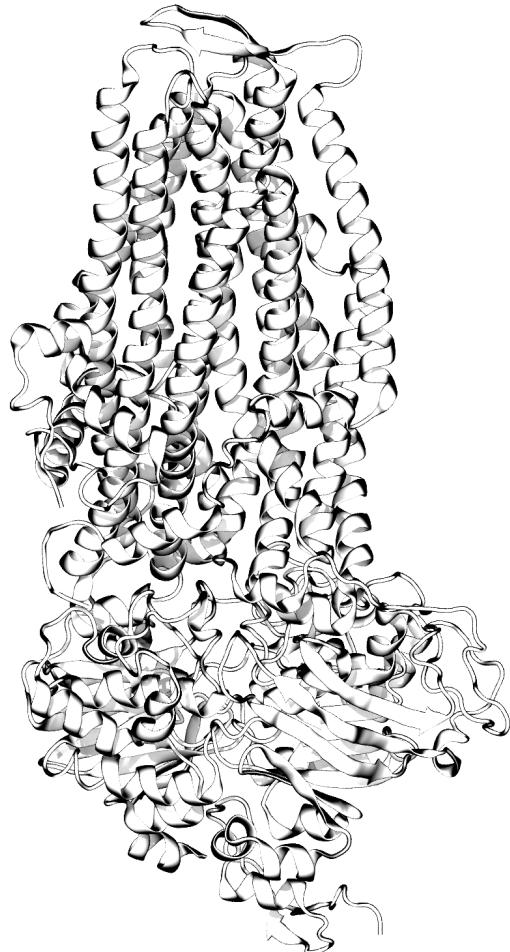


Computer Modelling the Root Cause of Cystic Fibrosis

by Miro Alexander Astore



*A thesis submitted in fulfilment of the
requirements for the degree of*

Doctor of Philosophy

The School of Physics
Faculty of Science
The University of Sydney
2022

Declaration of Original contribution

of the dissertation submitted by

Miro Alexander Astore

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.



09/09/2022

Miro Alexander Astore, Author

Date

Abstract

Sophisticated quantitative and physical thinking is revolutionising biology. The refinement of theoretical models in physical chemistry, alongside the growing capabilities computational engines and biochemical techniques are giving unprecedented insight into the structure and dynamics of living things. Here we will present a small example of this paradigm shift, by demonstrating how a physical mindset can contribute to the study of a disease—Cystic Fibrosis.

Cystic Fibrosis is the most common fatal genetic condition in Caucasians. It is a debilitating disease, significantly shortening the life span of patients and degrading their quality of life. It is caused by deleterious mutations to a protein known as the Cystic Fibrosis Transmembrane conductance Regulator (CFTR). This protein acts as an anion channel—in order to balance the levels of salts across epithelial membranes, it must conduct chloride and bicarbonate.

Over the last decade there have been an increasing number of clinically approved small molecule drugs, called CFTR modulators, which act directly on CFTR in order to restore its function. Unfortunately, since CF is a rare disease and there are more than 400 mutations which cause it, there is insufficient clinical data to determine which patients will respond to modulator therapy. This leaves many patients with rare mutations unable to access these life changing drugs.

In this work, we used extensive molecular dynamics (MD) and free energy calculations to study how CFTR works and characterise the numerous ways it can malfunction. This was done in close collaboration with *in vitro* and clinical experiments, in order to understand what types of molecular defects may be treated by existing modulators with a process known as therotyping.

Our findings indicate that each CFTR mutation is largely unique, with each one exhibiting its own set of molecular interactions which cause pathogenesis. What is then remarkable, is that these different mutations all appear to respond to CFTR modulators—both *in vitro* and in some cases, in the clinic as well. This strongly indicates the possibility that a larger population of those afflicted with rare forms CF may respond to the right choice of modulator therapy. This physical understanding of the underlying cause of CF leads us to suggest that patients carrying missense mutations should be systematically theretyped, and the choice of modulators could be informed by a molecular understanding of the malfunction in CFTR.

In loving memory of Madeline Jennifer Dell

Fear cuts deeper than swords.

– Arya Stark [1]

Acknowledgment of Country

Before beginning, I'd like to acknowledge the traditional owners of the land where this research was conducted, to pay my respects to the elders of these lands, both past and present.

This research was produced while I lived and worked on the lands of the Gadigal people of the Eora Nation. The computer with which I performed simulations is named "Gadi". This means "to search for" in the language of Ngunnawal people on whose land this scientific equipment is located. This name is intended to pay homage to the continued traditions of learning and knowledge sharing which have been practiced by Aboriginal and Torres Strait Islander peoples for millennia.

Acknowledgment of People

Daniel Golestan, a wise man, once told me that to be given the opportunity to create this thesis was a gift. It was. It was a gift given to me by every friend, colleague, teacher, mentor and family member I've spent any time with. This list of thanks is by nature incomplete. If it was you'd be reading about a conversation I had with a middle aged public service woman in a hostel north of San Francisco, but that has little to do with Cystic Fibrosis. Since they would be rightfully offended if I forgot them, I will start with my supervisors.

To Shafagh Waters. For your vision, your drive, and all your advice. You brought me a truly fascinating PhD project and I benefited greatly from your mentorship. Full of helpful suggestions and ambition, you really helped me figure out *how* to do this kind of research and your approach to science has made me more confident and more capable.

To Serdar, a brilliant mind and a patient boss. Thank you for giving me the best possible experience at grad school I could have asked for during the pandemic. Your willingness to let me pursue self directed projects with a guided hand is a privilege during a PhD and I'm all the better for having gotten it from one of the best. I'm excited to carry some of your deep physical insight into biological systems to future research projects.

To Poker Chen. I am a better human being in every conceivable way for having known you. Your wisdom, intelligence and kindness are boundless. You have taught me an inordinate number of things. And yes, I do mean inordinate.

To Jeffry Setiadi. You got me started with these weird and wacky simulations. Thank you for your continued tutelage and patience, even after you crossed the pacific ocean.

To Renate Griffith for all your input, even from Tasmania, as well as Katelin Allen, Sharon Wong, Laura Fawcett, and the rest of the Waters lab. For teaching me so much chemistry and cell biology, and helping me write manuscripts. Bridging the gap between cell biology and molecular physics is something that will happen more in the future and I'm lucky to have met such a dedicated lab to teach me to do so.

To the patients and their families who participated in the studies designed by the

Waters lab, and also the support of the Cystic Australia Research Trust and the Peter Domachuck Memorial Scholarship. I hope there is some day a cure for Cystic Fibrosis and that our work might be a small step toward that goal.

To my parents. You raised me with not only academic rigor in mind but also a respect for the arts which has served me strangely well. The hours I spent rendering the images in this thesis I was thinking of you.

To Nonno and Nonna, I don't think you'll read this. I'm sad that you won't understand what I've done but I think you'd be proud if you did. Living in Condell park did more for me than you could know. Far from war-torn Beirut or dirt poor Orria, I'm writing this in a well lit office with a full stomach and few worries. Sometimes this luck makes my head spin.

Of course, I must also thank many friends and family for the inspiration and motivation they provided along the way. Zac, Alon, Harrison, Loretta, Melissa, Daniela, Chris, Josh, Georgio, Markus, Joseph, Deb, Ollie, Harry, Ashwin, Calida, Tilly, Helena, Hway, David, Benjamin, Nicholas, Alex, Amy, Gareth, Shirley, Nick, Natasha, Aleksa, Frank and my dear sister Carla.

Finally. Maddy, I love you. I miss you every day. You couldn't have imagined what it was like to finish this after losing you. I carry so much of you with me and I wish I had more. Thank you for your intelligence, your warmth and your love, I miss them all terribly.

To all of you, thank you for your help along the way. You're all in my Loop [2] and I hope I'm in some of yours.

Permission for the Inclusion of Published Work

Dr Shafagh Waters
 Translational Medicine
 School of Biomedical Sciences
 University of New South Wales
 Sydney, Australia



9 Sep. 22

Miro's PhD work was interdisciplinary. He performed extensive simulations MD in silico to make a substantial contribution to our understanding of CF pathogenesis. Chapters 4 and 5 of his PhD thesis have been peer reviewed and published in *IScience* and *The American Journal of Respiratory Cell and Molecular Biology* respectively. While chapter 6 is in preparation for submission to *Frontiers in Pediatrics*.

- (Chapter 4) Wong, Sharon L.[^], Nikhil T. Awatade[^], **Miro A. Astore**[^], Katelin M. Allan, Michael J. Carnell, Iveta Slapetova, Po-chia Chen, et al. "Molecular Dynamics and Functional Characterization of I37R-CFTR Lasso Mutation Provide Insights into Channel Gating Activity." *IScience* 25, no. 1 (January 2022): 103710. <https://doi.org/10.1016/j.isci.2021.103710>.
- (Chapter 5) Wong, Sharon L.[^], Nikhil T. Awatade[^], **Miro A. Astore**[^], Katelin M. Allan, Michael J. Carnell, Iveta Slapetova, Po-Chia Chen, et al. "Molecular Dynamics and Therotyping in Airway and Gut Organoids Reveal R352Q-CFTR Conductance Defect." *American Journal of Respiratory Cell and Molecular Biology*, April 2022. <https://doi.org/10.1165/rcmb.2021-0337OC>.
- (Chapter 6) Katelin M. Allan, **Miro A. Astore**, Laura Fawcett, Po-Chia Chen, Renate Griffith, Adam Jaffe, Serdar Kuyucak, Shafagh Waters. "Unique S945L-CFTR defect restored by CFTR modulator co-therapy in vitro correlates with in vivo biomarkers post-therapy" (In preparation for submission to *Frontiers in Pediatrics*)

[^] These authors contributed equally to this work.

For the in silico components, Po-chia Chen, Jeffry Setiadi, Renate Griffith and Serdar Kuyucak assisted Miro with the design and interpretation of his results and all authors on these publications assisted him with preparing the manuscripts for publication. Detailed descriptions of each author's contribution can be found in each publication.

As corresponding author on these publications, I give Miro permission to include content from these manuscripts in his PhD thesis.

I concur with A. Prof Serdar Kuyucak that Miro independently designed and executed the in silico components of each of these studies. This computational modelling contributed to our understanding of rare Cystic Fibrosis mutations and the function of the CFTR protein overall.

Kind regards,

A handwritten signature in black ink, appearing to read 'Shafagh Waters'.

Dr Shafagh Waters
 UNSW Scientia Senior Lecturer
 Head of Molecular Integrative Cystic Fibrosis (miCF) Research Laboratory,
 School of Biomedical Sciences, UNSW and Sydney Children's Hospital
 UNSW Early Career Academic Network (ECAN) Executive Committee and
 Co-Chair of UNSW Medicine ECAN
 E: shafagh.waters@unsw.edu.au | W:<https://research.unsw.edu.au>

Publication Authorship Attribution

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.



09/09/2022

Miro Alexander Astore, Student

Date

As the supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.



