules into close proximity with the TG.

Previous studies on TG within a lipid droplet core (bulk) identified six distinct conformations, classified by the orientation of the acyl tails around the central glycerol carbon (Fig 7.3A) (Bacle *et al.*, 2017). We adopted this framework to discretize our simulation frames into six states: Stacker, Trident, T, Hand, Fork, and Chair. Orientations that significantly deviated from these six were assigned to an 'Other' category.

The original classification protocol utilized by Bacle et al., which was based on the dot product of three orientation vectors, frequently resulted in the majority of conformations—especially for medium- and short-chain TG—being categorized as 'Other'. This protocol was modified by basing the classification on the angle between the vectors, rather than the dot product. Principal Component Analysis (PCA)-based clustering of the resulting angular distribution indicated that the clusters for the different conformations were very close to one another (Fig 7.3B). Despite this proximity, discrete clusters corresponding to each of the six target conformations could still be successfully identified.

Using this refined classification scheme, we assigned a specific conformation to every frame in the simulation. For long- and medium-chain TG, all six target conformations (Stacker, Trident, T, Hand, Fork, and Chair) were successfully sampled from both the water and cyclohexane simulations. However, for short-chain TG, only the Hand, Stacker, and T conformations could be sampled. These specific conformational topolo-

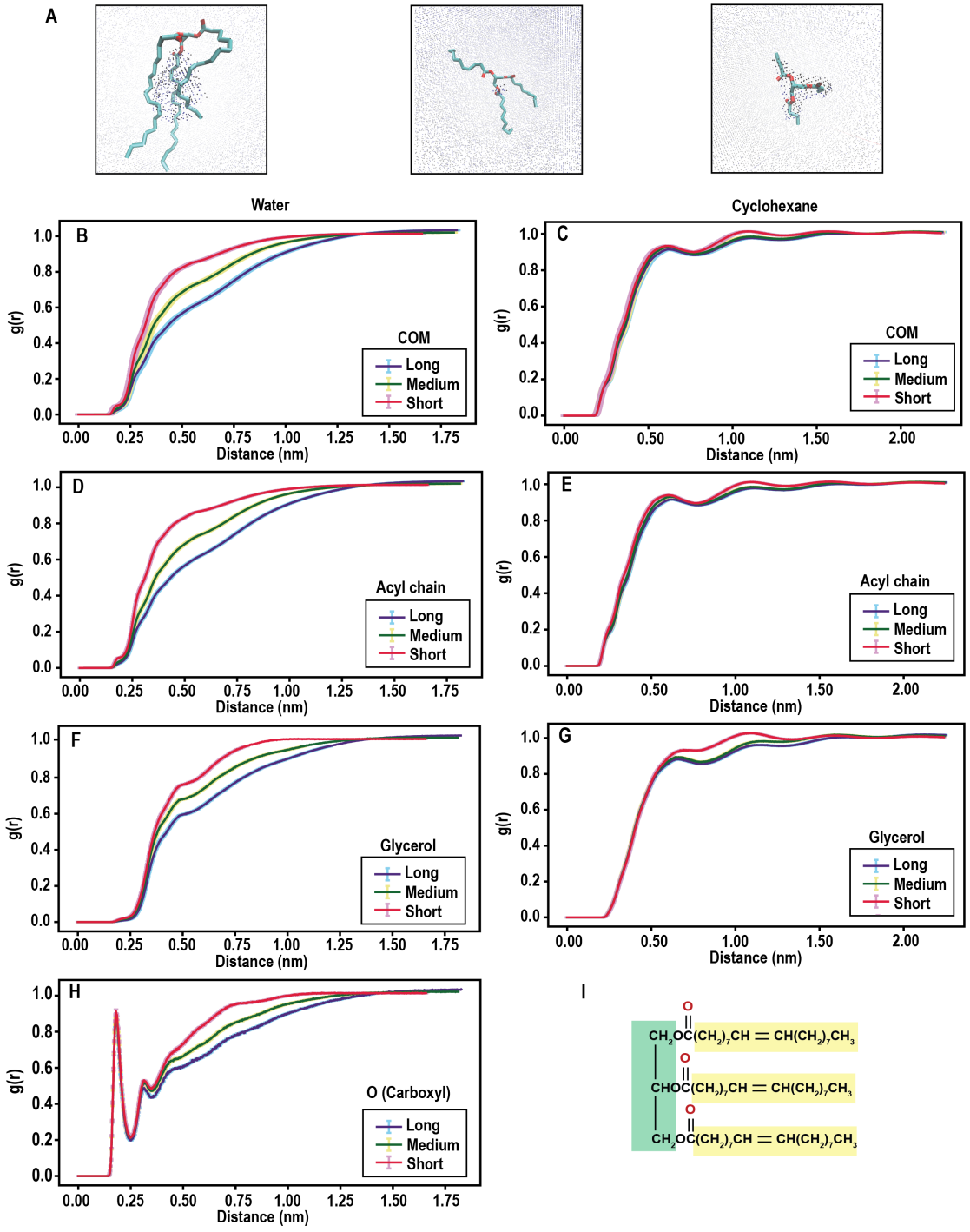



**Figure 7.2. Solvation dynamics of a single TG A.** System setups depicting a long, medium, and short chain TG in a water box. Radial distribution function (RDF) profile of **B.** water and **C.** cyclohexane around the center of mass of the TG, varying in acyl chain length. RDF of **D.** water and **E.** cyclohexane around the acyl chain of TG. RDF of **F.** water and **G.** cyclohexane around glycerol group of TG. **H.** RDF of water around the carboxyl atom of TG. **I.** Schematic of a TG molecule with its various functional groups highlighted. Solid lines represent the mean values, while error bars depict the absolute deviation from the mean calculated across three replicates of the simulation.

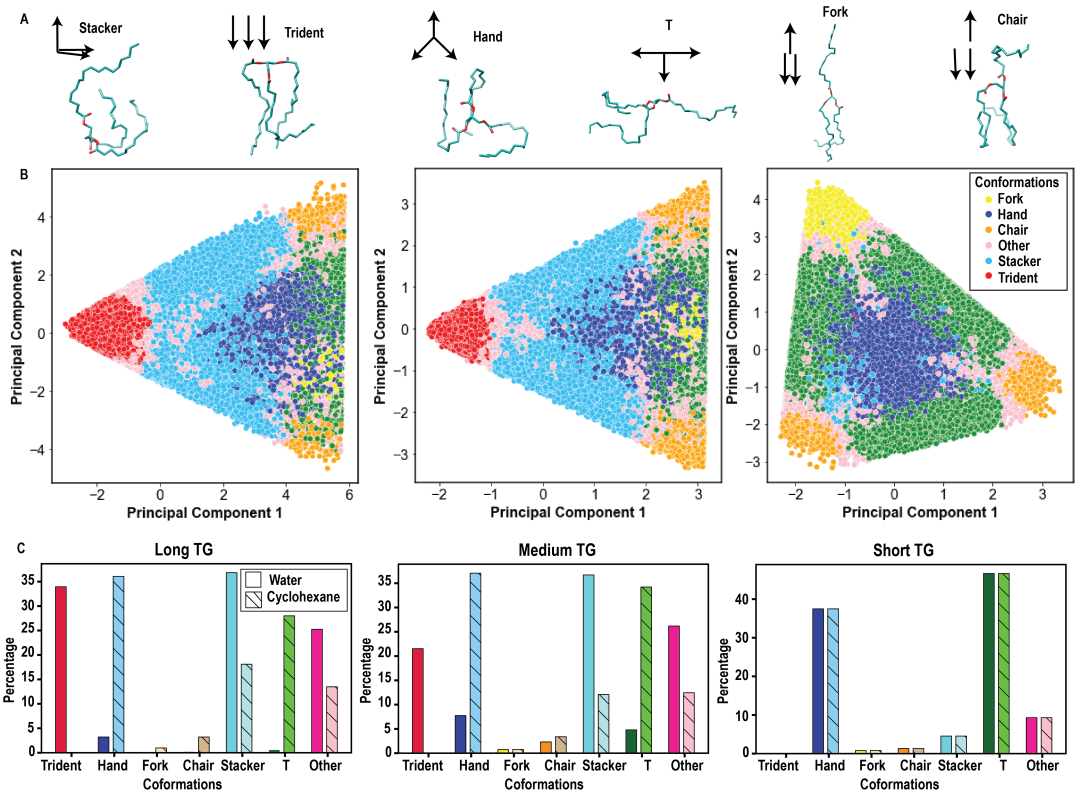



**Figure 7.3. Distribution of the different conformations of a TG. A.** The six conformations that a TG molecule exists in. Panel shows a schematic representation along with a representative image for each of the six conformations (Adapted from Bacle et al. (Bacle *et al.*, 2017). **B.** Clustering of the six conformations based on principal components. **C.** Percentage of each of the six conformations of a TG sampled within simulation when solvated in water or cyclohexane

gies were subsequently used as the starting structures for further analysis (Fig. 7.3C).

### 7.3.2 TG entry within the tunnel is dependent on its conformation

To investigate the effect of conformation on TG entry, the six distinct TG conformations (Hand, T, Stacker, Trident, Fork, & Chair), previously sampled for each acyl chain length, were positioned at the C-barrel mouth of CETP. Each was then pulled through the tunnel across four phases, as previously described. Simulations were conducted until at least six independent successful entry events were recorded for each starting conformation. To manage computational resources, an upper limit of 100 trials per setup was also imposed. This success-based stopping rule naturally resulted in unequal replicate counts; conformations that entered readily required fewer trials, while those that were more resistant to entry required significantly more, with some potentially reaching the 100-trial cap before achieving six successes. It was observed that the different conformations of TG had varying propensities to enter the CETP tunnel. These propensities only represent the ability of a conformation to enter the tunnel rather than absolute entry probabilities.

Long TG: Fork >> Chair > T > Trident > Stacker > Hand.