09/09/2022

Serdar Kuyucak, Supervisor

Date

Contents

List of Abbreviations	xiii
List of Figures	xiii
List of Tables	xvi
1 A Physical Introduction to Theoretical Biology	1
1.1 Thesis and Chapter Summary	1
1.2 What is Physics?	2
1.3 The Physics Inside your Cells	4
1.4 Using Ion Channels as Natural Laboratories to Study Biophysics	5
1.5 Studying Cystic Fibrosis; Toward a Molecular Theory of Disease.	8
1.6 The Future is Biological	10
2 From Protons to Proteins: Methods to Simulate the Inside of a Cell.	12
2.1 Quantum Mechanics is Not Tractable at the Scale of Biology.	12
2.1.1 A Full Quantum Mechanical Treatment	13
2.1.2 The Born-Oppenheimer approximation.	14
2.2 Classical MD; Molecular Motions Without Quantum Mechanics	15
2.2.1 Philosophy of Different Molecular Mechanics forcefields.	18
2.3 Periodic Boundaries To Pretend A Protein is Inside a Cell	21
2.4 Controlling the Temperature and Pressure in a Simulation	24
2.4.1 Hot and Cold with the Nosé-Hoover Thermostat	24
2.4.2 Under Pressure with the Parrinello-Rahman Barostat	27
2.5 The Process of Preparing an MD Simulation	28
2.6 Choosing an Appropriate Time Step	29
2.6.1 Verlet Leap-Frog Integration	30
2.7 Free Energy Calculations and Enhanced Sampling: Making Simulations More Useful	31
2.7.1 Umbrella Sampling	32
2.7.2 Metadynamics	37
2.8 Short Comings of Classical MD	38
2.8.1 The Problem with Forcefields	38
2.8.2 The Problem with Sampling	39
2.9 Conclusion	40
3 Review of Cystic Fibrosis and The Molecular Basis of its Treatment	43

3.1	Introduction	43
3.2	Clinical Realities of Living with Cystic Fibrosis	45
3.3	The Misfunction of CFTR Causes Cystic Fibrosis	47
3.3.1	CFTR structure and function	48
3.3.2	Previous MD Studies of CFTR	49
3.3.3	Catalytic and Conformational Changes During the Gating Cycle	52
3.3.4	Anion Selectivity	54
3.3.5	Different classes of Cystic Fibrosis Phenotype	56
3.4	Modulators Act Directly on CFTR to Restore its Function	60
3.4.1	Correctors	61
3.4.2	Potentiators	62
3.5	Patients with Rare Mutations Struggle to Access Modulator Therapy	64
3.6	Patient Derived Organoids: A Pre-Clinical Model to Assess Modulators	65
3.6.1	Forskolin Induced Swelling to Assess CFTR Function	66
3.6.2	Electrophysiology Directly Measures CFTR Gating and Conduction	67
3.6.3	Western Blotting Assesses CFTR Trafficking	69
3.7	Conclusion	71
4	Molecular Dynamics and Functional Characterization of I37R-CFTR Lasso Mutation Provide Insights into Channel Gating Activity	72
4.1	Introduction	73
4.2	Results	75
4.2.1	I37R-CFTR Baseline Activity in Patient-Derived Rectal Biopsies and Intestinal Organoids	75
4.2.2	I37R-CFTR Functional Response to CFTR Modulator Monotherapy in Intestinal Organoids	78
4.2.3	I37R-CFTR Perturbs the Noose Structure of the Lasso Motif	82
4.2.4	I37R Mutation Strengthens Lasso Motif Interaction with the R-domain	83
4.3	Discussion	85
4.4	Supplementary Information	89
4.5	Method Details	94
4.6	Acknowledgements	98
4.7	Author Contributions	98
4.8	Declaration of Interests	98
5	Molecular Dynamics and Therotyping in Airway and Gut Organoids Reveal R352Q-CFTR Conductance Defect	99
5.1	Introduction	100
5.2	Results	101
5.2.1	R352Q-CFTR Baseline Activity and Response to CFTR Modulators in Nasal Epithelial Cells	101
5.2.2	R352Q-CFTR Baseline Activity and Response to CFTR Modulators in Intestinal Organoids	105
5.2.3	Correlation of CFTR Modulator Response in hNECs with Participant-matched Intestinal Organoid Swelling Assay Outcomes	108

5.2.4	Maturation of R352Q-CFTR in Nasal Epithelial Cells and Intestinal Organoids	109
5.2.5	Molecular Dynamics Simulations and Free Energy Calculations Reveal the Dysfunction of the R352Q Mutant	110
5.3	Discussion	113
5.4	Supplemental Information	116
5.5	Funding Disclosure	116
5.6	Author Contributions	117
6	Unique S945L-CFTR defect Restored by CFTR Modulator Co-Therapy In Vitro, Which Correlates with In Vivo Biomarkers Post-Therapy	118
6.1	Introduction	119
6.2	Results	121
6.2.1	In silico characterisation of the Misfunction of S945L-CFTR	121
6.3	Discussion	123
6.4	Supplementary Information	125
6.5	Method Details	126
6.5.1	Molecular Dynamics	126
6.5.2	Umbrella Sampling Protocols	127
7	Resolving a Conducting Conformation of CFTR Using Free Energy Calculations	128
7.1	Introduction	129
7.2	Results	132
7.2.1	Long Simulations Reveal Motions Which Dilate the CFTR Pore	132
7.2.2	OPES-Metadynamics Finds a Stable Conformation with an Open Extracellular Pore	133
7.2.3	The Open Conformation is Stabilised by the R334-E1126 Salt Bridge	135
7.2.4	Proving the Open Conformation is Capable of Ion Conduction with Umbrella Sampling	137
7.2.5	The Open Conformation Can be Used to Study Disease Causing Mutations in the Selectivity Filter, such as R334W	139
7.3	Discussion	139
7.3.1	The Dilated Structure Is Consistent with Experiments	141
7.3.2	Study Limitations	141
7.3.3	Proposed Experiments to Test the R334-E1126 Salt Bridge	143
7.3.4	The Open Conformation is Important in the Study of Rare Mutations	144
7.4	Conclusion	145
7.5	Methods Details	145
7.5.1	System Composition	145
7.5.2	Unbiased MD	146
7.5.3	Principal Component Analysis and Choice of Collective Variables	146
7.5.4	OPES-MetaD	148
7.5.5	Assessing the Stability of the Dilated Conformation	149
7.5.6	Umbrella Sampling	149

8 A Physical Perspective on The Rescue of Mutant CFTR	153
8.1 Introduction	153
8.2 Summary Of Rare Mutations Studied so Far	154
8.3 Simulating the Common G551D Gating Mutation	156
8.4 The Many Molecular Modes of Misfunction in CFTR	157
8.5 A Physics Motivated Approach to Personalised Medicine in Cystic Fibrosis	159
8.6 Therotyping Q1291H-CFTR With Atomic Precision	161
8.7 Summary	163
8.8 Method Details	163
9 Conclusion: Where Next for Molecular Studies of CFTR	164
9.1 Summary: The Three Categories for Future CFTR Studies	164
9.2 Outstanding Basic Questions of CFTR Function	165
9.3 Mechanistic Understanding of Mutations	166
9.4 Elucidating Modulator Action on CFTR	166
10 Afterword	169
Bibliography	173

List of Figures

1	Nerve Cross Section by David Goodsell	xviii
2	The Position That Biology Occupies Compared to the Rest of Physics	xix
1.1	The Structure of DNA has Periodic and Aperiodic Elements	3
1.2	The Action Potential is a Solution to the Hodgkin-Huxley Model	6
1.3	Different Ion Channels Amenable to Molecular Simulation	7
1.4	A Physical Model to Understand Protein Evolution	10
2.1	The Bonded Interactions Calculated In Classical Forcefields	17
2.2	Comparison Between Potentials in Quantum and Classical Forcefields	19
2.3	The Lennard-Jones Potential	20
2.4	An Example of a Solvated Biomolecular Environment Ready for Simulation	22
2.5	Particle Mesh Ewald Summation	23
2.6	Enhanced Sampling Allows Access to More Protein States	31
2.7	Illustration of Umbrella Sampling	34
2.8	Illustration of Metadynamics	36
3.1	Cystic Fibrosis is a Debilitating Disease Whose Cause is Genetic	45
3.2	CF Clinical Progress	47
3.3	CFTR Structure	48
3.4	NBD Organisation	52
3.5	Chloride Passage through CFTR	55
3.6	Rare Mutations Occur Across the CFTR Protein	57
3.7	Modulators Bound to CFTR	60
3.8	Forskolin Induced Swelling	66
3.9	Patch Clamp Apparatus	68
3.10	Ussing Chamber Setup	69
3.11	Western Blot Explanation	70
4.1	Graphical Abstract. Integration of in silico and in vitro experiments for personalised medicine	73
4.2	Characterization of I37R-CFTR residual function in rectal biopsies and intestinal organoids	76
4.3	Characterization of I37R-CFTR functional response to corrector or potentiator monotherapy in intestinal organoids	79
4.4	Characterization of I37R-CFTR functional response to dual potentiator or corrector therapy or corrector(s)-potentiator(s) co-therapy in intestinal organoids	81

4.5 Placement of I37R within the lasso motif and the resulting changes to its conformation	82
4.6 I37R interacts with a previously unresolved section of the R domain . .	84
4.S3 Comparison between the stability of the lasso motif in I37R-CFTR and WT- CFTR at 310K and 350K.	89
4.S4 Comparison of the stability of different WT-CFTR domains at 310K and 350K.	90
4.S5 Comparison of the stability of the bilayer surrounding WT-CFTR at 310K and 350K	91
4.S6 Results from All Tested R-domain Models	92
4.S7 A comparison between the prediction of the structure of human CFTR by AlphaFold2 and the modelled segment of the R domain	93
4.S8 Comparison between the stability of the transmembrane domains in I37R-CFTR and WT-CFTR at 310K.	93
5.1 Functional response of R352Q-CFTR to cystic fibrosis transmembrane conductance regulator (CFTR) modulators in differentiated human nasal epithelial cells (hNECs).	103
5.2 Functional response of R352Q-CFTR to CFTR modulators in intestinal organoids.	106
5.3 Correlation between individual in vitro treatment-modulated CFTR function in participant-matched differentiated hNECs and intestinal organoids.	109
5.4 Simulations provide a mechanistic understanding of R352Q-CFTR dysfunction	111
5.E4 Stability of Key Salt Bridges within the CFTR Channel	116
6.1 S945L	122
6.S1 Detailed View of the Hydrogen Bond Netowrk in WT-CFTR	125
6.S2 Simulations of WT-CFTR and S945L-CFTR at Physiological Temperature	125
6.S3 Local Conformational Change of the Elbow Between TM8 and the R-domain.	126
7.1 Constriction in Cryo-EM Structure (6MSM)	130
7.2 Principal Component Analysis Reveals Large Motions in CFTR Transmembrane Helices	133
7.3 Free Energy Calculations Discover a Dilated Conformation of CFTR. .	134
7.4 The Dilated Conformation of CFTR Gives Rise to the R334-E1126 Salt Bridge	136
7.5 The Dilated Conformation of CFTR is Capable of Conducting Anions .	138
7.6 R334W Inhibits the Conduction of Anions	140
7.7 The Alpha Carbon Atoms Manipulated to Dilate CFTR	148
7.8 OPES MetaD Free Energy Surfaces 400ns-900ns	150
7.9 OPES MetaD Free Energy Surfaces 1000ns-1500ns	151
7.10 Demonstration of OPES-MetaD Convergence	152
8.1 Summary of our Rare Mutation Studies	154
8.2 Pathogenesis in G551D-CFTR	156

8.3	Granular grouping of CF pathogenesis	158
8.4	Conceptual Framework for the Pathogenesis of Mutations and The Action of Potentiators	160
8.5	Pathogenesis in Q1291H-CFTR	162
9.1	CFTR Has Many Proposed Drug Binding Sites	168

List of Tables

2.1	Timescales of Motions in a Molecular System	30
2.2	Classification of Common Free Energy Techniques	33
7.1	Amino acids used to analyse and manipulate the outer pore of CFTR .	147

Foreword

I've never held a pipette

The more I wrote of this thesis the more I found myself writing down things I wish I'd known when I first started studying biophysics. I was writing for my past self. So I've tried my best to write each chapter to be understandable to those with a loose grasp of undergraduate physics. The most difficult thing for this audience will not be the mathematics or technical content herein, but rather the breadth of knowledge from different fields to understand the scope of the contents. I have not had time to write an introduction to molecular biology, so I recommend a physics based introduction to those concepts such as those found in "The Physical Biology of the Cell" [3]. For more intensive biology focussed introduction there is "Cell Biology" [4]. In this spirit of trying to give the reader an interdisciplinary roadmap, I have taken some care to name certain authors as waypoints, to give the reader anchors to help them navigate the literature. Similarly, when a technical concept is mentioned, I have searched for a useful review article, book or explainer. So please take such citations as reading recommendations [5, 6].

When I first started performing simulations of living systems, I didn't even know what a protein was, and as I'm writing this, I have still never taken a formal course in biology or chemistry. Despite this, for the past 4 years I have been captivated by the seemingly infinite complexity in biological systems. For some examples, please take this opportunity to treat yourself to the fabulous work of David Goodsell in Figure 1.¹

Overall, I've found that the mindset for solving biological problems feels very different to the narrowed focus which we cultivate in students to study idealised problems in mathematics and physics. The problems are more specific and the solutions are often less general. For example, plasma physicists may use the same mathematical tools to describe a diverse set of materials, from the dense stellar core, to the sparse intergalactic nebulae. The density of these objects spans 28 orders of magnitude [11].² Would that we were so lucky in biology, where we struggle to apply same physical models to deal with phenomenon across a single order of magnitude. All the same, it seems increasingly clear to me that living things as we understand them occupy a special place in the universe (Figure 2).

¹<https://pdb101.rcsb.org/sci-art/goodsell-gallery>

²From $10^{34} - 10^6$ charged particles per meter cubed respectively.

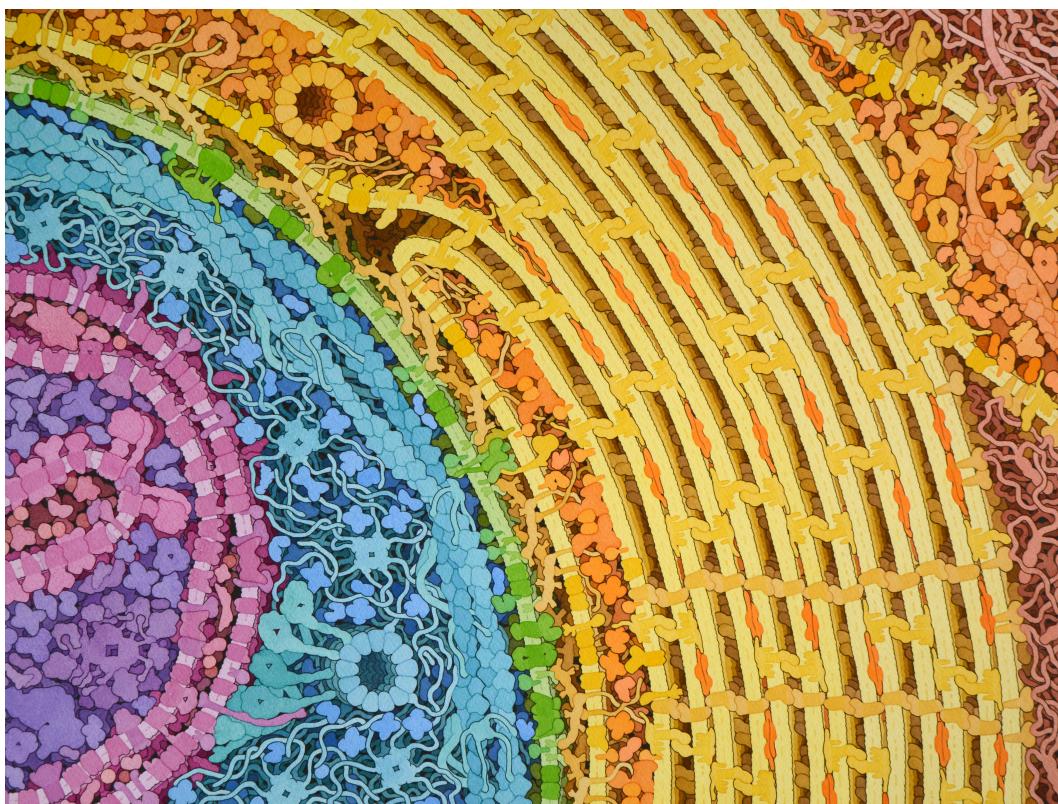


Figure 1: Nerve Cross Section by David Goodsell

David Goodsell is an artist who produces water paintings of cellular environments. Many of his works can be downloaded and used [for free](#). He has also written many books and articles which would serve as a light, layperson friendly introduction to molecular biology [7–9]. This particular painting depicts the cross section of a nerve. The nerve fibre (blue) is wrapped in an insulating myelin sheath (yellow). Electrical signals propagate perpendicular the page, via voltage gated sodium and potassium protein ion channels at the edge of the nerve[10]. We will discuss the discovery and details of this mechanism in the next chapter, as they were critical to the development of biophysics.

Biological systems are not homogeneous. If you look at your hand, you will notice hair, pores, dry skin, dead skin, perhaps even tendons and muscles twitching beneath a ghostly web of blood vessels. If you were to pluck a single cell from anywhere in this hierarchy and view it under a microscope, you would just be able to make out some of the organelles inside that cell. The size, shape and function of the organelles would vary depending on what kind of cell you happened to choose. Within and between each those organelles is a wet, salty dance of molecular machines called proteins. It is this kind of molecule which I have trained to study in detail with computational models. The length scale of this journey, from your arm to a single protein, spans 8 orders of magnitude.³ At each step along the way, if you searched the literature you would find an army of experts studying an array of phenomenon at each length scale.

Note for example how the publications arising from this thesis have many authors.

³From the size of your hand to the size a single protein spans from 10^{-1}m to 10^{-9}m respectively.

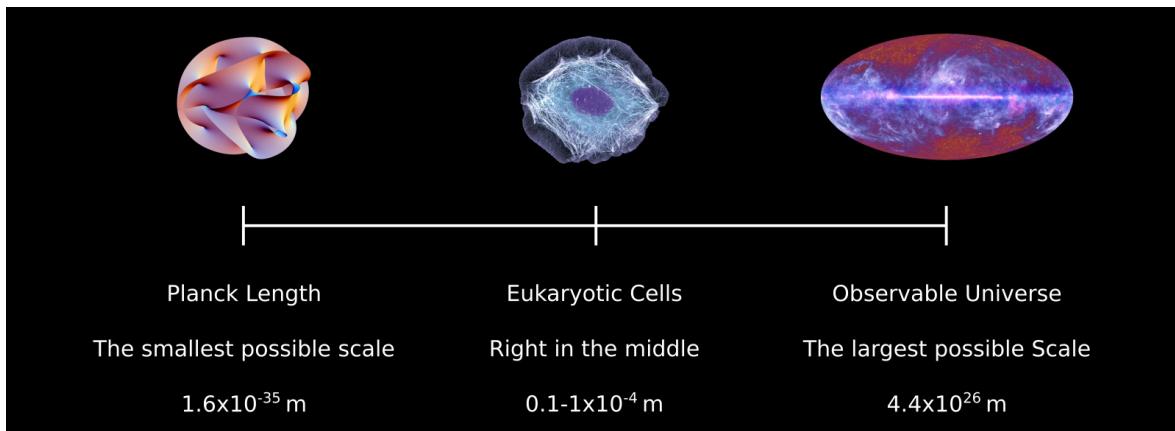


Figure 2: The Position That Biology Occupies Compared to the Rest of Physics
It just so happens that if you plot the size of everything in the universe on a log scale, eukaryotic cells fall right in the middle. A physicist can talk about both ends of this scale, we should learn something about the middle too.

Each researcher specialises, not unlike their cells, in a specific discipline. It is only by working together that can we hope to understand the whole organism.

If the reader is anything like myself they will find the amount of required knowledge to study biology a substantial barrier to entry. It is quite difficult at first to figure out which questions to ask or even which subfields to study to alleviate confusion. These challenges can be tackled by cultivating a broad coalition of connections—speak to physicians, clinicians, evolutionary biologists [5, 12], philosophers, molecular biologists, biochemists, cell biologists[4], geneticists, bioinformaticians, theoretical chemists, computer scientists, neuroscientists, physicists, mathematicians—everybody. It will take time but remain patient and you will find that a physics motivated approach can indeed explain and eventually even predict outcomes in biological experiments. A broad world view awaits you and it's really quite fun.

If this is indeed read by a future trainee, I hope the physics focussed philosophy in the introduction of chapter 1 and the literature review of simulation techniques in chapter 2 can serve as a road map—but a physicist studying biology will also be served well by nurturing a strong base in electrodynamics and statistical mechanics [13–15]. One particularly thorny issue for such a reader is that the field is now progressing so quickly that I'm sure much of this thesis will be out of date by the time it's read by anybody I'd hope to train. But really, that's just one of the things that makes biophysics so exciting. Shoot me an email if you want help miro.astore@gmail.com. Hopefully I'm still in the biophysics game.

The task ahead is immense. We're still not exactly sure how many unique proteins are expressed by the human genome, but it's upward of 20 thousand [16]. By contrast, this thesis represents an all consuming effort by a single PhD student—running decades worth of simulations to make incremental progress in the study of just one of these genes. There is so much to do. Good luck, I suggest you pack a towel [17].

1

A Physical Introduction to Theoretical Biology

Whatever “complexity” means, most people agree that biological systems have it.

– Frauenfelder and Wolynes [18]

1.1 Thesis and Chapter Summary

This thesis seeks to apply a philosophy of molecular biophysics, to demonstrate its capability to investigate pressing problems in biology and medicine. In particular, we will use molecular dynamics (MD) to look at how a specific gene misfunctions to cause disease. The disease in question is Cystic Fibrosis (CF), and it is caused by the malfunction of a protein called the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). These MD techniques will allow us to formulate a model of CFTR’s malfunction which we hope will direct research efforts and allow more patients suffering from CF to access life saving medication.

In this first short chapter we will quickly build a philosophy which outlines how to look at biology through the lens of a physicist. To do this we will first outline what the goal of a physicist is: to create abstract formalisms which can be used to model the natural world. We will then observe what makes the construction of such formalisms so difficult for biological problems.

These difficulties have often led biophysicists to study ion channels. Because of their simplicity, and their importance to cellular function, these molecules have served as laboratories to understand more complex biological systems, such as whole cells or organisms.

Chapter 2 will describe the chemical and numerical simulation techniques we have used to study the CFTR protein system in detail, while chapter 3 gives an overview of the CFTR system itself, and also a set of *in vitro* assays which compliment our computational modelling. Chapters 4, 5, 6 and 7 demonstrate the details of how a diverse application of the simulation techniques in chapter 2 can be used to discover the

unique modes of malfunction in CFTR. In combination with *in vitro* cellular techniques, these simulation results prove that these existing drug regimens can rescue many kinds of molecular defect. Finally, chapter 8 ties together these results to argue for a physical model which elucidates the mechanism of action for cystic fibrosis drugs. Armed with this model we will work through the available literature and identify some priorities for future studies using molecular modelling for cystic fibrosis research.

We hope this small example can demonstrate the utility of physics expertise for the field of molecular medicine. We anticipate that such methods will only grow in power with improvements in computational and experimental techniques.

1.2 What is Physics?

Before I started university I described physics as “the study of how things move”. Although intuitive, this description does not shed light on the philosophy of doing physics which make it such a powerful tool for understanding the natural world. When we create a predictive physical theory, we first carefully define a formalism motivated by patterns we see in the world around us. Then, using mathematics, the implications of this formalism are built up to make predictions about measurable phenomena. Should the predictions from the formalism agree with experimental results, it validates the formalism, giving rise to a physical theory. This is what makes physics feel like the most “fundamental” of the natural sciences.

These formalisms make use of many mathematical objects. Examples you might be familiar with include:

- The gas of hard spheres, which we use to derive the Boltzmann’s kinetic theory of gasses.

$$\frac{\partial f_1}{\partial t} + \frac{\mathbf{F}}{m} \cdot \nabla_{\mathbf{r}} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f_1 = \int \int (f'_1 f'_2 - f_1 f_2) \tilde{\kappa} d\tilde{\kappa} d\mathbf{v}_2. \quad (1.1)$$

- The interacting electric and magnetic vector fields in Maxwell’s laws of electromagnetism [13].

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} & \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} & \nabla \cdot \mathbf{B} &= \mu_0 (\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}). \end{aligned} \quad (1.2)$$

- The Riemannian manifolds which define the curvature of spacetime in Einstein’s theories of relativity [19].

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}. \quad (1.3)$$

- The complex probability waves which evolve according to Schröedinger’s equation, describing quantum mechanics [20].