Medium TG: Fork & Chair > Hand > T > Trident > Stacker

Short TG: T > Hand >> Stacker (could not enter) (Table 7.3)

Specifically, short-chain TG showed a significantly reduced overall propensity to enter the tunnel compared to long and medium TGs, with the Stacker conformation being unable to enter at all (Table 7.3). Despite these differences in entry propensity, once successfully inside the tunnel, every TG conformation, regardless of its acyl chain length, was able to traverse the entire CETP tunnel successfully. Notably, the long TG-Fork conformation exhibited the highest propensity to enter the tunnel (0.85), followed by the medium TG-Fork and Chair conformations (both 0.46). All other conformations demonstrated a lower entry propensity than these top performers (Table 7.3).

| Chain Length | Conformation | Phase1 | Phase2 | Phase3 | Phase4 |
|---|---|---|---|---|---|
| Long TG (18:1/18:1/18:1) | Hand | 3/100 = 0.03 | 3/3 = 1 | 3/3 = 1 | 3/3 = 1 |
| | T | 6/23 = 0.26 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Stacker | 6/45 = 0.13 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Trident | 6/31 = 0.19 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Fork | 6/7 = 0.85 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Chair | 6/19 = 0.31 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| Medium TG (10:0/10:0/10:0) | Hand | 6/39 = 0.15 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | T | 6/47 = 0.12 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Stacker | 2/100 = 0.02 | 2/2 = 1 | 2/2 = 1 | 2/2 = 1 |
| | Trident | 6/57 = 0.10 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Fork | 6/13 = 0.46 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Chair | 6/13 = 0.46 | 6/6=1 | 6/6=1 | 6/6=1 |
| Short TG (4:0/4:0/4:0) | Hand | 6/34 = 0.17 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | T | 6/27 = 0.22 | 6/6 = 1 | 6/6 = 1 | 6/6 = 1 |
| | Stacker | 0/100 = 0 | - | - | - |

**Table 7.3.** Table summarizing the propensity of various TG conformations to enter and traverse the tunnel.

### 7.3.3 Rate of TG transfer is dependent on acyl chain length

To determine the impact of acyl chain length on TG traversal through CETP, we examined three representative chain lengths: long-chain TG (54 C), medium-chain TG (30 C), and short-chain TG (12 C). Three similar simulation systems were established, with each TG (long, medium, or short) placed at the C-terminal opening and pulled through the tunnel, as detailed in the methodology. The number of contacts per TG with CETP residues within 0.6 nm was monitored, and contact probabilities were calculated by normalizing each contact's duration to the total simulation time.

It was observed that long-chain TG exhibited both a higher total number of contacts and larger contact probabilities compared to medium-chain and short-chain TG. This suggests a direct correlation between the acyl chain length and both the number and duration of contacts formed (Fig. 7.4 A). The distribution of contact probabilities also varied by chain length. The distributions for long-chain TG were left-skewed (Fig. 7.4 B ). In contrast, the distribution for medium & short-chain TG was normally distributed (Fig. 7.4 C& D). Interestingly, the majority of contacts, regardless of chain length, shared a similar contact probability of 0.15 (Fig. 7.4 B-D). However, a small subset of contacts

made by long-chain TG reached a significantly high contact probability of 0.4 (Fig. 7.4 B). Despite the differences in length and the variability in the exact paths taken, all three TGs interacted with a similar cluster of residues (Fig. 7.4 A).

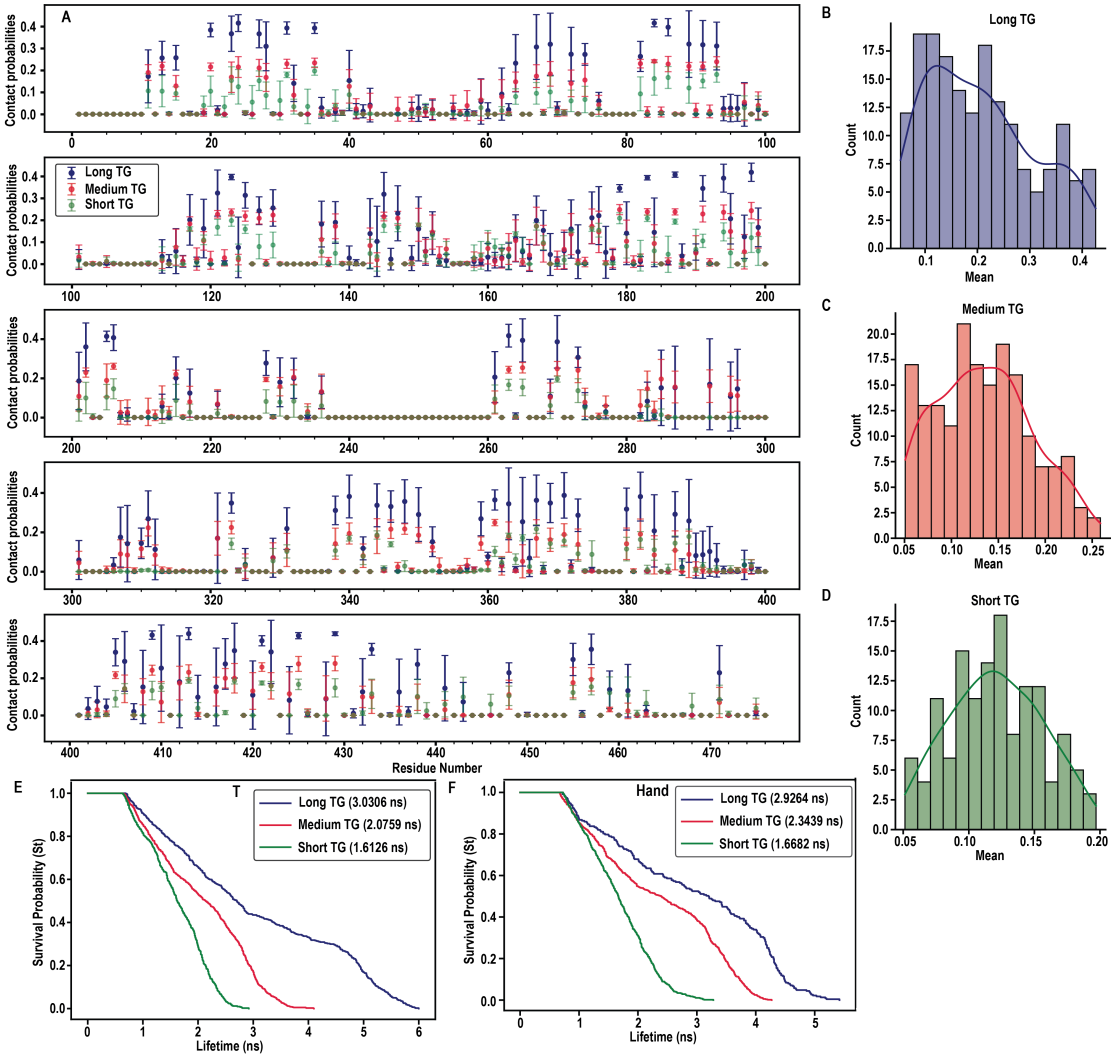



**Figure 7.4. Contacts made by TG varying in acyl chain length with CETP tunnel residues.** **A.** Probability of contact between CETP tunnel residues and TG (long/medium/short) within 0.6 Å. Distribution of contact probabilities of **B.** long **C.** medium **D.** short TG moving through CETP tunnel. Survival probability function of contacts between TG varying in acyl chain length and CETP tunnel residues for **E.** T and **F.** Hand conformations

To further quantify the effects of contact number and duration, we calculated the residence time for each of the three TGs inside the tunnel. This residence time was represented by the autocorrelation function of the time series data for each TG-protein residue contact pair throughout the simulation. A steeper or faster decay of the function indicates a short residence time, reflecting rapid decorrelation of the time series data. A slower decay indicates a longer correlation and a long residence time. Given only T and

Hand conformations of short-chain TG entered the CETP tunnel (Table 7.3), we have compared the residence time of these two conformations. Our results clearly showed that short-chain TG had the shortest decay time, with no contacts lasting longer than 2.5-2.7 ns (Fig. 7.4 E& F). Medium-chain TG followed, with no contacts exceeding 3.7-4.0 ns (Fig. 7.4 E & F). Long-chain TG had the longest decay time, decaying over approximately ≈ 5.5-6.0 ns (Fig. 7.4 E & F).

The mean residence time—defined as the time at which 50% of the contacts break—was significantly greater for long-chain TG (3.03 ns), intermediate for medium-chain TG (2.07 ns), and shortest for short-chain TG (1.61 ns). This confirms that longer acyl chains result in longer, more stable interactions within the CETP tunnel, thereby slowing down the transfer rate.

### 7.3.4 Conformational flexibility of lipid allows movement inside the tunnel

To minimize the influence of the pulling force on the TG conformation during traversal, different conformations were steered through the tunnel using varied pulling points. Specifically, Fork, Chair, T, and Hand were pulled via their acyl tails, while Stacker and Trident were pulled via their glycerol group. Despite this careful approach, it was observed that as the TG conformations traversed the different domains of the CETP tunnel, they fluctuated according to the spatial and volume constraints within the tunnel. The specific conformational transitions varied based on acyl chain length: For long TG; Hand & Chair maintained Chair conformation, while Stacker & Trident maintained the Trident conformation, and both Fork and T conformations maintained the Fork conformation for the majority of the simulation as they moved through the tunnel (Fig. 7.5 A).

In the case of medium TG, Chair, Fork, and T traversed as Fork, while Hand converted to Chair to move through the tunnel. Interestingly, Stacker converted to Trident, while Trident converted to Stacker, allowing them to effectively move through the tunnel (Fig. 7.5 B). Even short-chain TG, despite its minimal length, fluctuated between the T and Chair conformations as it moved through the tunnel (Fig. 7.5 C).

Furthermore, the residence time for each TG conformation within the tunnel was

distinct and depended on the number of hydrophobic interactions it formed. The mean residence times varied in the following orders: For long TG, Fork > T > Hand > Chair > Stacker = Trident (Fig. 7.5 D). While for medium TG, Fork = Chair = Hand > T > Stacker = Trident (Fig. 7.5 E).

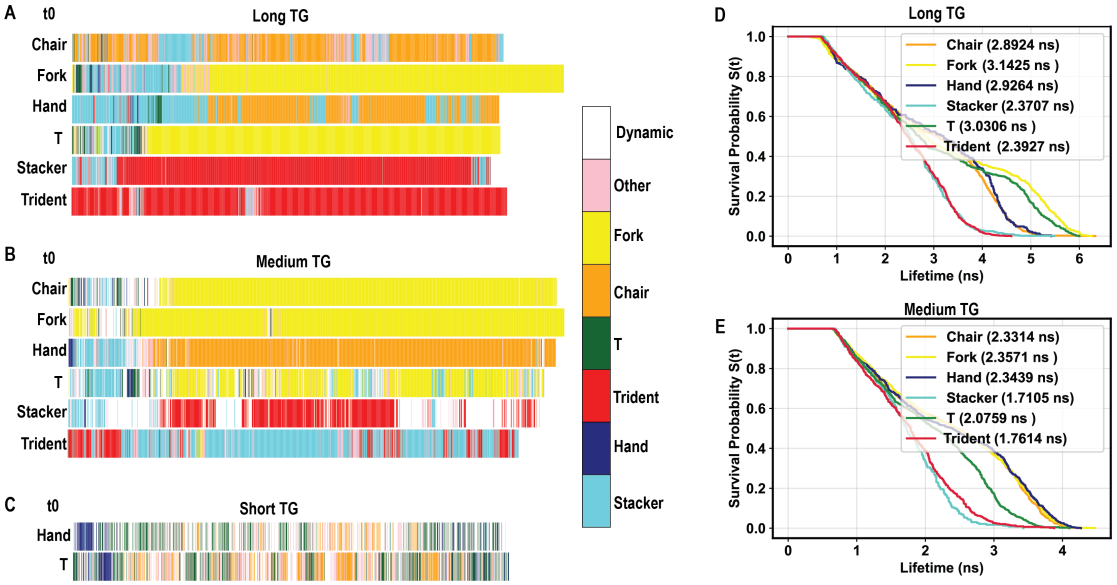



**Figure 7.5. Conformational flexibility of TG.** Fluctuations in **A.** long **B.** medium and **C.** short chain TG conformations as they move through the CETP tunnel. Dynamic conformations are assigned to those frames in which no consensus conformation is obtained across 50% of replicates. Survival probability function of contacts between the six conformations of **D.** long and **E.** medium TG and CETP tunnel residues.