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = \hat{H} |\psi(t)\rangle. \quad (1.4)$$

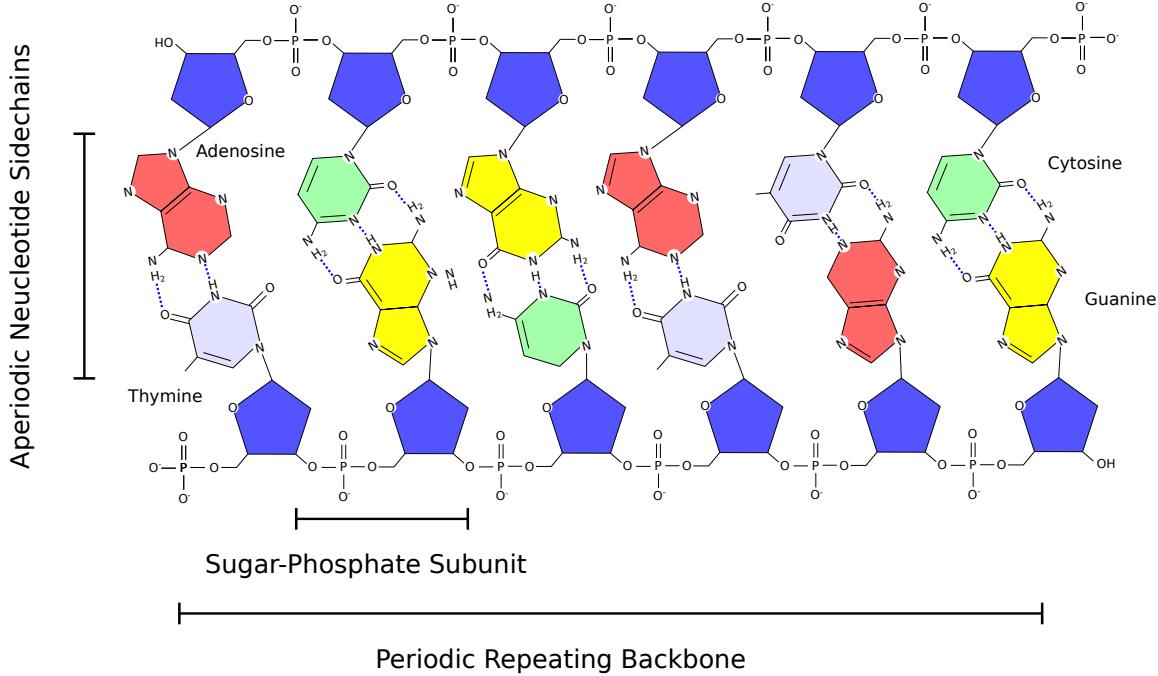


Figure 1.1: The Structure of DNA has Periodic and Aperiodic Elements

Although not a crystal, the structure of DNA has some periodic and some aperiodic elements. This is reminiscent of Schrödinger's speculation that genetic information is chemically encoded in an aperiodic crystal.

As biophysicists we would wish to find the most basic formalisms for biological phenomenon, so we can fully understand the function of an organism. The quest for such a formalism has an interesting origin. In 1944 Erwin Schrödinger wrote an essay titled "What is Life?" This remarkable work was written before the discovery of DNA's structure or the maturation of information theory. Schrödinger used first principles in thermodynamics and quantum mechanics to speculate at the nature of life at the atomic level. The most remarkable thing about this essay is how much the author got right.

He observed that since organisms exist at temperatures on the order of 300 Kelvin, the physical encoding of genetic information inside cells must be chemical in nature—since at such temperatures the physical arrangements of atoms would be unstable without chemical bonds. By estimating the amount of information that could be stored in an arrangement of atoms, Schrödinger posits the existence of what he calls an "aperiodic crystal". Although not a crystal in the rigorous sense, the double helix structure of DNA is not far from such an analogy. As Figure 1.1 shows, the DNA double helix is a combination of periodic and aperiodic elements, not too different from how one might visualise a one-dimensional aperiodic crystal [21]. This allegory is an example of how physical principals can be used to direct questions in fundamental biology. In fact, James Watson credited Schrödinger's book as one of his influences in how he visualised the structure of the DNA before its discovery through X-ray crystallography [22].

In this example, Schrödinger naturally chose a formalism of interacting atoms, mostly

consistent with the statistical mechanics in equation 1.1, to arrive at his model of an aperiodic crystal. In this thesis we will use a more careful, but similar formalism to study the function of the CFTR protein. We will outline this formalism in detail in chapter 2. The goal of chapters 3-7 is then to collect evidence in order to develop a physics inspired model of for the treatment of Cystic Fibrosis disease, which we will propose in chapter 8. The model we arrive at is more abstract than physicists are used to, and the next section will explore why this is often the case for biological systems.

1.3 The Physics Inside your Cells

Why can't I write down an equation which will tell me how long I will live? Or how tall I will grow?

These might seem like odd questions but if you asked a physicist how much power it would take to ionise a gas or how long it will take a black hole to evaporate itself out of existence, they would have have a tool kit of highly accurate models to help them calculate solutions.

To understand why the first set of problems are so much more difficult for physicists, we will analyse the cases where physics fails. Any physical theory may fail for one of the following three reasons:

- The physicist does not have sufficient computational power to integrate the formalisms in the relevant theory, so they cannot make predictions about a specific phenomenon.
- The physicist lacks sufficient data, so they cannot specify reasonable initial conditions for the calculations with the formalism.
- The phenomenon in question lies outside the applicable energy or length of the physical theory. This would occur, for example, if we used Newton's theories of gravity to predict the motion of an object around a super-massive black hole—here we would instead need Einstein's general theories of relativity [23].

From the perspective of fundamental physics, quantitative theories of biology fail for one of the first two reasons. Given the energy and length scale of biological systems, we currently possess the necessary theoretical tools to simulate every interaction inside a living being [24].

The difficulty of studying biology then, does not arise from a lack of physical understanding. As we will see in chapter 2 the physics molecular interactions in living things is surprisingly simple. Rather, the difficulty of biological problems arises from the sheer number of interactions we must consider. Inside cells we find proteins, lipids, solvents, salts each with their own physical properties. Therefore, biophysics distinguishes itself from more traditional physics by considering systems that are both highly heterogeneous and anisotropic. This aspect of biological systems makes it difficult to scale up our formalisms to make predictions using human tractable mathematical tools—meanwhile, computational engines can only help so much. Hence, we have to be careful about which systems we study with our available toolkit. When their computational

or theoretical tools were lacking, physicists often turned to ion channels as laboratories to develop and test biophysical theories [25, 26].

1.4 Using Ion Channels as Natural Laboratories to Study Biophysics

Ion channels are a special kind of protein which allow the passage of charged particles to pass across the membrane of a cell. They are excellent laboratories for the study of biophysics for two reasons. Firstly, it is very easy to measure their activity with a technique called electrophysiology which we will briefly explore in chapter 3 [27]. Secondly, they are critical to the health and function of cells. As cell biology has advanced, it has become increasingly clear that the level of polarisation (potential difference from the inside to the outside) in a cell is critical to its function. Cell polarisation regulates many chemical reactions [28–31].¹

It is perhaps then not surprising but nonetheless remarkable that ion channels are the targets of 19% of clinically approved drugs [32]. However, there is much more work to be done, as candidate drugs are often insufficiently selective for the desired ion channel, leading to sometimes lethal side effects [33–35].

Historically, ion channels have served as a testing ground for biophysical models. The first interest in modelling their behaviour comes from the experiments of two biophysicists, Andrew Hodgkin and Alan Huxley. They threaded silver wires through the thick nerves of a giant squid and measured the current running through the nerve in response to electrical stimulation. What they found was intriguing. Signals would only propagate down the nerve when the input signal was of a sufficient high voltage. They managed to match their experimental data with a model comprised of the following set of ordinary differential equations:

$$\begin{aligned} I &= C_m \frac{dV}{dt} + \bar{g}_K n^4 (V - V_K) + \bar{g}_{Na} m^3 h (V - V_{Na}) + \bar{g}_l (V - V_l), \\ \frac{dn}{dt} &= \alpha_n(V)(1 - n) - \beta_n(V)n, \\ \frac{dm}{dt} &= \alpha_m(V)(1 - m) - \beta_m(V)m, \\ \frac{dh}{dt} &= \alpha_h(V)(1 - h) - \beta_h(V)h. \end{aligned} \tag{1.5}$$

Here, the n , m and $h \in [0, 1]$ parameters are associated with potassium channel subunit activation, sodium channel subunit activation, and sodium channel subunit inactivation.

¹A back of the envelope calculation reveals that a -100mV resting potential is a 2.5 kcal/mol barrier to the transport of molecular species with a charge of -1e into the cell. This is not an insignificant energy contribution.

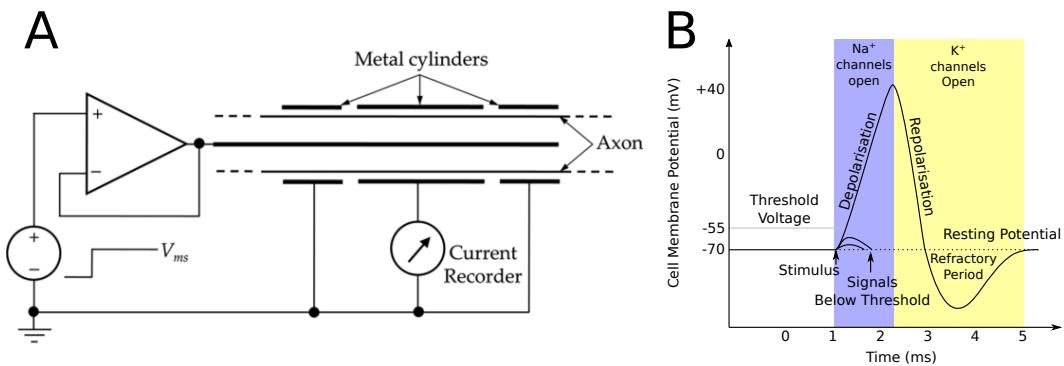


Figure 1.2: The Action Potential is a Solution to the Hodgkin-Huxley Model
A) The “voltage clamp” experimental set up which was used to study nerve impulses [37]. A giant squid nerve was submerged in sea water and an electrode was threaded through it as shown. The voltage was recorded as impulses were fired into the nerve. B) The shape of the action potential is a familiar sight in many physiology textbooks. The physical basis for its shape was realised through the mathematical modelling of Hodgkin and Huxley which won them the 1963 Nobel Prize in physiology or medicine. This discovery is an example of the physiological importance of ion channels and also how deep theoretical insight can lead to predictive models of living systems [38–42].

tion, respectively. C_m is the capacitance of the lipid membrane per unit area, and \bar{g}_i is the maximal conductance allowed across the membrane, per unit area. The terms V_i denote either the total voltage, or the contribution to the total from a specific charged species.

The solutions to the Hodgkin-Huxley model allow us to mathematically discover and describe several important cellular functions. The model encodes both the existence of a cell’s resting potential and the behaviour of selective voltage-gated ion channels. Even today, the molecular mechanisms behind these discoveries are used to understand protein and cellular function [36]. This curve created by the action potential is a common sight in many physiology textbooks, and it is in fact a solution to this coupled set off ODES [4].

The equations in the model above are an example of the development of a mathematical formalism which is not fundamental in the same way as the equations at the beginning of section 1.2. But rather, it is built for an express purpose: to model the propagation of signals through a nerve. This model shows how quantitative thinking can lead to insights in biology. The sheer complexity of biology demands this of us. We cannot create complete theories so we must find useful formalisms for small domains of the problem space.

In this thesis we aim to do something similar. By building up from the fundamental physics outlined in chapter 2, we will build a model for the dysfunction of a single gene (CFTR), to understand a disease (CF). Again, we do not possess sufficient computational power to produce a complete physical model of CF, so we will have to settle for a qualitative model, which we will outline in chapter 8.

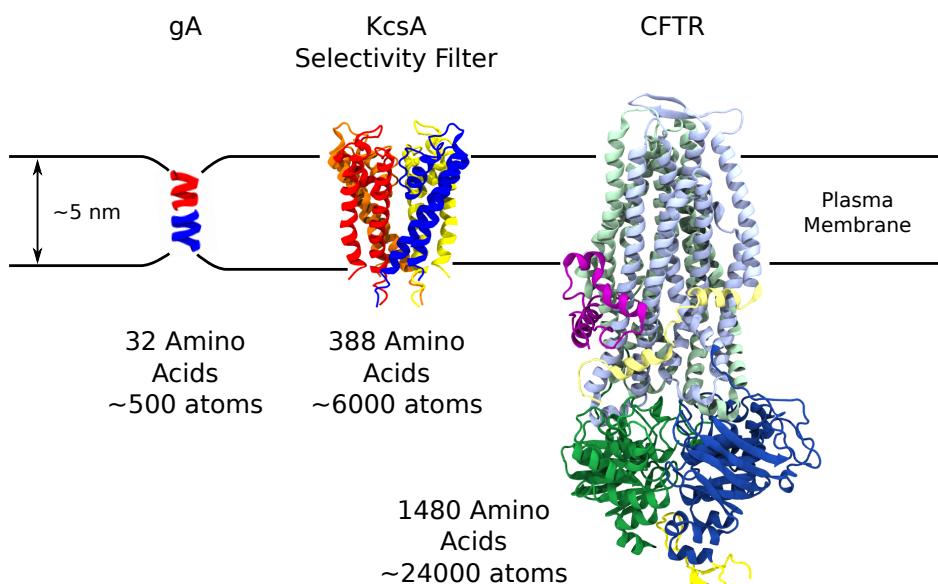


Figure 1.3: Different Ion Channels Amenable to Molecular Simulation

Gramicidin A was initially used as a toy model to test different *in silico* modelling techniques (PDB ID 1NT5) [43]. KcsA, is a bacterial potassium channel. This structure only comprises the pore domain, also called the selectivity filter (PDB ID 1BL8) [44]. This structure drew interest because potassium channels are play an important role in the function of cells. Finally, the CFTR anion channel which this thesis is written about (PDB ID 6MSM) [45]. In the course of 30 years that we have gone from simulating just a few nanoseconds of Gramicidin A to collecting almost half a millisecond of data for this thesis. This heralds an exciting future for computational biophysics [46]. What is even more exciting will be the next 30 years. At the time of writing, a computational engine named Anton 3 has come online. This purpose built computer could perform all the calculations in this thesis in less than a week [47, 48]. That's not a week in parallel, that's in serial!

Computer Modelling the Dynamics of Ion Channels

Motivated by the advances in cell biology that arose from the Hodgkin-Huxley model, early adopters of computational biophysics such as Martin Karplus, Benoît Roux, Klaus Schulten, J. Andrew McCammon, Shin-Ho Chung, Mark Sansom, Helmut Grubmüller, Serdar Kuyucak and Toby Allen devoted significant parts of their career to studying ion channels [46, 49–57].

These early studies usually focussed on gramicidin A (gA) because of its small size, its amenability to structural characterisation and the opportunity to compare experimental measurements with *in silico* calculations [58–61]. This small antibacterial peptide assembles into a coil which selectively allows the passage of potassium ions in the cell walls of gram-positive bacteria. This bleeds the bacteria internal potassium stores which are critical to cellular function [62]. For a long time, gA was the only protein ion channel with its structure determined to atomic resolution. Because ion channels are membrane proteins, it took further advancements in experimental structural biology techniques by Roderick Mackinnon and others to determine the atomistic structures of other channels [44, 63, 64].

The careful work on gA enabled the development of theoretical and computational methods which can now be transferred to the more biologically important channels, such as potassium and sodium ion channels [65–68]. With the “resolution revolution” in structural biology, driven to the maturation of cryo-EM structural biology and even more recently AI based structure predictions, we are entering an era of structural abundance—there are now literally thousands of biologically relevant targets which we can study with simulations [69–71].

In this new era, ion channels will likely continue to be important clinical and biophysical targets. In addition to the propagation of nerve signals we saw in section 1.4, they play a critical role in all 5 senses, allowing cells to react to pressure (and hence touch) [72], temperature [73], pain [74], acids [75] and smell (in insects) [76]. There is also an exciting subfield emerging which investigates the role of ion channels in the bioelectrical networks which regulate cell morphogenesis and growth [30, 31, 77, 78].

As we have shown, the historical goal of computational biophysicists has been to use whatever physical models and structural information is available, to understand living things at the molecular level (Figure 1.3) [79, 80]. In this thesis, we seek to continue this trend, by applying the fundamental physical models described in chapter 2 to study a disease, Cystic Fibrosis, at the molecular level.

1.5 Studying Cystic Fibrosis; Toward a Molecular Theory of Disease.

The sad truth of Cystic Fibrosis (CF) is that those afflicted are extremely unlucky. A small change to their genome causes their lungs fill with sticky mucus and become infected with bacteria—each breath becomes cumbersome [81]. Historically, diseases have been diagnosed based on symptoms and not causes [82]. As section 1.2 outlines,

this is antithetical to how we would like to study a physical process, even a biological one.

To build a predictive model, we would like to understand the root cause of a disease so we can predict how to treat it. Discovering this root cause can be extremely difficult and often requires decades of clinical enquiry [83, 84]. CF has the helpful characteristic of being a monogenic disease, so our molecular theory of this disease can begin by considering a single protein system and then slowly build outward, to consider other genes which effect the disease as well.

In this way, my motivations for studying the CFTR protein aren't solely focussed on treating disease. This problem is also an interesting opportunity to develop molecular theories of biology. This is what motivated the basic biophysical studies in chapter 7.

When understanding CFTR, a helpful perspective can be gained from the following model of protein evolution. It states that the stability of a protein's structure contributes to the overall fitness of an organisms by a formula [85]:

$$W(\Delta G) \propto \exp\left(\left[-\frac{\Delta G - \Delta G_{opt}}{\sigma_{\Delta G}}\right]^4\right) + c. \quad (1.6)$$

Here, W represents the evolutionary fitness of an organism, ΔG is the folding energy of the protein with a given gene sequence and ΔG_{opt} is the folding energy of the protein in the average (fit) population. The parameter $\sigma_{\Delta G}$ controls how broad this distribution will be and depends significantly on the protein physics of the gene. Figure 1.4 demonstrates the types of random walks of a gene through sequence space which this model predicts.

There are over 400 recorded mutations which cause CF [86]. This indicates that the CFTR gene exhibits an exceptionally small $\sigma_{\Delta G}$, so the band in the modified gaussian in figure 1.4a is very narrow. This means that CFTR sits at the precipice of a daunting cliff in sequence space. So, by taking small steps in sequence space and figuring out what factors have caused W to decrease, we can try to understand how we might push the needle of protein stability back into the optimal zone. Once we understand how we can do this for CFTR, we can use such thinking to build our methods outward and apply it to other diseases, such as sickle cell anemia, Alzheimer's disease, and Huntington's disease [85].

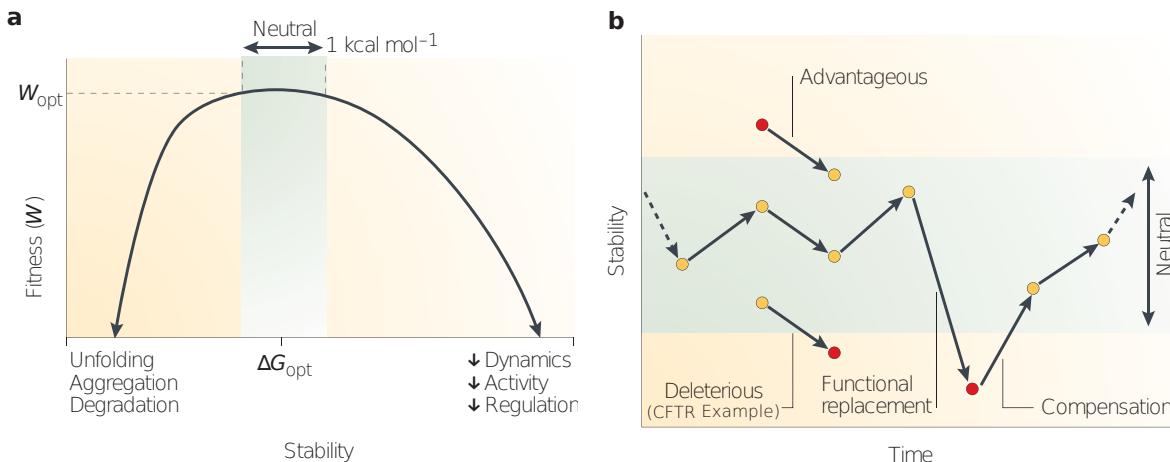


Figure 1.4: A Physical Model to Understand Protein Evolution

a) The stability of a protein can be measured by its folding free energy ΔG . This is related to the fitness of an organism W by equation 1.6. This model gives rise to a peaked distribution which helps us understand how so many mutants can give rise to cystic fibrosis. It appears as though the CFTR gene has an exceptionally narrow $\sigma_{\Delta G}$ and so small changes to the sequence of the gene have a comparatively dramatic effect on the evolutionary fitness of the organism. b) The model in equation 1.6 Gives rise to a random walk through in sequence space for subsequent generations. In the case of CF, it would appear as though *homo sapiens* are stuck at a specific snapshot in evolutionary time where CFTR may easily lose function. So, without gene therapies which can modify a gene sequence *in vivo*, we must use small molecule drugs and other medical interventions to somehow broaden the peak in equation 1.6.

CF is an excellent disease to begin such a project. Due to the array of disease causing mutations which occur across the CFTR protein, there is a large body of literature on its unique function [87]. This presents the opportunity to simultaneously perform basic biophysical research, while directly assisting in the improvement of patient outcomes. This is the sort of inquiry which drives basic science forward, combining interesting experimental data into theoretical models to make testable predictions. The aim of this thesis is to build a model to help direct efforts which will hopefully produce better patient outcomes. In chapter 8, we will combine evidence from the existing literature with our biochemical and simulation results to build our predictive model. It is hoped that this model can help direct future strategies to treat CF.

1.6 The Future is Biological

Biological sciences have important applications in medicine, agriculture and increasingly, manufacturing [88, 89]. We are in the midst of developing biology from a descriptive, to a predictive science [69, 90–93]. This transition is being driven by the combination of refined experimental techniques, rich datasets, strong theories and powerful computational engines. This is exciting because biological systems have happened upon ingenious solutions to some very difficult problems with the help evolution [5, 12]. It has much of the hard work for us, and as we understand it better, we can begin to apply its solutions to our own problems [94–96].

Throughout science, the integration of experimental data with theoretical models leads to new discoveries. In the case of biology, wet lab scientists take advantage of experimental techniques which allow them to understand the structure and dynamics of living things from the top down. The finer the experimental instrument, the finer the detail they may resolve. Conversely, computational and theoretical biologists take a bottom up approach. Like the physicist, they aim to take the granular details of a system and integrate them upwards, to model and predict the macroscopic behaviour of a system.

With more powerful computers and more detailed theoretical models, theorists can make predictions about the behaviour of more complex systems. What is so exciting about the current era of biological research is that the domains of these two approaches are beginning to overlap, where they can synergize and drive further breakthroughs [88].

This has been happening in physics for decades, resulting in world changing breakthroughs like telecommunications, nuclear power and transistors [97]. The reason this has been possible for physicists is that the systems they usually study are homogeneous. They include just a few, albeit sometimes complex components. Hence, it is much easier to integrate a theoretical formalism upward and predict measurable phenomenon.

The difference with biological systems is that they have so many different components. Hence, finding an analytic or even computationally tractable solution is usually impossible. However, as we collect more data with increasingly advanced experimental techniques² and build more powerful computers, more biological problems will become tractable. These, in turn, inform more detailed theoretical techniques which help direct the material efforts of experimental studies.

Onward and Upward

This introductory chapter has given us an overview of the philosophy behind physical biology, how we intend to apply it to study CF, and the exciting advances we can expect from future developments in biophysics. It is this same structure that we will use throughout the thesis. Chapter 2 will use our philosophy of physics to take a formalism of interacting atoms and integrate it upward to model biomolecular systems. We will then collect a large number of molecular details about the molecular cause and treatment of CF in chapter 3. This will give us the starting point for our own contributions to the study of CF in chapters 4, 5, 6 and 7. These results will allow us to collect evidence to help formulate the molecular model in 8 to understand the treatment of CF. With this roadmap, we will now begin our journey upward in chapter 2, by diving deep into the molecular dynamics methods which will allow us to model the molecular basis of Cystic Fibrosis.

²Important biophysical techniques one will encounter often in the literature are cryogenic electron microscopy (Cryo-EM) [98–100], electrophysiology [36], nuclear magnetic resonance (NMR) spectroscopy [101], confocal and fluorescence microscopy [102], X-ray Crystallography [103, 104], and genetic engineering (explanations of CRISPR-Cas9 or inverse PCR based techniques can be found in [105] and [106], respectively)

2

From Protons to Proteins: Methods to Simulate the Inside of a Cell.

Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.

– Richard P. Feynman [107]

This chapter is written for somebody who has studied undergraduate physics and now wishes to model biological systems at the molecular level. Care is taken to dive deeper into the mathematical formulations of simulation methods than is conventionally given in introductory texts. Essentially, this is the understanding of simulation techniques I wish I had had when I started studying them. An excellent overview which I would recommend as first reading for any new student can be found in an article by Braun et al. [108]. Other important topics like statistical rigor in simulations are covered by Gapsys et al. [109] or Grossfield et al. [110], while theory and practice in free energy calculations are covered by Pohorille et al. [111] or Chipot [112].