## 7.4 Discussion

This chapter establishes a clear, multifaceted link between the acyl chain length of TG, its preferred conformation, and the efficiency of its transfer through the CETP tunnel. The length of the acyl chains is a primary determinant of the TG's kinetic profile inside CETP, as evidenced by a progressive change in contact dynamics: Long-chain TG (54 C) formed a higher number of contacts with CETP residues and exhibited larger contact probabilities compared to medium- (30C) and short-chain TGs (12C). This enhanced interaction directly correlated with the longest decay time for the autocorrelation function ($\approx$ 6 ns) and the highest mean residence time (3.03 ns) for long-chain TGs. This demonstrates that longer chains promote stronger, more persistent hydrophobic interactions within the tunnel, which may lead to a slower exchange rate, an important factor in lipoprotein remodeling.

Further, the ability of a TG to enter the CETP tunnel is heavily dependent on its instantaneous conformation, which is itself constrained by chain length. Specifically, long-chain TG showed the highest entry propensity in the Fork conformation (0.85), suggesting that an open, splayed arrangement of the acyl tails is optimal for initial threading into the C-barrel mouth. Medium-chain TG similarly favored Fork and Chair conformations for entry (0.46). In contrast, short-chain TG displayed a significantly reduced overall propensity to enter, highlighting that the physical size and reduced hydrophobicity profoundly impact the initial steering process inside the CETP tunnel, which is primarily guided by hydrophobic interactions. Interestingly, once inside the constrained tunnel environment, TGs demonstrated significant conformational flexibility, undergoing spontaneous conformational transitions (e.g., Hand $\rightarrow$ Chair, Trident $\rightleftharpoons$ Stacker) to minimize energetic strain and maximize fit as they navigated the varying spatial constraints of the tunnel. This flexibility is essential for successful, end-to-end traversal, regardless of the initial entry conformation. The mean residence time for TGs inside the tunnel was also confirmation-dependent, with Fork and T generally having longer residence times than Stacker and Trident.

Ultimately, the finding that long-chain TGs move through the CETP tunnel slowly is highly relevant to plasma lipoprotein remodeling. Since CETP typically exchanges cholesteryl esters (CE) from HDL for TG from VLDL and LDL, a slow TG exchange rate means that plasma CETP may operate less efficiently as a dual exchanger and more as a CE exchanger, where CE transfer activity exceeds TG transfer activity (Morton and Izem, 2015). This imbalance results in the CE enrichment of VLDL/LDL, which are then rapidly acted upon by lipases to form smaller, denser LDL particles (sdLDL). The formation of sdLDL is more atherogenic than large, buoyant LDL because it can more easily penetrate the arterial wall and is less readily cleared by the LDL receptor, thereby contributing directly to the development of Cardiovascular Disease (CVD). In support of this connection, elevated plasma TG levels, particularly those associated with Triglyceride-Rich Lipoproteins (TRLs) which are an independent risk factor for CVD (Libby, 2021; Nordestgaard and Varbo, 2014), are comprised predominantly of long-chain TGs (chain length > 50C) in CVD patients, as demonstrated by lipidomics in large clinical studies (Naushin, 2024). Furthermore, long-chain TGs, formed by long-chain fatty acids (LCFAs > 13C), are the prevalent form of fat largely derived from common dietary sources such as edible oils (soybean, corn, olive), animal fats (beef, pork), butter,

and processed foods (Abedi and Sahari, 2014). Therefore, high consumption of these LCFAs, often combined with a poor lifestyle, leads to hypertriglyceridemia and high circulating TRLs containing long-chain TGs. This subsequently drives adverse lipoprotein remodeling, mediated in part by the slow kinetics of long-chain TG transfer through CETP, thus escalating CVD risk. In light of these results, the consumption of medium-chain or short-chain fatty acids may prove to be more beneficial or neutral (Perna and Hewlings, 2023). Intriguingly, while LCFAs dominate, consumption of very long chain fatty acids ($>$ 20 C) has been shown to reduce the risk of heart failure, although the underlying biophysical mechanisms require further study (Lemaitre and King, 2022).

# CHAPTER 8

# Conclusion

This thesis culminates in the elucidation of the mechanism by which CETP facilitates lipid transfer, highlighting the roles of integral structural components and thermodynamic factors in the process. Our findings fundamentally challenged prior models, establishing that lipid traverses the entire tunnel as it is exchanged. It also provides compelling evidence in favor of the ternary mechanism of CETP-mediated lipid transfer, where CETP simultaneously bridges two lipoproteins. Furthermore, CETP tunnel characteristics, particularly its hydrophobicity, emerge as a crucial factor that gives rise to significant structural features for passive, directed lipid diffusion through the tunnel. The study also identifies PL-plugs as important co-factors that strongly influence CETP function, with their ability to modulate tunnel hydrophobicity, proving themselves as indispensable for efficient lipid exchange

## 8.1 Development and Validation of CICLOP for Internal Cavity Characterization

To overcome the experimental and computational limitations associated with studying internal protein cavities, a novel algorithm, CICLOP, was developed (Chapter 4). This tool offers an ingenious, accurate, and fully automated method for detecting and characterizing protein internal cavities, including pores, channels, tunnels, and voids at the atomistic resolution.

The key advantages of this method include:

- **Atomistic Resolution and Accuracy:** CICLOP prioritizes the accurate identification of residues lining the cavity, which are critical for determining the cavity's physico-chemical properties and functional role.
- **Automated and Robustness:** The algorithm requires no user intervention (e.g., seed residues) and is rigorously tested to be insensitive to protein orientation and free from issues related to grid spacing sensitivity, a common pitfall of other grid-based methods. It continues running until all empty voxels are identified, ensuring that no cavity is missed.

- **Superior Boundary Handling:** It avoids pitfalls related to boundary issues, allowing for a clear distinction between the cavity interior and the protein boundary, making it superior to methods like PoreWalker, MSpocket, and Mole.

However, a primary limitation of CICLOP is its focus on internal cavities only; it performs suboptimally when identifying surface cavities, such as pockets or clefts. Furthermore, while the method is used for dynamic characterization throughout this thesis by analyzing single frames from Molecular Dynamics (MD) trajectories, it does not currently take an entire trajectory as input.

## 8.2 Characterizing the CETP tunnel: A Dynamic Structural Component

On analyzing CETP using CICLOP, we found that the protein contains a single, continuous hydrophobic cavity that runs its entire length (Chapter 4). This finding contradicted earlier hypotheses, which had suggested that CETP's tunnel was formed by the fusion of several small (diameter >5Å), isolated cavities upon protein conformational changes. While CICLOP confirmed that the cavity's volume changes with protein conformation, it remained continuously connected. This important observation demonstrates that the CETP cavity is not a passive channel but rather a dynamic structural component essential for lipid transfer.

Our findings also demonstrated that the tunnel volume is not static but dynamically reshapes to facilitate lipid movement (Chapter 6). Specifically, as a lipid (TG or CE) traverses the cavity, the barrel segment ahead of the cargo dilates (volume increases), while the segment behind constricts (volume decreases). This concerted volume change strongly suggests a directional "push-pull" mechanism that mimics peristaltic movement, actively generating a structural wave to propel the lipid forward and assist transit. The mechanism is characterized by non-specificity at the termini (Chapter 6).

Furthermore, CETP permits the entry of both TG and CE through either the N- or C-barrel (Chapter 6). Lipid entry is primarily governed by the surface availability of neutral lipids within the lipoprotein environment and the acyl chain length and conformation of lipid (Chapter 7). Once inside, the lipid follows a single, conserved physical path, ir-

respective of the lipid type or entry barrel, and contacts the same set of residues. This evidence suggests that the overall tunnel integrity is the critical determinant of transfer efficiency, not specific, residue-level lipid recognition. Therefore, single-residue mutations are more likely to impair function by destabilizing the entire protein structure and tunnel architecture than by blocking a specific lipid interaction site.

## 8.3 The Indispensable Cog in the Lipid Exchange Machinery: PL plugs

The major finding of this work is the identification of PL plugs as structural cofactors that are essential for regulating CETP's mechanical and functional behaviors (Chapter 5). The presence of PL plugs significantly enhances the tunnel's overall hydrophobicity by effectively sealing the two central orifices. This increased hydrophobicity is critical, as it results in a more thermodynamically favorable free energy of entry ($\sim$−4.5 kcal/mol) for TG into the CETP tunnel. Furthermore, PLs increase the tunnel volume, particularly in the neck region and the adjacent C-barrel, which is essential for accommodating the larger lipid cargo.

During lipid movement, the PL acyl tails, along with three key Phenylalanine residues (F263, F265, F270) that form flaps in the neck region, form a coordinated system to prevent kinetic trapping (Chapter 5). The Phe gates open concertedly, serving as a directional valve to prevent backflow (regurgitation) of the lipid, which is essential for a diffusion-driven transport process. Further, interactions between the lipid and the PL acyl tails in the neck region create a highly ordered hydrophobic plane. This interaction facilitates the lipid's reorientation, allowing it to glide forward seamlessly into the next compartment, thereby accelerating movement and preventing it from becoming trapped in a potential energy well at the neck.

PLs also induce synchronized motion between the N- and C-terminal β-barrels (Chapter 5). This domain synchrony, characterized by the "squeezing and stretching" motions, is proposed to be the prerequisite for lipoprotein docking and the formation of the ternary complex (Lipoprotein-CETP-Lipoprotein). The displacement of these crucial PLs by inhibitors, such as anacetrapib and torcetrapib, likely disrupts this barrel synchrony,

providing another mechanistic link for how these drugs impede lipid transfer.

## 8.4 Mechanistic and Pathophysiological Insights into CETP-Mediated Lipid Exchange

This work also addressed the kinetic and thermodynamic basis of lipid transfer, providing insights directly relevant to the pathophysiology of cardiovascular disease (CVD).

Firstly, the free energy profiles showed that TG movement is thermodynamically bidirectional, with similar barrier heights for C-to-N and N-to-C traversals (Chapter 6). However, the movement exhibits directional asymmetry in kinetics, with traversal from the N-barrel to the C-barrel being intrinsically slower for both TG and CE. More importantly, the discovery that two lipids can traverse the tunnel concurrently, coupled with the observed sharp, high exit barrier (~+5 kcal/mol) for a single-lipid transfer (Chapter 5), provides strong evidence that CETP operates via a "through the tunnel" counter-exchange mechanism. This mechanism requires the attachment of a second lipoprotein at the opposite terminus to act as a thermodynamic sink, lowering the exit barrier and enabling the complete transfer to be physiologically feasible.

The acyl chain length of TG is a primary determinant of its kinetic profile (Chapter 7). Long-chain TGs (LCTs) exhibit the highest residence time (≈3.03 ns) and form a greater number of stable contacts with tunnel residues. It is essential to note that the residence times and contact probabilities reported herein are derived from steered molecular dynamics (SMD) simulations performed at a non-equilibrium velocity of 1 nm/ns. Consequently, these metrics reflect the mechanical resilience and kinetic hindrance of the triglyceride-protein interface under external load, rather than thermodynamic equilibrium affinities. Therefore, the observed residence times may include artifacts of the pulling speed. These values, however, represent the time required for a contact to withstand the increasing tension of the virtual spring before forced dissociation occurs. To obtain absolute, physiologically relevant values for binding free energies and residence times, these SMD-derived work distributions must be processed using the Jarzynski Equality or the Crooks Fluctuation Theorem. These theorems allow for the recovery of the equilibrium free energy profile ($\Delta G$), while the absolute residence times ($\tau$) can

be estimated by integrating the resulting potential of mean force (PMF) within a diffusional framework, such as Kramers' theory, to account for the actual timescales of thermal escape.

Nevertheless, these SMD-derived values are far from meaningless; they provide a mechanical fingerprint of the translocation pathway. While the absolute timescales are compressed, the relative ranking of residence times and contact persistence remains a robust indicator of which residues constitute the primary structural barriers. This allows us to identify the specific regions that dictate resistance within CETP tunnel. By capturing how the increased van der Waals footprint of LCTs leads to greater mechanical persistence than shorter chains. This increased interaction may lead to a slower exchange rate compared to medium- and short-chain TGs. This slower kinetic profile of LC-TG, the prevalent form of fat in the plasma of CVD patients, drives adverse lipoprotein remodeling, leading to the formation of pro-atherogenic small, dense LDL (sdLDL) particles. This finding establishes a mechanistic link between dietary long-chain fatty acids (LCFAs), slow CETP kinetics, and increased CVD risk, suggesting that diet could directly modulate CETP-mediated lipoprotein remodeling.

## 8.5 Summary and Future Directions

In summary, CETP is an extraordinary exchange protein whose function relies on a precisely orchestrated interplay of tunnel architecture, optimal hydrophobicity, and dynamic plasticity. The PL plugs act as essential cofactors that facilitate lipid exchange by increasing tunnel hydrophobicity for entry, physically occluding orifices to prevent lipid cargo loss, introducing disorder to prevent kinetic trapping, and reorienting the lipid to accelerate traversal.

These insights not only advance our fundamental understanding of lipid transfer within the BPI/LBP protein family, all members of which contain a continuous hydrophobic tunnel and central orifices, but also provide a powerful mechanistic foundation for designing and evaluating next-generation CETP-targeted therapeutics. While current inhibitor strategies focus on tunnel occlusion with hydrophobic drug molecules, this work suggests an alternative, smarter therapeutic strategy: targeting the more solvent-accessible PL-binding pockets. Modulating PL plug activity (e.g., by disrupting key salt

bridges) or regulating tunnel integrity may be a more effective way to control CETP function in patients with atherosclerosis. This thesis also emphasizes the power of molecular dynamics simulations to provide fundamental, valuable atomistic insights necessary to unravel complex biological mechanisms.

A primary limitation of the work is that it considers CETP and lipid dynamics in isolation rather than simulating the CETP-lipoprotein complex as a whole. Future work should focus on addressing the current limitations, specifically through molecular dynamics simulations that include the CETP-lipoprotein interface to understand binding kinetics. Furthermore, it should aim to definitively determine the kinetic and thermodynamic parameters of the concurrent bidirectional movement of CE & TG.

# REFERENCES


1. **Abedi, E.** and **M. Sahari** (2014). Long-chain polyunsaturated fatty acid sources and evaluation of their nutritional and functional properties. *Food Sci Nutr.*, **2**(5), 443–463.

2. **Abraham, M. J.**, **T. Murtola**, **R. Schulz**, **S. Páll**, **J. C. Smith**, **B. Hess**, and **E. Lindahl** (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1**, 19–25.

3. **Adachi, J.**, **P. J. Waddell**, **W. Martin**, and **M. Hasegawa** (2000). Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast dna. *Journal of molecular evolution*, **50**(4), 348–358.

4. **Aicart, E.**, **G. Tardajos**, and **M. D. Peña** (1981). Isothermal compressibility of cyclohexane+ n-tridecane and+ n-pentadecane at 298.15, 308.15, 318.15, and 333.15 k. *The Journal of Chemical Thermodynamics*, **13**(8), 783–788.

5. **Altschul, S. F.**, **W. Gish**, **W. Miller**, **E. W. Myers**, and **D. J. Lipman** (1990). Basic local alignment search tool. *Journal of molecular biology*, **215**(3), 403–410.

6. **Ashkenazy, H.**, **S. Abadi**, **E. Martz**, **O. Chay**, **I. Mayrose**, **T. Pupko**, and **N. Ben-Tal** (2016). Consurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, **44**(W1), W344–W350.

7. **Bacle, A.**, **R. Gautier**, **C. L. Jackson**, **P. F. J. Fuchs**, and **S. Vanni** (2017). Interdigitation between triglycerides and lipids modulates surface properties of lipid droplets. *Biophysical Journal*, **112**(7), 1417–1430. ISSN 1542-0086.

8. **Bairoch, A.** and **R. Apweiler** (1996). The swiss-prot protein sequence data bank and its new supplement trembl. *Nucleic acids research*, **24**(1), 21–25.

9. **Bakan, A.**, **L. Meireles**, and **I. Bahar** (2011). Prody: Protein dynamics inferred from theory and experiments. *Bioinformatics*, **27**, 1575–1577.

10. **Barter, P.** (2000). Cetp and atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **20**(9), 2029–2031.

11. **Barter, P.**, **S. Nicholls**, **J. Kastelein**, and **K. Rye** (2015). Is cholesteryl ester transfer protein inhibition an effective strategy to reduce cardiovascular risk? cetp inhibition as a strategy to reduce cardiovascular risk: The pro case. *Circulation*, **132**, 423–432.

12. **Bauer, J.**, **J. Pavlović**, and **V. Bauerová-Hlinková** (2019). Normal mode analysis as a routine part of a structural investigation. *Molecules*, **24**, 3293.

13. **Beamer, L.**, **S. Caroll**, and **D. Eisenberg** (1997). Crystal structure of human bpi and two bound phospholipids at 2.4 angstrom resolution. *Science*, **236**(276), 1861–1864.

14. **Beckstein, O.**, **E. Denning**, **J. Perilla**, and **T. Woolf** (2009). Zipping and unzipping of adenylate kinase: Atomistic insights into the ensemble of open to closed transitions. *Journal of Molecular Biology*, **394**, 160–176.

15. **Bekker, H.**, **H. Berendsen**, **E. Dijkstra**, and **et al** (1993). Gromacs - a parallel computer for molecular dynamics simulations. *4th International Conference on Computational Physics*, **(PC 92)**, 252–256.

16. **Berendsen, H.**, **P. van**, and **W. e. a. Van Gunsteren** (1984*a*). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, **81**, 3684–3690.

17. **Berendsen, H.**, **D. van der Spoel**, and **R. van Drunen** (1995). Gromacs: A message-passing parallel molecular dynamics implementation. *Computer physics communications*, **91**, 43–56.

18. **Berendsen, H. J.**, **J. Postma**, **W. F. van Gunsteren**, **A. DiNola**, and **J. R. Haak** (1984*b*). Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, **81**(8), 3684–3690.

19. **Best, R.**, **X. Zhu**, **J. Shim**, **P. Lopes**, **J. Mittal**, **M. Feig**, and **A. MacKerell** (2012). Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone $\phi$, $\psi$ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. *Journal of Chemical Theory and Computation*, **8**, 3257–3273.

20. **Bhatnagar, A.** (2017). Environmental determinants of cardiovascular disease. *Circulation research*, **121**(2), 162–180. ISSN 0009-7330.

21. **Bouillet, B.**, **T. Gautier**, **D. Blache**, **J.-P. Pais de Barros**, **L. Duvillard**, **J.-M. Petit**, **L. Lagrost**, and **B. Vergès** (2014). Glycation of apolipoprotein c1 impairs its cetp

inhibitory property: pathophysiological relevance in patients with type 1 and type 2 diabetes. *Diabetes Care*, **37**(4), 1148–1156. ISSN 1935-5548.

22. **Bowman, L.**, **J. Hopewell**, and **F. e. a. Chen** (2017). Effects of anacetrapib in patients with atherosclerotic vascular disease. *New England Journal of Medicine*, **377**(13), 1217–1227.

23. **Buša, J.**, **S. Hayryan**, **C.-K. Hu**, and **et al** (2010). Molaxis: efficient and accurate identification of channels in macromolecules. *Computational Physics communications*, **181**, 2116–2125.

24. **Campomanes, P.**, **J. Prabhu**, **V. Zoni**, and **S. Vanni** (2021). Recharging your fats: Charmm36 parameters for neutral lipids triacylglycerol and diacylglycerol. *Biophysical Reports*, **1**, 100034.

25. **Cao, G.**, **T. Beyer**, **Y. Zhang**, **R. Schmidt**, **Y. Chen**, and **S. e. a. Cockerham** (2011). Evacetrapib is a novel, potent, and selective inhibitor of cholesteryl ester transfer protein that elevates hdl cholesterol without inducing aldosterone or increasing blood pressure. *Journal of Lipid Research*, **52**, 2169–2176.