2.1 Quantum Mechanics is Not Tractable at the Scale of Biology.

Living things are made of atoms, and atoms themselves are composed of many particles—protons, neutrons and electrons. These constituent particles are governed by quantum mechanics and unfortunately, performing simulations for the number of atoms involved in proteins and other cellular components with quantum mechanical accuracy is impossible. Therefore, we will show how to take the fundamental formulation of atomic interactions in the Schrödinger wave equation and apply approximations in order to produce a model which is capable of simulating macromolecular systems at biologically relevant timescales.

We will gradually integrate upwards. Beginning with the interactions in a single molecule we will build our way up to a complex macromolecular system with lipids,

water, salts and of course, proteins. Ultimately this section rationalises the treatment of atoms as point charges in classical molecular dynamics simulations.

2.1.1 A Full Quantum Mechanical Treatment

Since we are dealing with atoms, which are governed by quantum mechanics we must begin our journey upwards with the time dependent form of the Schrödinger wave equation

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{x}, t) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}, t) \right] \Psi(\mathbf{x}, t). \quad (2.1)$$

In quantum systems we treat all particles as waves hence the use of the wave function $\Psi(\mathbf{x}, t)$. The complex amplitude of the wave function $|\Psi(\mathbf{x}, t)|^2$ tells us the likelihood of detecting the particle at time t and at position \mathbf{x} . The terms in brackets correspond to the kinetic energy of the particle ($-\frac{\hbar^2}{2m} \nabla^2$) and any externally applied potentials to the system ($V(\mathbf{x}, t)$) respectively, while m is the mass of the particle. Given that the left hand term $i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{x}, t)$ contains a gradient with respect to time, it governs how the wave function will evolve in time.

When the external potential V has no explicit dependence on time, this equation reduces to the familiar time independent form

$$E\Psi(\mathbf{x}, t) = \left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{x}) \right] \Psi(\mathbf{x}, t) = H\Psi(\mathbf{x}, t). \quad (2.2)$$

Here, E is an eigenvalue of the Hamiltonian operator H . Note that the wave function $\Psi(\mathbf{x}, t)$ is still allowed to evolve in time.

In atomic systems there are two types of particles—nuclei, which we will denote with the subscript n and electrons, which will be denoted by e . In order to treat these elements separately we decompose the Hamiltonian of the system into a few components

$$H = \underbrace{T_n + U_{n-n}}_{H_n} + \underbrace{T_e + U_{e-e} + U_{n-e}}_{H_e}, \quad (2.3)$$

where T_n and T_e denote the kinetic energy of the nuclei and electrons respectively. While U_{n-n} , U_{n-e} , U_{e-e} denote the potential energy for interactions between nuclei, between electrons and nuclei, and between electrons respectively.

Since the potential terms all describe charged species, they follow Coulomb's law and have the form

$$U_{n-n} = \sum_{i>j} \frac{q_e^2 z_i z_j}{|\mathbf{R}_i - \mathbf{R}_j|}, \quad U_{n-e} = - \sum_{i,l} \frac{q_e^2 z_i}{|\mathbf{r}_l - \mathbf{R}_i|}, \quad U_{e-e} = \sum_{l>k} \frac{q_e^2}{|\mathbf{r}_l - \mathbf{r}_k|}. \quad (2.4)$$

Here, the z_i represent the atomic number (and thus the charge) of the i th nucleus and q_e is the unit charge of the electron. The reason for the separate coordinates R_i and r_l is to separate out the treatment of nuclei and electrons, which will be important once we apply the Born-Oppenheimer approximation in the next section.

Meanwhile, the kinetic energy terms are of the form

$$T_n = - \sum_i \frac{\hbar^2}{2M_i} \nabla_i^2, \quad T_e = - \sum_l \frac{\hbar^2}{2m_e} \nabla_l^2, \quad (2.5)$$

where M_i represents the mass of the i th nucleus and m_e represents the mass of an electron. The operator $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$. The separate subscripts i and l are due to the different coordinates which we use to denote the positions of the nuclei and the electrons. The reason for this will become clear when we derive the Born-Oppenheimer approximation, to separate the wave functions and treat them separately to reduce computational expense.

2.1.2 The Born-Oppenheimer approximation.

In order to reach the Born-Oppenheimer approximation, we start with the observation that electrons have a mass 3-4 orders of magnitude smaller than the nuclei. This motivates two simplifications. The “clamped nuclei assumption” where we solve the Schrödinger equation whilst nuclei are fixed in space and do not move. And a related assumption known as the “adiabatic assumption” which postulates that the electrons will respond instantaneously to any changes in the positions of the nuclei. Combining these physical approximations we derive the “Born-Oppenheimer approximation” for the Schrödinger equation which can be used to simplify calculations involving several atoms at once.

We begin the derivation by examining the time-independent form of the electronic Schrödinger wave equation where the nuclei are fixed at positions R_i

$$H_e(\mathbf{r}_l, \mathbf{R}_i)\psi_e(\mathbf{r}_l, \mathbf{R}_i) = U_e(\mathbf{R}_i)\psi_e(\mathbf{r}_l, \mathbf{R}_i). \quad (2.6)$$

Fixing the nuclei in this way gives the “clamped nuclei” approximation [113]. To solve the wave function for the whole system Ψ_{tot} , we use an *ansatz* which decomposes the wave function with an electronic basis into two components: $(\psi_e)_k$ and $(\psi_n)_k$ which are the k th eigenfunction solutions to H_e and H_n , respectively

$$\Psi_{tot}(\mathbf{r}_l, \mathbf{R}_i, t) = \sum_{k=0}^{\infty} \psi_e(\mathbf{r}_l, \mathbf{R}_i)_k \psi_n(\mathbf{R}_i)_k. \quad (2.7)$$

Note that there is an implied direct product between the wave functions $\psi_e(\mathbf{r}_l, \mathbf{R}_i)$ and $\psi_n(\mathbf{R}_i)$. When we substitute this expression into the full Schrödinger equation 2.1 we find the following expression for the k th nuclear eigenfunction [114]

$$i\hbar \frac{\partial}{\partial t} \psi_n(\mathbf{R}_i)_k = \left[-\sum_i \frac{\hbar^2}{2M_i} \nabla_i^2 + U_e(\mathbf{R}_i)_k \right] \psi_n(\mathbf{R}_i)_k + \sum_j C_{kj} \psi_n(\mathbf{R}_i)_j, \quad (2.8)$$

where we have coupled the electronic wave functions to each other with the operator

$$C_{kj} = \int (\psi_e)_k^* \left[\sum_i \frac{\hbar^2}{2M_i} \nabla_i^2 \right] (\psi_e)_j d\mathbf{r} + \frac{1}{M_i} \sum_i \left[\int (\psi_e)_k^* [-\hbar i \nabla_i] (\psi_e)_j d\mathbf{r} \right] [-\hbar i \nabla_i]. \quad (2.9)$$

Using the “adiabatic assumption” [114], the off-diagonal terms of C_{kj} can be set to 0 as they represent the interactions between the electrons and the nuclei. This completely decouples the wave function into two components

$$\Psi_{tot}(\mathbf{r}_l, \mathbf{R}_i, t) = \psi_e(\mathbf{r}_l, \mathbf{R}_i)_k \psi_n(\mathbf{R}_i, t)_k. \quad (2.10)$$

A further approximation is justified by ignoring the diagonal terms C_{kk} as they are 4 orders of magnitude smaller than the other terms in equation 2.8 [113].

We now write the Born-Oppenheimer approximated wave equation for an atomic system

$$i\hbar \frac{\partial}{\partial t} \psi_n(\mathbf{R}_i)_k = \left[-\sum_i \frac{\hbar^2}{2M_i} \nabla_i^2 + U_e(\mathbf{R}_i)_k \right] \psi_n(\mathbf{R}_i)_k. \quad (2.11)$$

The separation between atoms in a molecule is on the order of $10^{-10}m$, while the de Broglie wavelength of the nuclei at room temperature is on the order of $10^{-11}m$. Hence, we treat the nuclei as point particles at the scale of the full molecule. So, by rearranging equation 2.11 and taking the derivative with respect to time, we can see how to use Newton’s equations of motion to calculate the forces on the nuclei from the surrounding electric potential

$$M_i \ddot{\mathbf{R}}_i(t) = -\nabla_i U_e(\mathbf{R}_i). \quad (2.12)$$

By choosing an appropriate time-step, one can simply iteratively solve this equation of motion to understand the dynamics of an atomic system. The nuclei will move according to their relative positions to each other and the electron clouds will rearrange in response to that motion. There is no need to explicitly treat the electrons at all. This is sufficient accuracy to simulate the low energy motions of molecules such as the environment found in biological systems.

2.2 Classical MD; Molecular Motions Without Quantum Mechanics

Following the Born-Oppenheimer approximation there are Hartree-Fock methods and density functional theory (DFT) which further simplify Schrödinger’s equation. These

more sophisticated physical methods allow us to simulate the organisation of electron clouds around small molecules, finding broad applications in chemistry and materials science [115]. These methods are known as *ab initio* MD.

However, even with these approximations, simulating a large number of atoms is still not computationally tractable. State of the art DFT methods can only simulate on the order of 10^3 atoms [116] and scales as $O(N^3)$ [117]. This is not sufficient to simulate proteins and their surrounding solvation environment where the molecular system is usually on the order of $10^4 - 10^6$ atoms. So, we must use another round of approximations to reach the spatial and time scales necessary to simulate biological molecules. We do this by creating a set of mathematical functions to simplify the calculations further. Here we use a set of virtual springs and other simple models for the energetic interactions between atoms. This creates what's known as an effective potential.

The CHARMM effective potential employed in all simulations in this thesis is similar to those found in all-atom classical molecular dynamics forcefields. The same functional forms are used in other forcefields such as AMBER, GROMOS and OPLS but with different parameters and design philosophies [118].

This formulation gives us classical molecular dynamics, sometimes referred to as molecular mechanics (MM). The aim of the classical forcefields discussed here is to use *ab initio* MD as an initial target for approximation and then refine the model to better match certain experimental quantities. This is discussed in detail in section 2.2.1.

We split up the molecular mechanics potential into several components dealing with the energies from covalent bonds, including bond stretching, twisting and bending as well as contributions associated with the forces that atoms exert on each other when they are not bonded together

$$U_{CHARMM} = \underbrace{U_{LJ} + U_{coulomb}}_{U_{non-bonded}} + \underbrace{U_{bonds} + U_{angles} + U_{dihedrals} + U_{impropers}}_{U_{bonded}}. \quad (2.13)$$

Interestingly, the bonded terms may be reasonably approximated by simple harmonic functions, with an exception we will discuss shortly,

$$\begin{aligned} U_{bonded} = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{Urey-Bradley} k_u(r_{UB} - r_{UB_0})^2 \\ & + \sum_{dihedrals} k_\varphi(1 + \cos(n\varphi - \delta)) + \sum_{improper-dihedrals} k_\phi(\phi - \phi_0)^2. \end{aligned} \quad (2.14)$$

Here, the k_i terms correspond to the strength of the restraint for a parameter, while the 0 subscript denotes the equilibrium position for that parameter. Even though this formulation is quite simple, it has empirically been shown to be a reasonable approximation for the potential energy functions of quantum mechanics in covalently bonded chemical species. An example can be seen in Figure 2.2.

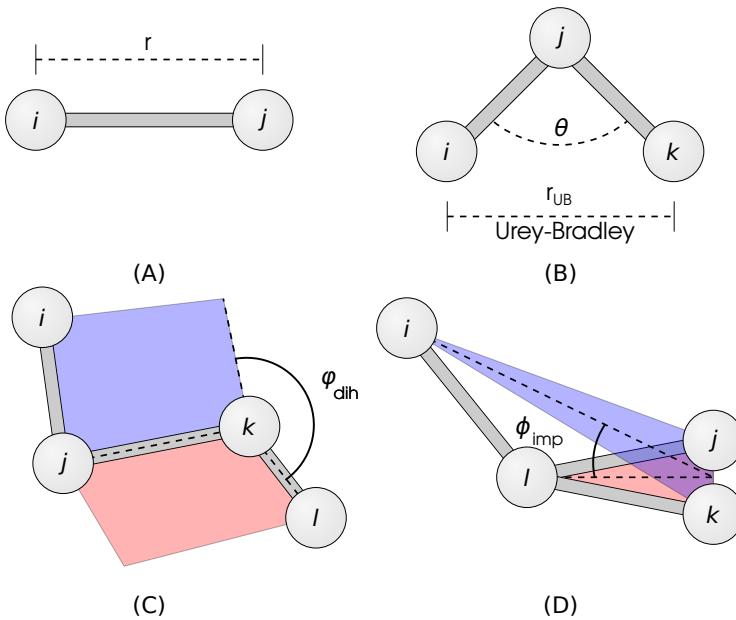


Figure 2.1: The Bonded Interactions Calculated In Classical Forcefields.