26. **Cesare Di, M.**, **P. Perel**, **S. Taylor**, **C. Kabdula**, **H. Bixby**, **T. Gaziano**, **D. McGhie**, and **J. Mwangi** (2024). The heart of the world. *Global Heart*, **19**(1), 11. ISSN 2211-8160.

27. **Charles, M. A.** and **J. P. Kane** (2012). New molecular insights into cetp structure and function: a review. *Journal of Lipid Research*, **53**(8), 1451–1458. ISSN 0022-2275.

28. **Chirasani, V.** and **S. Sanjib** (2017). How cholesteryl ester transfer protein can also be a potential triglyceride transporter. *Scientific Reports*, **7**(7), 6159.

29. **Chirasani, V. R.**, **P. D. Revanasiddappa**, and **S. Senapati** (2016*a*). Structural plasticity of cholesteryl ester transfer protein assists the lipid transfer activity. *The Journal of Biological Chemistry*, **291**(37), 19462–19473. ISSN 0021-9258.

30. **Chirasani, V. R.**, **R. Sankar**, and **S. Senapati** (2016*b*). Mechanism of inhibition of cholesteryl ester transfer protein by small molecule inhibitors. *The Journal of Physical Chemistry B*, **120**(33), 8254–8263. ISSN 1520-6106.

31. **Choudhuri, S.**, *Bioinformatics for beginners: genes, genomes, molecular evolution, databases and analytical tools*. Elsevier, 2014.

32. **Chovancova, E.**, **A. Pavelka**, **P. Benes**, and **et al** (2012). Caver 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Computational Biology*, **8**, e1002708.

33. **Cilpa-Karhu, G.**, **M. Jauhiainen**, and **M.-L. Riekkola** (2015). Atomistic md simulation reveals the mechanism by which cetp penetrates into hdl enabling lipid transfer from hdl to cetp. *Journal of Lipid Research*, **56**(1), 98–108. ISSN 1539-7262.

34. **Coleman, R.** and **K. Sharp** (2009). Finding and characterizing tunnels in macromolecules with application to ion channels and pores. *Biophysical journal*, **96**, 632–645.

35. **Dangas, K.**, **A. Navar**, and **J. Kastelein** (2022). The effect of cetp inhibitors on new-onset diabetes: a systematic review and meta-analysis. *European Heart Journal of Cardiovascular Pharmacotherapy*, **8**, 622–632.

36. **Darden, T.**, **D. York**, and **L. Pedersen** (1993). Particle mesh ewald: An n·log(n) method for ewald sums in large systems. *Journal of Chemical Physics*, **98**, 10089–10092.

37. **Dautin, G.**, **Z. Soltani**, **D. Ducloux**, **T. Gautier**, **J. P. Pais de Barros**, and **P. e. a. Gambert** (2007). Hemodialysis reduces plasma apolipoprotein c-i concentration making vldl a better substrate for lipoprotein lipase. *Kidney International*, **72**, 871–878.

38. **Davidson, M. H.**, **A. Hsieh**, and **J. J. P. Kastelein** (2024). Cholesteryl ester transfer protein inhibition: a pathway to reducing risk of morbidity and promoting longevity. *Current Opinion in Lipidology*, **35**(6), 303. ISSN 0957-9672.

39. **Dixit, S. M.**, **M. Ahsan**, and **S. Senapati** (2019). Steering the lipid transfer to unravel the mechanism of cholesteryl ester transfer protein inhibition. *Biochemistry*, **58**(36), 3789–3801. ISSN 0006-2960.

40. **Dobson, C. M.** (2019). Biophysical techniques in structural biology. *Annual Review of Biochemistry*, **88**(1), 25–33.

41. **Dorighello, G. G.**, **L. H. P. Assis**, **T. Rentz**, **J. Morari**, **M. F. M. Santana**, **M. Passarelli**, **N. D. Ridgway**, **A. E. Vercesi**, and **H. C. F. Oliveira** (2022). Novel role of

cetp in macrophages: Reduction of mitochondrial oxidants production and modulation of cell immune-metabolic profile. *Antioxidants*, **11**(9), 1734. ISSN 2076-3921.

42. **Dudev, T.** and **C. Lim** (2010). Factors governing the na+ vs k+ selectivity in sodium ion channels. *Journal of american chemial society*, **132**, 2321–2332.

43. **Dumont, L.**, **T. Gautier**, **J. P. de Barros**, **H. Laplanche**, **D. Blache**, and **e. a. Ducoroy, P.** (2005). Molecular mechanism of the blockade of plasma cholesteryl ester transfer protein by its physiological inhibitor apolipoprotein ci. *Journal of biological chemistry*, **280**, 38108–38116.

44. **Edgar, R. C.** (2004). Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**(5), 1792–1797.

45. **Fan, J.**, **Y. Liu**, **S. Yin**, **N. Chen**, **X. Bai**, **Q. Ke**, **J. Shen**, and **M. Xia** (2019). Small dense ldl cholesterol is associated with metabolic syndrome traits independently of obesity and inflammation. *Nutrition & Metabolism*, **16**, 7. ISSN 1743-7075.

46. **Farrer, S.** (2018). Beyond statins: Emerging evidence for hdl-increasing therapies and diet in treating cardiovascular disease. *Advances in Preventive Medicine*, **6024747**.

47. **Fincham, D.** (1992). Leapfrog rotational algorithms. molecular simulation. *Molecular Simulation*, **8**, 165–178.

48. **Forrest, M.**, **D. Bloomfield**, **R. Briscoe**, **P. Brown**, **A.-M. Cumiskey**, **J. Ehrhart**, and **J. e. a. Hershey** (2008). Torcetrapib-induced blood pressure elevation is independent of cetp inhibition and is accompanied by increased circulating levels of aldosterone. *British Journal of Pharmacology*, **154**, 1465–1473.

49. **Foucher, C.**, **L. Lagrost**, **V. Maupoil**, **M. le Meste**, **L. Rochette**, and **P. Gambert** (1996). Alterations of lipoprotein fluidity by non-esterified fatty acids known to affect cholesteryl ester transfer protein activity. an electron spin resonance study. *European Journal of Biochemistry*, **236**(2), 436–442. ISSN 0014-2956.

50. **Garg, P.**, **S. Sacher**, **P. Mrinal, Gautam**, **Atul**, and **A. Ray** (2022). Ciclop: a robust and accurate computational framework for protein inner cavity detection. *Bioinformatics*, **38**, 2153–2161.

51. **Ginsberg, H. N.**, **C. J. Packard**, **M. Chapman**, **J. Aguilar-Salinas**, and **M. e. a. Averna** (2021). Triglyceride-rich lipoproteins and their remnants: metabolic insights, role in atherosclerotic cardiovascular disease, and emerging therapeutic strategies—a consensus statement from the european atherosclerosis society. *European Heart Journal*, **42**(47), 4791–4806.

52. **Gora, A.**, **J. Brezovsky**, and **J. Damborsky** (2013). Gates of enzymes. *Chemical Reviews*, **113**, 5871–5923.

53. **Gowers, R.**, **N. Linke**, and **M. e. a. Barnoud** (2016). Mdanalysis: A python package for the rapid analysis of molecular dynamics simulations. *PROC. OF THE 15th PYTHON IN SCIENCE CONF. (SCIPY 2016) ; 2016-07-11 - 2016-07-11*.

54. **Green, S.** and **R. Pittman** (1991). Comparative acyl specificities for transfer and selective uptake of high density lipoprotein cholesteryl esters. *Journal of Lipid Research*, **32**(3), 457–467. ISSN 0022-2275.

55. **Grubmüller, H.**, **H. Heller**, **A. Windemuth**, and **K. Schulten** (1991). Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molecular Simulation*, **6**, 121–142.

56. **Grundy, S.** and **M. Denke** (1990). Dietary influences on serum lipids and lipoproteins. *Journal of Lipid Research*, **31**(7), 1149–1172.

57. **Hall, A.**, **J. Repakova**, and **I. Vattulainen** (2008). Modeling of the triglyceride-rich core in lipoprotein particles. *The Journal of Physical Chemistry B*, **112**(44), 13772–13782.

58. **Hao, D.**, **H. Wang**, **Y. Zang**, **L. Zhang**, **Z. Yang**, and **S. Zhang** (2022). Mechanism of glycans modulating cholesteryl ester transfer protein: Unveiled by molecular dynamics simulation. *Journal of Chemical Information and Modeling*, **62**(21), 5246–5257. ISSN 1549-9596.

59. **Harder, C.**, **P. Lau**, **A. Meng**, **S. C. Whitman**, and **R. McPherson** (2007). Cholesteryl ester transfer protein (cetp) expression protects against diet induced atherosclerosis in sr-bi deficient mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **27**(4), 858–864. ISSN 1524-4636.

60. **Hendlich, M.**, **F. Rippmann**, and **G. Barnickel** (1997). Ligsite: automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graph*, **15**, 359–363.

61. **Hess, B.**, **H. Bekker**, **H. Berendsen**, and **J. Fraaije** (1998). Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, **18**, 1463–1472.

62. **Humphrey, W.**, **A. Dalke**, and **K. Schulten** (1996). Vmd: visual molecular dynamics. *Journal of Molecular Graphics*, **14**, 33–38, 27–28.

63. **Ivanova, E. A.**, **V. A. Myasoedova**, **A. A. Melnichenko**, **A. V. Grechko**, and **A. N. Orekhov** (2017). Small dense low-density lipoprotein as biomarker for atherosclerotic diseases. *Oxidative Medicine and Cellular Longevity*, **2017**, 1273042. ISSN 1942-0900.

64. **Johns, D.**, **S. Wang**, **R. Rosa**, **J. Hubert**, **S. Xu**, **Y. Chen**, **T. Bateman**, and **R. Blaustein** (2019). Impact of drug distribution into adipose on tissue function: The cholesteryl ester transfer protein (cetp) inhibitor anacetrapib as a test case. *Pharmacology Research Perspective*, **7**, e00543.

65. **Johnson, L. S.**, **S. R. Eddy**, and **E. Portugaly** (2010). Hidden markov model speed heuristic and iterative hmm search procedure. *BMC bioinformatics*, **11**(1), 1–8.

66. **Jonckheere, A. I.**, **J. A. M. Smeitink**, and **R. J. T. Rodenburg** (2012). Mitochondrial ATP synthase: architecture, function and pathology. *Journal of Inherited Metabolic Disease*, **35**(2), 211–225. ISSN 0141-8955.

67. **Jones, P. J.**, **A. S. Pappu**, **L. Hatcher**, **Z.-C. Li**, **D. R. Illingworth**, and **W. E. Connor** (1996). Dietary cholesterol feeding suppresses human cholesterol synthesis measured by deuterium incorporation and urinary mevalonic acid levels. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **16**(10), 1222–1228.

68. **Jorgensen, W.**, **J. Chandrasekhar**, **J. Madura**, **R. Impey**, and **M. Klein** (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, **79**, 926–935.

69. **Kaminski, G. A.**, **R. A. Friesner**, **J. Tirado-Rives**, and **W. L. Jorgensen** (2001). Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, **105**(28), 6474–6487.

70. **Kastelein, J.**, **J. Besseling**, **S. Shah**, **J. Bergeron**, **G. Langslet**, and **G. e. a. Hovingh** (2015). Anacetrapib as lipid-modifying therapy in patients with heterozygous familial hypercholesterolaemia (realize): a randomised, double-blind, placebo-controlled, phase 3 study. *Lancet*, **385**, 2153–2161.

71. **Kastelein, J.**, **A. Hsieh**, **M. Dicklin**, **M. Ditmarsch**, and **M. Davidson** (2023). Obicetrapib: Reversing the tide of cetp inhibitor disappointments. *Current Atherosclerosis Reports*, **26**, 35–44.

72. **Kawano, K.**, **S.-C. Qin**, **M. Lin**, **A. R. Tall**, and **X.-c. Jiang** (2000). Cholesteryl ester transfer protein and phospholipid transfer protein have nonoverlapping functions in vivo. *Journal of Biological Chemistry*, **275**(38), 29477–29481. ISSN 0021-9258.

73. **Kerimov, A.** and **T. Apaev** (1972). Experimental values of density for hexene-1, octene-1, cyclohexane, and methylcyclohexane in varying temperatures and pressures. *Teplofiz. Svoistva Veshchestv Mater*, **5**, 26–46.

74. **Keshavamurthy, A.**, **P. D. Revanasiddappa**, **S. M. Dixit**, and **G. R. Priyanka** (2023). Bound phospholipids assist cholesteryl ester transfer in the cholesteryl ester transfer protein. *Journal of Chemical Information and Modeling*, **63**(10), 3054–3067. ISSN 1549-960X.

75. **Khatana, C.**, **N. K. Saini**, **S. Chakrabarti**, **V. Saini**, **A. Sharma**, **R. V. Saini**, and **A. K. Saini** (2020). Mechanistic insights into the oxidized low-density lipoprotein-induced atherosclerosis. *Oxidative Medicine and Cellular Longevity*, **2020**, 5245308. ISSN 1942-0900.

76. **Kim, S.**, **M. I. Oh**, and **J. M. J. Swanson** (2021). Stressed lipid droplets: How neutral lipids relieve surface tension and membrane expansion drives protein association. *The journal of physical chemistry. B*, **125**(21), 5572–5586. ISSN 1520-6106.

77. **Klauda, J.**, **R. Venable**, **J. Freites**, **J. O'Connor**, **D. J. Tobias**, and **C. e. a. Mondragon-Ramirez** (2010). Update of the charmm all-atom additive force field for lipids: Validation on six lipid types. *Journal of Physical Chemistry B.*, **114**, 7830–7843.

78. **Koivuniemi, A.**, **T. Vuorela**, **P. T. Kovanen**, **I. Vattulainen**, and **M. T. Hyvönen** (2012). Lipid exchange mechanism of the cholesteryl ester transfer protein clarified

by atomistic and coarse-grained simulations. *PLOS Computational Biology*, **8**(1), e1002299. ISSN 1553-7358.

79. **Kontush, A.** (2015). Hdl particle number and size as predictors of cardiovascular disease. *Fronteirs in Pharmacology*, **6**, 218.

80. **Kumar, S.**, **J. Rosenberg**, **D. Bouzida**, and **et al** (1992). The weighted histogram analysis method for free-energy calculations on biomolecules i. the method. *Journal of Computational Chemistry*, **13**, 1011–1021.

81. **Kwiterovich, P. O.** (2000). The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: a current review. *American Journal of Cardiology*, **86**(12), 5–10. ISSN 0002-9149, 1879-1913.

82. **Lagrost, L.** and **P. Barter** (1991*a*). Effects of various non-esterified fatty acids on the particle size redistribution of high density lipoproteins induced by the human cholesteryl ester transfer protein. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, **1082**(2), 204–210.

83. **Lagrost, L.** and **P. Barter** (1991*b*). Effects of various non-esterified fatty acids on the transfer of cholesteryl esters from hdl to ldl induced by the cholesteryl ester transfer protein. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, **1085**(2), 209–216.

84. **Laskowski, R.** (1995). Surfnet: A program for visualising molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graph*, **13**, 322–330.

85. **Lauer, M. E.**, **A. Graff-Meyer**, **A. C. Rufer**, **C. Maugeais**, **E. von der Mark**, **H. Matile**, **B. D'Arcy**, **C. Magg**, **P. Ringler**, **S. A. Müller**, **S. Scherer**, **G. Dernick**, **R. Thoma**, **M. Hennig**, **E. J. Niesor**, and **H. Stahlberg** (2016). Cholesteryl ester transfer between lipoproteins does not require a ternary tunnel complex with cetp. *Journal of Structural Biology*, **194**(2), 191–198. ISSN 1095-8657.

86. **Le, S. Q.** and **O. Gascuel** (2008). An improved general amino acid replacement matrix. *Molecular biology and evolution*, **25**(7), 1307–1320.

87. **Le Guilloux, V.**, **P. Schmidtke**, and **P. Tuffery** (2009). Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics*, **10**(1), 1–11.

88. **Lee, H.**, **A. Akhmedov**, and **C. Chen** (2022). Spotlight on very-low-density lipoprotein as a driver of cardiometabolic disorders: Implications for disease progression and mechanistic insights. *Front. Cardiovasc. Med.*, **9**.

89. **Lei, D.**, **M. Rames**, and **X. e. a. Zhang** (2016). Insights into the tunnel mechanism of cholesteryl ester transfer protein through all-atom molecular dynamics simulations. *Journal of Biological Chemistry*, **291**, 14034–14044.

90. **Lemaitre, R.** and **I. King** (2022). Very long-chain saturated fatty acids and diabetes and cardiovascular disease. *Curr Opin Lipidol.*, **33**(1), 76–82.

91. **Lewis, G.** and **D. Rader** (2005). New insights into the regulation of hdl metabolism and reverse cholesterol transport. *Circulation Research*, **96**(12), 1221–1232.

92. **Libby, P.** (2021). The changing landscape of atherosclerosis. *Nature*, **592**, 524–533.

93. **Lichtinger, S. M.** and **P. C. Biggin** (2023). Tackling hysteresis in conformational sampling: How to be forgetful with memento. *Journal of Chemical Theory and Computation*, **19**(12), 3705–3720. PMID: 37285481.

94. **Lindahl, E.**, **B. Hess**, and **D. van der Spoel** (2001). Gromacs 3.0: a package for molecular simulation and trajectory analysis. *Molecular Modelling*, **7**, 306–317.

95. **Liu, S.**, **A. Mistry**, **J. M. Reynolds**, **D. B. Lloyd**, **M. C. Griffor**, **D. A. Perry**, **R. B. Ruggeri**, **R. W. Clark**, and **X. Qiu** (2012). Crystal structures of cholesteryl ester transfer protein in complex with inhibitors. *The Journal of Biological Chemistry*, **287**(44), 37321–37329. ISSN 0021-9258.

96. **Luo, R.**, **L. David**, **H. Hung**, and **et al** (1999). Strength of solvent-exposed salt-bridges. *Journal of Physical Chemistry B*, **103**, 727–736.

97. **Ma, W.**, **C. Tang**, and **L. Lai** (2005). New insights into the regulation of hdl metabolism and reverse cholesterol transport. *Biophysical journal*, **89**, 1183–1193.

98. **MacLean, P.**, **S. Vadalamudi**, **E. Hao**, and **H. Barakat** (2000). Differential expression of cholesteryl ester transfer protein in the liver and plasma of fasted and fed transgenic mice. *The Journal of Nutritional Biochemistry*, **11**(6), 318–325. ISSN 0955-2863.

99. **Mahoney, W., M. W.and Jorgensen** (2000). A five-site model for liquid water. *Journal of Chemical Physics*, **112**, 8910–8922.

100. **Mark, P.** and **L. Nilsson** (2001*a*). Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, **105**, 9954–9960.

101. **Mark, P.** and **L. Nilsson** (2001*b*). Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *The Journal of Physical Chemistry A*, **105**(43), 9954–9960.

102. **Martin, L.**, **P. Connelly**, **D. Nancoo**, **N. Wood**, **Z. Zhang**, **G. Maguire**, **E. Quinet**, **A. Tall**, **Y. Marcel**, and **R. McPherson** (1993). Cholesteryl ester transfer protein and high density lipoprotein responses to cholesterol feeding in men: relationship to apolipoprotein e genotype. *Journal of Lipid Research*, **34**(3), 437–446. ISSN 0022-2275.

103. **Mason, S.** and **B. Chen** (2018). Exploring protein cavities through rigidity analysis. *Molecules*, **23**, 351.

104. **Masson, D.**, **A. Athias**, and **L. Lagrost** (1996). Evidence for electronegativity of plasma high density lipoprotein-3 as one major determinant of human cholesteryl ester transfer protein activity. *Journal of Lipid Research*, **37**(7), 1579–1590.

105. **Mayrose, I.**, **D. Graur**, **N. Ben-Tal**, and **T. Pupko** (2004). Comparison of site-specific rate-inference methods for protein sequences: empirical bayesian methods are superior. *Molecular biology and evolution*, **21**(9), 1781–1791.

106. **Mehta, N.**, **K. Dangas**, **M. Ditmarsch**, **P. C. N. Rensen**, **M. R. Dicklin**, and **J. J. P. Kastelein** (2023). The evolving role of cholesteryl ester transfer protein inhibition beyond cardiovascular disease. *Pharmacological Research*, **197**, 106972. ISSN 1043-6618.

107. **Mercadante, D.**, **F. Gräter**, and **C. Daday** (2018). Conan: A tool to decode dynamical information from molecular interaction maps. *Biophysical Journal*, **114**, 1267–1273.

108. **Metzinger, M. P.**, **S. Saldanha**, **J. Gulati**, **K. V. Patel**, **A. El-Ghazali**, **S. Deodhar**, **P. H. Joshi**, **C. Ayers**, and **A. Rohatgi** (2020). Effect of anacetrapib on cholesterol efflux capacity: A substudy of the define trial. *Journal of the American Heart Association*, **9**(24), e018136. ISSN 2047-9980.

109. **Michaud-Agrawal, N.**, **E. Denning**, **T. Woolf**, and **O. Beckstein** (2011). Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of Compututational Chemistry*, **32**, 2319–2327.