(A) The energy of Bond Stretching is approximated as a harmonic oscillator with respect to their separation r . (B) Angles between neighbouring covalently bonded atoms are also approximated as a harmonic oscillator with respect to the angle θ . In some forcefields such as CHARMM there is a correction term for these angular interactions known as Urey Bradley forces. This is calculated using the separation between the non-bonded atoms $i-k$ in the triplet with the parameter r_{UB} . (C) The dihedral angle between four atoms is calculated by constructing two planes. Each plane is constructed to contain three of the four atoms in the set. One plane encompasses atoms i, j and k here colored in blue and the other plane contains the j, k and l atoms colored in red. The dihedral angle is then calculated by taking the angle between these two planes along the line they intersect, the line formed by the $j-k$ bond. (D) The improper dihedral angles enforce the planarity of a molecular configuration. A plane is constructed to contain the i, j and k (blue) atoms and another plane is constructed to contain the j, k and l atoms (red). The improper angle is then calculated as the angle between these two planes.

Over time several additions have been made to this formalism in order to reproduce experimental measurements of chemical systems [119]. We will discuss two of the major additions to the CHARMM formalism. The first is energy correction maps, abbreviated “CMAP corrections” concerning the dihedral parameters [120, 121]. These were motivated by the observation that MD simulations were not correctly reproducing protein secondary structure, and resulted in errant Ramachandran distributions [120, 122]. Mathematically, these corrections take the form of a two dimensional grid which is used to interpolate results from *ab initio* quantum mechanics calculations for each torsion angle. Continued adjustments to $U_{dihedral}$ in the form of these CMAP corrections have resulted in substantial improvements to the CHARMM forcefield, particularly for disordered proteins [123].

Non Bonded Interactions

The term $U_{non-bonded}$ captures interactions which arise when atoms are not covalently bound to each other. Namely, Coulomb forces due to electric charges on the atoms, attractive van der Waals interactions and repulsion due to Pauli Exclusion,

$$U_{non-bonded} = \underbrace{\sum_{i>j} \epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)}_{U_{Lennard-Jones}} - \underbrace{\sum_{i>j} \frac{q_i q_j}{r_{ij}}}_{U_{coulomb}}. \quad (2.15)$$

Note how the repulsive Pauli Exclusion and attractive dispersion forces have been combined into one term known as the Lennard-Jones potential or U_{LJ} . The σ parameter denotes the location of the local minima in the Lennard-Jones potential. This is the optimum distance that two atoms will rest against each other in the absence of other effects. The ϵ parameter denotes the depth of the potential well, or how stable the two atoms will be in the minimum energy configuration. This is very important for certain physical parameters such as osmotic pressure [124].

Meanwhile, the partial charge assignments q_i for each atom are very important in a biological context, for stability of protein conformations of salt bridges and the solvation energy of different molecules [125].

By focusing on adjusting the charges of an atom to fit the solvation energy of a molecule and adjusting the Lennard-Jones parameters to fit the osmotic pressure measurements, we can isolate and determine these non-bonded parameters fairly well.

Another modification to the strict CHARMM formalism can be found in its treatment of Lennard-Jones parameters. Over time there have been an increasing number of “non-bonded fix” or “NBFIX” corrections to the Lennard-Jones parameters in CHARMM. These parameters modify the Lennard-Jones parameters between *specific* atom types. For example, in order to reproduce experimental values of osmotic pressure for sodium chloride, the values of ϵ and σ for these ions are modified values *only* when calculating the Lennard-Jones potential between sodium and chloride atoms. These modifications fixed issues with certain charged species, such as sodium and chloride, which were associating too much in bulk solution or the overly stable salt bridges within proteins [123, 124, 126, 127].

2.2.1 Philosophy of Different Molecular Mechanics forcefields.

At the time of writing, the four popular forcefields for the simulation of biomolecules are CHARMM, AMBER, GROMOS and OPLS. Each of these have a slightly different philosophy in their formulation. They may be bottom up, as in the case of AMBER and CHARMM or top down, in the case of OPLS. Bottom up forcefields take the results from *ab initio* MD calculations as an initial guess and them by tweaking the parameters in the functional form in equation 2.13. Further refinement is often performed to match experimental results. Conversely, top down forcefields tweak these parameters with reference to experimental measurements. Ultimately, the results from *in silico* experiments must match those of wet lab experiments, so the development of all

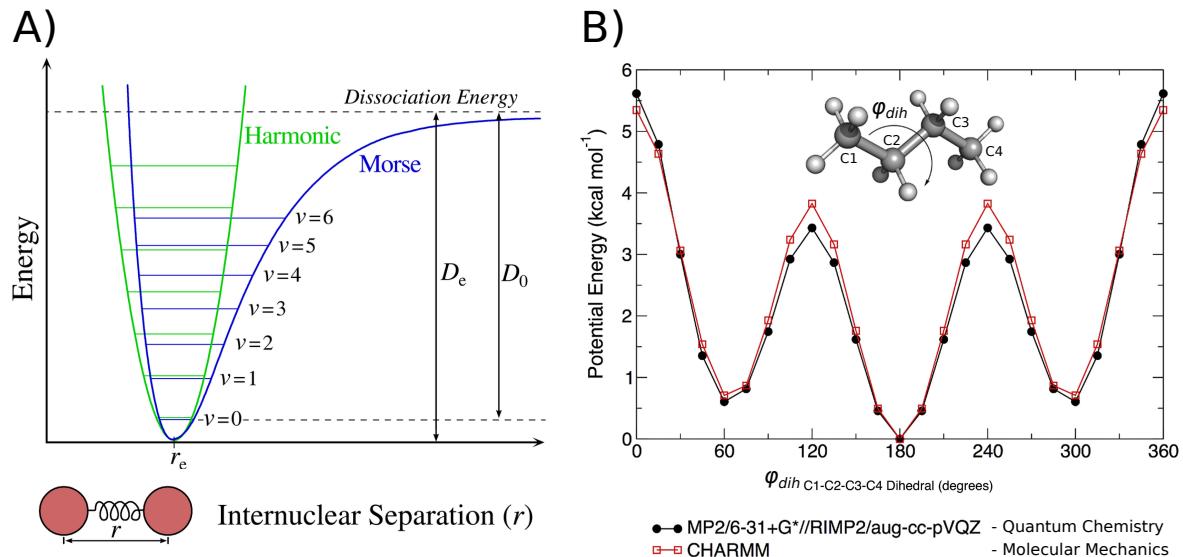


Figure 2.2: Comparison Between Potentials in Quantum and Classical Forcefields

A) The Morse potential was formulated to approximate the potential energy surface associated with the stretching of covalent bonds (blue). At low temperatures (the ground state, $v = 0$) like those found in classical MD there is good agreement between the Morse potential and the harmonic oscillator (green) (Credit, Mark Somoza 2006).

B) Here the potential of the dihedral angle between the atoms C1,C2,C3 and C4 in a butane molecule is calculated using two methods: Quantum Chemical calculations and approximations using the functional form in equation 2.14 [118]. Note how the appropriate choices of k_φ , n and δ have closely approximated the results in the more accurate quantum mechanical calculations.

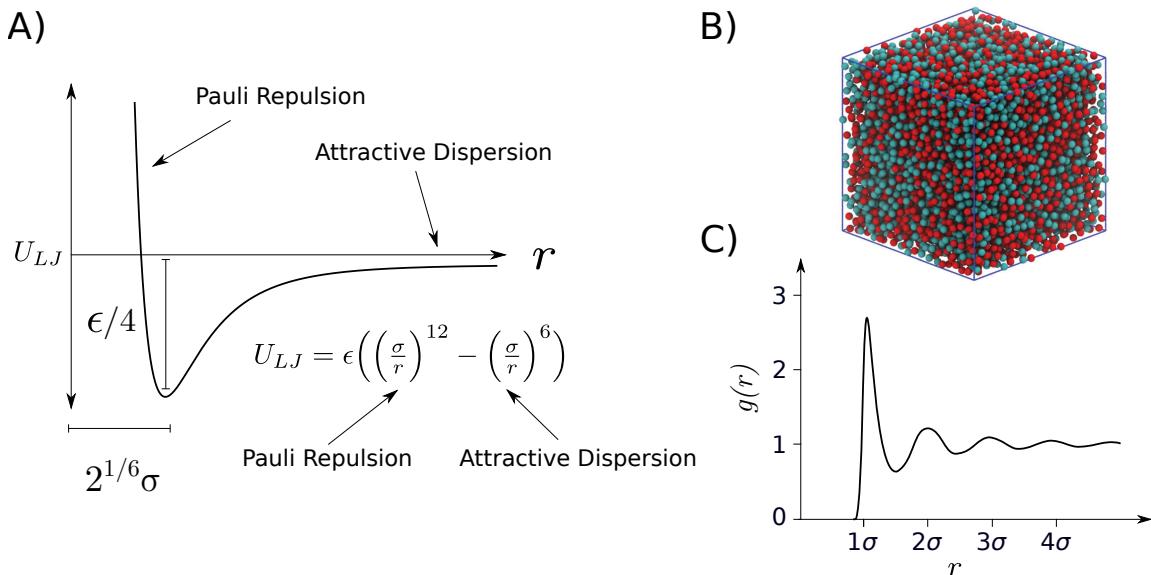


Figure 2.3: The Lennard-Jones Potential

A) The Lennard-Jones potential function has two regimes, the far region one dominated by attractive dispersion forces and the close region dominated by repulsion. In the case of atomic systems this is due to the Pauli exclusion principle. B) An example of a fluid modelled with Lennard-Jones particles [128]. C) The radial distribution function (g) for a Lennard-Jones fluid [129]. Note that the peaks in the distribution are spaced roughly 1σ apart.

forcefields has elements of both philosophies. All forcefields assign the partial charges to atoms using the results of *ab initio* MD calculations. The rest of the parameters are derived with other methods, such as attempting to match the known secondary structures of peptides [123]. Below is a short summary of the philosophy for the four major forcefields taken from the review by Justin Lemkul [118].

- CHARMM: The most popular all atom forcefield. To build CHARMM, QM optimized geometries and molecular dipole moments are compared to those found from classical MD simulations. Molecular degrees of freedom such as dihedral angles are also fit with QM energy profiles, an example can be seen in 2.2. Macroscopic experimental quantities are also used to validate the parameters in this forcefield, such as solvation energies, crystal geometries, heats of vaporization and conformational sampling of biomolecules [123]. One of the reasons for this forcefield's popularity is its considerable library of supported compounds, especially when combined with its generalisation module CGENFF [130].
- AMBER: An all atom forcefield which is built from the ground up from results from quantum mechanical calculations and results from spectroscopy. A version of this forcefield has become the favorite for the simulation of disordered proteins [131, 132]. There is also a generalised version of the AMBER forcefield known as GAFF [133, 134]. Comparisons between generalised forcefields can be found in [135].
- OPLS: An all atom forcefield. The OPLS forcefield takes the philosophy that

since many biomolecules share similar geometries to certain organic liquids, biomolecules can be accurately parameterised by creating a forcefield which correctly reproduces the experimental measurements for these species. Parameters are derived to accurately reproduce the liquid density of certain organic liquids. These parameters are then used as roots to construct larger biomolecules by drawing analogies between similar molecular geometries.

- GROMOS: A united atom forcefield, where hydrogen atoms are typically merged into the heavy atom they are bound to. Hence, they are not explicitly treated. This is known as a united atom forcefield. Charge assignment is done with DFT. Interestingly, GROMOS uses a quartic form of the bond stretching term

$$U_b = \frac{1}{4}k_b(b^2 - b_0^2)^2. \quad (2.16)$$

The parameters are adjusted to match results from experiments, such as measurements of solvation energies and liquid densities. GROMOS forcefields are popular in some contexts because of their ability to reproduce partition coefficients between polar and non-polar media, a similar chemical context to what is found at the interface between a lipid membrane and bulk water.