110. **Miller, K.** and **D. Small** (1983). Surface-to-core and interparticle equilibrium distributions of triglyceride-rich lipoprotein lipids. *Journal of Biological Chemistry*, **258**(22), 13772–13784.

111. **Mineo, C.**, **H. Deguchi**, **J. Griffin**, and **P. Shaul** (2006). Endothelial and antithrombotic actions of hdl. *Circulation Research*, **98**, 1352–1364.

112. **Morton, R.** and **J. Steinbrunner** (1990). Concentration of neutral lipids in the phospholipid surface of substrate particles determines lipid transfer protein activity. *Journal of Lipid Research*, **31**, 1559–1567.

113. **Morton, R.** and **D. Zilversmit** (1983). Inter-relationship of lipids transferred by the lipid-transfer protein isolated from human lipoprotein-deficient plasma. *The journal of biological chemistry*, **258**(19), 11751–11757.

114. **Morton, R. E.** and **D. J. Greene** (2003*a*). Cetp and lipid transfer inhibitor protein are uniquely affected by the negative charge density of the lipid and protein domains of ldl. *Journal of Lipid Research*, **44**(12), 2287–2296. ISSN 0022-2275.

115. **Morton, R. E.** and **D. J. Greene** (2003*b*). The surface cholesteryl ester content of donor and acceptor particles regulates cetp: a liposome-based approach to assess the substrate properties of lipoproteins. *Journal of Lipid Research*, **44**(7), 1364–1372. ISSN 0022-2275.

116. **Morton, R. E.** and **L. Izem** (2014). Cholesteryl ester transfer proteins from different species do not have equivalent activities. *Journal of Lipid Research*, **55**(2), 258–265. ISSN 0022-2275.

117. **Morton, R. E.** and **L. Izem** (2015). Modification of cetp function by changing its substrate preference: a new paradigm for cetp drug design. *Journal of lipid research*, **56**(3), 612–619.

118. **Morton, R. E.** and **Y. Liu** (2020*a*). The lipid transfer properties of cetp define the concentration and composition of plasma lipoproteins. *Journal of Lipid Research*, **61**(8), 1168–1179. ISSN 1539-7262.

119. **Morton, R. E.** and **Y. Liu** (2020*b*). The lipid transfer properties of cetp define the concentration and composition of plasma lipoproteins. *Journal of lipid research*, **61**(8), 1168–1179.

120. **Murphy, A.**, **K. Woollard**, **A. Hoang**, **N. Mukhamedova**, **R. Stirzaker**, **S. McCormick**, **A. T. Remaley**, **D. Sviridov**, and **J. Dusting** (2008). High-density lipoprotein reduces the human monocyte inflammatory response. *Arterioscleorsis Thrombosis Vascular Biology*, **28**, 2071–2074.

121. **Nagano, M.**, **S. Yamashita**, **K.-I. Hirano**, **M. Takano**, **T. Maruyama**, **M. Ishihara**, **Y. Sagehashi**, **T. Kujiraoka**, **K. Tanaka**, **H. Hattori**, **N. Sakai**, **N. Nakajima**, **T. Egashira**, and **Y. Matsuzawa** (2004). Molecular mechanisms of cholesteryl ester transfer protein deficiency in japanese. *Journal of Atherosclerosis and Thrombosis*, **11**(3), 110–121. ISSN 1340-3478.

122. **Naushin, S.** (2024). *Multiomics approach to identify potential biomarkers of coronary artery disease Indian perspective*. PhD dissertation, Academy of Scientific and Innovative Research (AcSIR)(CSIR-IGIB).

123. **Nayal, M.** and **B. Honig** (2006). On the nature of cavities on protein surface: Applications to the identification of drug-binding sites. *Proteins: Structural, Functional Bioinformatics*, **63**, 892–906.

124. **Negre-Salvayre, A.**, **N. Dousset**, **G. Ferretti**, **T. Bacchetti**, **G. Curatola**, and **R. Salvayre** (2006). Antioxidant and cytoprotective properties of high-density lipoproteins in vascular cells. *Free Radical Biology and Medicine*, **41**, 1031–1040.

125. **Nicholls, S.** and **K. Bubb** (2020). The mystery of evacetrapib - why are cetp inhibitors failing? *Expert Review of Cardiovascular Therapy*, **52**, 127–130.

126. **Nicholls, S. J.**, **M. Ditmarsch**, **J. J. Kastelein**, **S. P. Rigby**, **D. Kling**, **D. L. Curcio**, **N. J. Alp**, and **M. H. Davidson** (2022). Lipid lowering effects of the cetp inhibitor obicetrapib in combination with high-intensity statins: a randomized phase 2 trial. *Nature Medicine*, **28**(8), 1672–1678. ISSN 1546-170X.

127. **Niessen, K.**, **M. Xu**, **A. Paciaroni**, and **et al** (2017). Moving in the right direction: Protein vibrations steering function. *Biophysical Journal*, **112**, 933–942.

128. **Nishida, H., H.I.and Arai** and **T. Nishida** (1993). Cholesterol ester transfer mediated by lipid transfer protein as influenced by changes in the charge characteristics of plasma lipoproteins. *Journal of Biological Chemistry*, **268**(22), 16352–16360.

129. **Nissen, S.**, **J.-C. Tardif**, **S. Nicholls**, **J. Revkin**, **C. Shear**, **W. Duggan**, **W. Ruzyllo**, **W. Bachinsky**, **G. Lasala**, **E. Tuzcu**, and **I. Investigators** (2007). Effect of torcetrapib on the progression of coronary atherosclerosis. *New England Journal of Medicine*, **13**, 1304–1316.

130. **Nordestgaard, B. G.** and **A. Varbo** (2014). Triglycerides and cardiovascular disease. *Lancet (London, England)*, **384**, 626–635.

131. **Oestereich, F.**, **N. Yousefpour**, **E. Yang**, **J. Phénix**, **Z. S. Nezhad**, **A. Nitu**, **A. Vázquez Cobá**, **A. Ribeiro-da Silva**, **P. Chaurand**, and **L. M. Munter** (2022). The cholesteryl ester transfer protein (cetp) raises cholesterol levels in the brain. *Journal of Lipid Research*, **63**(9), 100260. ISSN 1539-7262.

132. **Packard, C.** (2023). Triglyceride-rich lipoproteins and their role in cardiovascular disease. *Br. J. Cardiol.*, **30**, S4–S9.

133. **Packard, C. J.**, **J. Boren**, and **M.-R. Taskinen** (2020). Causes and consequences of hypertriglyceridemia. *Frontiers in Endocrinology*, **11**, 252. ISSN 1664-2392.

134. **Parini, P.** and **L. L. Rudel** (2003). Is there a need for cholesteryl ester transfer protein inhibition? *Arteriosclerosis, Thrombosis, and Vascular Biology*, **23**(3), 374–375.

135. **Parrinello, M.** and **A. Rahman** (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, **52**, 7182–7190.

136. **Pathirannahalage, S.**, **M. Nastaran**, **A. Elbourne**, **A. C. G. Weiss**, **C. McConville**, and **A. e. a. Padua** (2021). Systematic comparison of the structural and dynamic properties of commonly used water models for molecular dynamics simulations. *Journal of Chemical Information and Modeling*, **61**, 4521–4536.

137. **Pattnaik, N.** and **D. Zilversmit** (1979). Interaction of cholesteryl ester exchange protein with human plasma lipoproteins and phospholipid vesicles. *Journal of Biological Chemistry*, **254**(8), 2782–2786.

138. **Pellegrini-Calace, M.**, **T. Maiwald**, and **J. M. Thornton** (2009*a*). PoreWalker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Computational Biology*, **5**(7). ISSN 1553-734X.

139. **Pellegrini-Calace, M.**, **T. Maiwald**, and **J. M. Thornton** (2009*b*). Porewalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comput Biol*, **5**(7), e1000440.

140. **Perna, M.** and **S. Hewlings** (2023). Saturated fatty acid chain length and risk of cardiovascular disease: A systematic review. *Nutrients*, **15**(1), 30.

141. **Pettersen, E.**, **T. Goddard**, **C. Huang**, and **et al** (2004). Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, **25**, 1605–1612.

142. **Pillois, X.**, **T. Gautier**, **B. Bouillet**, **J. P. Pais de Barros**, **A. Jeannin**, and **B. e. a. Vergès** (2012). Constitutive inhibition of plasma cetp by apolipoprotein c1 is blunted in dyslipidemic patients with coronary artery disease. *Journal of lipid research*, **53**, 1200–1209.

143. **Price, D.** and **C. Brooks** (2004). A modified tip3p water potential for simulation with ewald summation. *Journal of Chemical Physics*, **121**, 10096–10103.

144. **Pupko, T.**, **R. E. Bell**, **I. Mayrose**, **F. Glaser**, and **N. Ben-Tal** (2002). Rate4site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**(suppl_1), S71–S77.

145. **Qiu, X.**, **A. Mistry**, and **M. e. a. Ammiratu** (2007). Crystal structure of cholesteryl ester transfer protein reveals a long tunnel and four bound lipid molecules. *Nature Structural and Molecular Biology*, **14**, 106–113.

146. **Ranalletta, M.**, **K. K. Bierilo**, **Y. Chen**, **D. Milot**, **Q. Chen**, **E. Tung**, **C. Houde**, **N. H. Elowe**, **M. Garcia-Calvo**, **G. Porter**, **S. Eveland**, **B. Frantz-Wattley**, **M. Kavana**, **G. Addona**, **P. Sinclair**, **C. Sparrow**, **E. A. O'Neill**, **K. S. Koblan**, **A. Sitlani**, **B. Hubbard**, and **T. S. Fisher** (2010). Biochemical characterization of cholesteryl ester transfer protein inhibitors. *Journal of Lipid Research*, **51**(9), 2739–2752. ISSN 1539-7262.

147. **Raposo, H. F.**, **P. Forsythe**, **B. Chausse**, **J. Z. Castelli**, **P. M. Moraes-Vieira**, **V. S. Nunes**, and **H. C. F. Oliveira** (2021). Novel role of cholesteryl ester transfer protein

(cetp): attenuation of adiposity by enhancing lipolysis and brown adipose tissue activity. *Metabolism*, **114**, 154429. ISSN 0026-0495.

148. **Revanasiddappa, P.**, **R. Sankar**, and **S. e. a. Senapati** (2018). Role of the bound phospholipids in the structural stability of cholesteryl ester transfer protein. *Journal of Physical Chemistry*, **122**, 4239–4248.

149. **Revanasiddappa, P. D.** (2023). Structural insights on the deformations induced by various mutations on cholesteryl ester transfer protein. *Biophysical Chemistry*, **301**, 107093. ISSN 1873-4200.

150. **Rhainds, D.**, **B. Arsenault**, **M. Brodeur**, and **J. Tardif** (2012). An update on the clinical development of dalcetrapib (ro4607381), a cholesteryl ester transfer protein modulator that increases hdl cholesterol levels. *Future Cardiology*, **8**, 513–531.

151. **Richelle, M.**, **Y. A. Carpentier**, and **R. J. Deckelbaum** (1994). Long- and medium-chain triacylglycerols in neutral lipid-exchange processes with human plasma low-density lipoproteins. *Biochemistry*, **33**(16).

152. **Rick, S.** and **S. Stuart** (2002). Potentials and algorithms for incorporating polarizability in computer simulations. *Reviews in Computational Chemistry.*, 89–146.

153. **Rouland, A.**, **D. Masson**, **L. Lagrost**, **B. Vergès**, **T. Gautier**, and **B. Bouillet** (2022). Role of apolipoprotein c1 in lipoprotein metabolism, atherosclerosis and diabetes: a systematic review. *Cardiovascular Diabetology*, **21**, 272.

154. **Rye, K.**, **C. Bursill**, **G. Lambert**, **F. Tabet**, and **P. Barter** (2009). The metabolism and anti-atherogenic properties of hdl. *Journal of Lipid Research*, **50**, 195–200.

155. **Sacher, S.**, **A. Mukherjee**, and **A. Ray** (2023). Deciphering structural aspects of reverse cholesterol transport: mapping the knowns and unknowns. *Biol Rev Camb Philos Soc*, **98**, 1160–1183.

156. **Salekeen, R.**, **A. N. Haider**, **F. Akhter**, **M. M. Billah**, **M. E. Islam**, and **K. M. Didarul Islam** (2022). Lipid oxidation in pathophysiology of atherosclerosis: Current understanding and therapeutic strategies. *International Journal of Cardiology. Cardiovascular Risk and Prevention*, **14**, 200143. ISSN 2772-4875.

157. **Sammett, D.** and **A. Tall** (1985). Mechanisms of enhancement of cholesteryl ester transfer protein activity by lipolysis. *Journal of Biological Chemistry*, **260**(11), 6687–6697.

158. **Schmidtke, P.**, **A. Bidon-Chanal**, **F. Luque**, and **X. Barril** (2011). Mdpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, **27**, 3726–3285.

159. **Schneider, C.**, **W. Rasband**, and **K. Eliceiri** (2012). Nih image to imagej: 25 years of image analysis. *Nature Methods*, **9**, 671–675.

160. **Schrauber, H.**, **F. Eisenhaber**, and **P. Argos** (1993). Rotamers: To be or not to be?: An analysis of amino acid side-chain conformations in globular proteins. *Journal of Molecular Biology*, **230**, 592–612.

161. **Sehnal, D.**, **V. R. Svobodová**, and **K. e. a. Berka** (2013). Mole 2.0: advanced approach for analysis of bio-macromolecular channels. *Journal of Chemical Information*, **5**, 39.

162. **Serdyuk, A. P.** and **R. E. Morton** (1999). Lipid transfer inhibitor protein defines the participation of lipoproteins in lipid transfer reactions: Cetp has no preference for cholesteryl esters in hdl versus ldl. *Arteriosclerosis, thrombosis, and vascular biology*, **19**(3), 718–726.

163. **Shannon, P.**, **A. Markiel**, and **O. e. a. Ozier** (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.

164. **Shrestha, S.**, **B. Wu**, **L. Guiney**, **P. Barter**, and **K. Rye** (2017). Cholesteryl ester transfer protein and its inhibitors. *Journal of Lipid Research*, **59**, 772–783.

165. **Simões, T.**, **D. Lopes**, **S. Dias**, **F. Fernandes**, **J. Pereira**, **J. Jorge**, **C. Bajaj**, and **A. Gomes** (2017). Geometric detection algorithms for cavities on protein surfaces in molecular graphics: A survey. *journal of the European Association for Computer Graphics*, **36**(1), 643–683.

166. **Smart, O.**, **J. Neduvelil**, and **X. Wang** (1996). Hole: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graph*, **14**, 354–360.

167. **Stange, E.**, **B. Agostini**, and **J. Papenberg** (1975). Changes in rabbit lipoprotein properties by dietary cholesterol, and saturated and polyunsaturated fats. *Atherosclerosis*, **22**(1), 125–148.

168. **Tall, A.** (1993). Plasma cholesteryl ester transfer protein. *The journal of biological chemistry*, **34**(8), 1255–1274.

169. **Tall, A.** and **D. Rader** (2018). Trials and tribulations of cetp inhibitors. *Circulation Research*, **122**, 106–112.

170. **Turton, D.**, **H. Senn**, **T. Harwood**, and **et al** (2014). Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nature Communications*, **5**, 3999.

171. **Venancio, T. M.**, **R. M. Machado**, **A. Castoldi**, **M. T. Amano**, **V. S. Nunes**, **E. C. R. Quintao**, **N. O. S. Camara**, **F. G. Soriano**, and **P. M. Cazita** (2016). Cetp lowers tlr4 expression which attenuates the inflammatory response induced by lps and polymicrobial sepsis. *Mediators of Inflammation*, **2016**, 1784014. ISSN 0962-9351.

172. **Wang, S.**, **L. P. Deng**, **M. L. Brown**, **L. B. Agellon**, and **A. R. Tall** (1991). Structure-function studies of human cholesteryl ester transfer protein by linker insertion scanning mutagenesis. *Biochemistry*, **30**(14), 3484–3490. ISSN 0006-2960.

173. **Wazir, M.**, **O. A. Olanrewaju**, **M. Yahya**, **J. Kumari**, **N. Kumar**, **J. Singh**, **A. Y. Abbas Al-itbi**, **K. Kumari**, **A. Ahmed**, **T. Islam**, **G. Varrassi**, **M. Khatri**, **S. Kumar**, **H. Wazir**, and **S. S. Raza** (2023). Lipid disorders and cardiovascular risk: A comprehensive analysis of current perspectives. *Cureus*, **15**(12), e51395. ISSN 2168-8184.

174. **Weber, O.**, **H. Bischoff**, **C. Schmeck**, and **M.-F. Böttcher** (2010). Cholesteryl ester transfer protein and its inhibition. *Cell Molecular Life Science*, **67**(18), 3139.

175. **Whelan, S.** and **N. Goldman** (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular biology and evolution*, **18**(5), 691–699.

176. **Wong, L.**, **A. Copic**, and **T. Levine** (2017). Advances on the transfer of lipids by lipid transfer proteins. *Trends Biochemistry Society*, **42**, 516–530.

177. **Xue, M.**, **T. Wakamoto**, and **C. Kejlberg** (2019). How internal cavities destabilize a protein. *PNAS*, **116**, 21031–21036.

178. **Yaffe, D., E.and Fishelovitch**, **H. Wolfson**, and **et al** (2008). Molaxis: efficient and accurate identification of channels in macromolecules. *Proteins*, **73**, 72–86.

179. **Yamashita, S.**, **K. Hirano**, **N. Sakai**, and **Y. Matsuzawa** (2000). Molecular biology and pathophysiological aspects of plasma cholesteryl ester transfer protein. *Biochimica Et Biophysica Acta*, **1529**(1), 257–275. ISSN 0006-3002.

180. **Yang, Z.**, **Y. Cao**, **D. Hao**, **X. Yuan**, **L. Zhang**, and **S. Zhang** (2018). Binding profiles of cholesterol ester transfer protein with current inhibitors: a look at mechanism and drawback. *Journal of Biomolecular Structure and Dynamics*, **10**, 2567–2580.

181. **Yang, Z.**, **R. Nielsen**, and **M. Hasegawa** (1998). Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular biology and evolution*, **15**(12), 1600–1611.

182. **Yetukuri, L.**, **S. Söderlund**, **A. Koivuniemi**, **T. Seppänen-Laakso**, **P. Niemelä**, and **M. e. a. Hyvönen** (2010). Composition and lipid spatial distribution of hdl particles in subjects with low and high hdl-cholesterol[s]. *Journal of Lipid Research*, **51**(8), 2341–2351.

183. **Zak, Z.**, **L. Lagrost**, **T. Gautier**, **D. Masson**, **V. Deckert**, **L. Duverneuil**, **J.-P. P. De Barros**, **N. Le Guern**, **L. Dumont**, **M. Schneider**, **V. Risson**, **P. Moulin**, **D. Autran**, **G. Brooker**, **J. Sassard**, and **A. Bataillard** (2002). Expression of simian cetp in normolipidemic fisher rats has a profound effect on large-sized apoe-containing hdl. *Journal of Lipid Research*, **43**(12), 2164–2171. ISSN 0022-2275.

184. **Zhang, L.**, **F. Yan**, **S. Zhang**, **D. Lei**, **M. A. Charles**, **G. Cavigiolio**, **M. Oda**, **R. M. Krauss**, **K. H. Weisgraber**, **K.-A. Rye**, **H. J. Pownall**, **X. Qiu**, and **G. Ren** (2012). Structural basis of transfer between lipoproteins by cholesteryl ester transfer protein. *Nature Chemical Biology*, **8**(4), 342–349. ISSN 1552-4469.

185. **Zhang, M.**, **R. Charles**, and **H. e. a. Tong** (2015). Hdl surface lipids mediate cetp binding as revealed by electron microscopy and molecular dynamics simulation. *Scientific Reports*, **5**, 8741.

186. **Zhang, M.**, **D. Lei**, **B. Peng**, **M. Yang**, **L. Zhang**, **M. A. Charles**, **K.-A. Rye**, **R. M. Krauss**, **D. G. Johns**, and **G. Ren** (2017). Assessing the mechanisms of cholesteryl

ester transfer protein inhibitors. *Biochimica Et Biophysica Acta. Molecular and Cell Biology of Lipids*, **1862**(12), 1606–1617. ISSN 1388-1981.

187. **Zhang, S.**, **J. Krieger**, and **Y. e. a. Zhang** (2021). Prody 2.0: increased scale and scope after 10 years of protein dynamics modelling with python. *Bioinformatics*, **37**, 3657–3659.

188. **Zhu, H.** and **M. Pisabarro** (2011). Mspocket: an orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*, **27**, 351–358.

189. **Zhu, X.**, **P. Lopes**, and **A. MacKerell** (2011). Recent developments and applications of the charmm force fields. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2**, 167–185.

190. **Ølnes, .**, **M. Teigen**, **J. Laerdahl**, and **et al** (2023). Variants in the cetp gene affect levels of hdl cholesterol by reducing the amount, and not the specific lipid transfer activity, of secreted cetp. *PLOS ONE*, **18**, e0294764.