2.3 Periodic Boundaries To Pretend A Protein is Inside a Cell

Inside cells, proteins are immersed in a large solvation environment composed of water and salts [3]. An example can be seen in figure 2.4. However, simulating a large amount of bulk solvent would be computationally wasteful, while simply truncating the system with vacuum at the boundaries leads to water molecules aligning their dipole moments along the boundary and perturbing equilibrium dynamics. In order to avoid such artefacts, we have to replicate the large cellular environment somehow [136]. We could make a simulation box large enough to replicate the behavior of a bulk solvent, but even with a large simulation box we can still observe artifacts associated with the vacuum at the boundaries [109]. So, to avoid these boundary effects, we use periodic boundary conditions (PBCs), allowing atoms to move between images in the simulation box. This replicates the molecular system infinitely in every direction (Figure 2.5A).

Using PBCs might remove vacuum from our molecular system but now we have a different problem. Effectively, with the PBCs, we have created a system with an infinite number of atoms. We have to somehow limit the number of computations we perform. We could simply truncate the calculation of interactions $U_{\text{non-bonded}}$ after a certain cut-off distance. This is not an issue for U_{LJ} because the $1/r^6$ and $1/r^{12}$ terms in equation 2.3 decay very quickly for large r . Inaccuracies due to this approximation can be further ameliorated with the use of a smooth switching function [137, 138]. On the other hand, the $1/r$ dependence in U_{coulomb} decays much more slowly so truncating it leads to a noticeable loss in accuracy [139–143]. So, we have to calculate the contributions of U_{coulomb} between all periodic images of every atom. Note that this periodicity requires

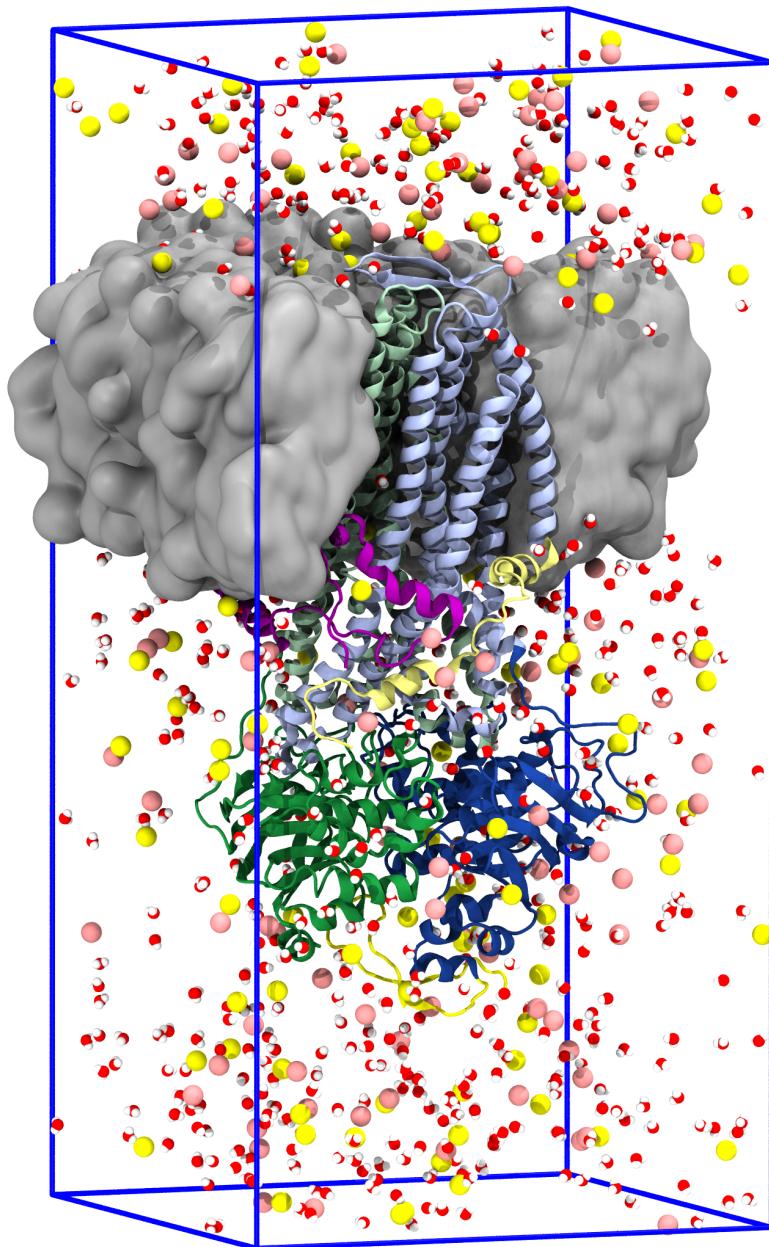


Figure 2.4: An Example of a Solvated Biomolecular Environment Ready for Simulation

This rendering shows a CFTR protein embedded in a lipid bilayer, immersed in a potassium chloride solvent. Half of the bilayer has been stripped away, as has much of the solvent so the protein is visible. The phospholipid lipid bilayer is colored grey, while the potassium chloride ions are colored pink and yellow, respectively. Water molecules are red and white. The blue box indicates the boundaries of the unit cell. This system contains roughly 190,000 atoms in total.

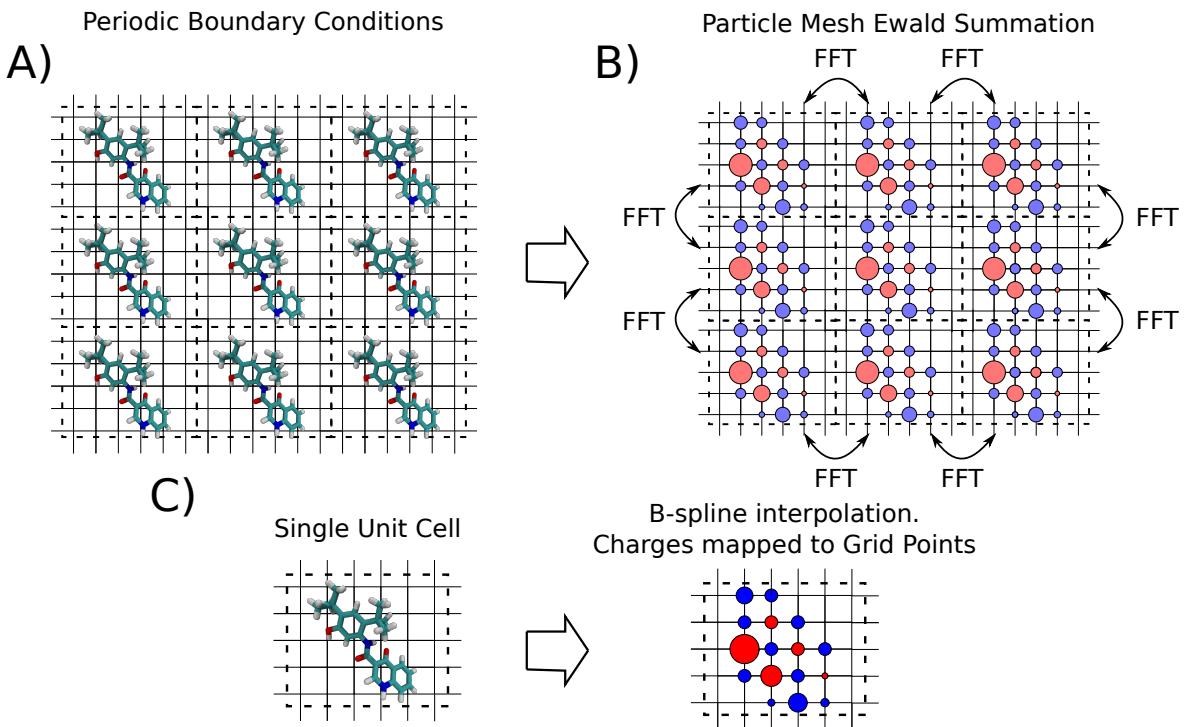


Figure 2.5: Particle-Mesh Ewald Summation

A) The molecular system is repeated infinitely along all axes, when atoms reach the edge of the simulation box, they are allowed wrap around to the other side of the box. B) The charges in the infinite periodic system are approximated onto a regular grid. Then the potential in the infinite system is calculated via a Fast Fourier Transform (FFT). C) A more detailed view of the charge mapping procedure. The charges in the system are interpolated onto the grid using B-spline interpolation.

that the unit cell is electrically neutral, else the contribution of potential energy from $U_{coulomb}$ will be infinite, leading to artefacts [144].

To calculate $U_{coulomb}$ in all periodic images and limit computational intensity of our calculations, we use a clever scheme known as Particle-Mesh Ewald summation (PME). Interestingly, this scheme ends up scaling better than the pairwise summation in equation 2.15 might imply. The direct summation scales with computational complexity of $O(N^2)$ with the number of atoms while the infinite PME scheme scales as $O(N \log N)$ [145], though there are some further considerations for large systems on parallel architectures [146]. Even with these sophisticated algorithms, the calculation of electrostatic potential in the infinite system still represents the largest computational bottle-neck in classical MD [146].

For a detailed review of different Particle-Mesh Ewald summation methods and the mathematics behind the method see [147]. A brief outline of the Smooth Particle Mesh Ewald summation algorithm is given below

1. An *ansatz* is used where $U_{coulomb}$ has Gaussian screening charges added to it and simultaneously subtracted away to create a smooth potential. The details can be

found in [147].

$$U_{coulomb} = U_{screening-charges} + U_{coulomb} - U_{screening-charges}. \quad (2.17)$$

Terms in these equations are then rearranged such that one term is evaluated with a Fourier transform and the other term is evaluated using a direct sum.

$$U_{coulomb} = U_{FT} + U_{direct-sum}. \quad (2.18)$$

2. The charges to be evaluated using U_{FT} are interpolated onto a grid using B-spline interpolation functions. This is the procedure demonstrated in figure 2.5.
3. The charge density functions for the charges on the grid are transformed into frequency space using Fast Fourier Transforms.
4. The Poisson equation is solved numerically in frequency space for these charges.

$$\nabla^2 \tilde{U} = 4\pi \tilde{\rho}(\mathbf{k}) \quad (2.19)$$

Where \tilde{U} is the component of U_{FT} we solve for in frequency space and $\tilde{\rho}$ is the Fourier transform of the smooth scalar function for the interpolated charge densities.

5. An inverse Fourier transform is calculated for the solution to \tilde{U} to transform it back into real space.
6. The interactions in $U_{direct-sum}$ are evaluated using a simple pairwise summation.
7. Now that $U_{coulomb}$ is known at every position in the unit cell, we can move atoms according to the contributions from this potential using Newton's second law.

2.4 Controlling the Temperature and Pressure in a Simulation

Living things are very sensitive to their external environment. Enzymes only work in a narrow range of temperatures and cells burst apart in the absence of pressure [148, 149]. As such, to correctly understand biological systems we not only need to simulate the dynamics of the atoms inside them, but we must also make sure that the virtual environment in our simulations matches what is found inside cells or in the laboratory. Our simulations should seek to approximate the environment of an open topped test-tube sitting in a pressure and temperature controlled laboratory. To do this, we make use of some statistical ensembles chosen for their performance in regulating the thermodynamic quantities in a simulation and their computational expense.

2.4.1 Hot and Cold with the Nosé-Hoover Thermostat

Recall that the temperature of a system is a direct function of the velocity of its constituent particles. So by regulating the ensemble of velocities we can control the

temperature. We begin a simulation by choosing the velocities of the atoms within the system from a Maxwell-Boltzmann distribution

$$f(v_i) = \left(\frac{m_i}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{m_i v_i^2}{2k_B T} \right), \quad (2.20)$$

where $f(v_i)$ is the proportion of particles with velocity v_i , k_B is Boltzmann's constant. Note that $i = 1, \dots, N_{df} = 3N$ as we choose a velocity component for x, y and z separately.

Despite starting from the same Cartesian coordinates, randomly sampling velocities from the Maxwell-Boltzmann means that replicate simulations will immediately begin from different points in phase space. Their coordinates will quickly diverge, raising questions around how long one should run a simulation and how many replicates they should run in order to collect reliable statistics. According to Knapp et al. [150], a good rule of thumb is to simulate between 5 and 10 replicates depending on the availability of computing resources¹. In this thesis we prioritised long time scales to sample slow motions, so 3 replicates with run times between 1 and 2 microseconds were produced for all systems.

After the initial choice of velocities, the temperature in a simulation is maintained by directly modulating the velocities of the atoms to maintain the target temperature T_0 . There are many schemes which attempt this. We will discuss the Nosé-Hoover thermostat in detail here because it was used during the production runs in this thesis [151–153]. However, for the equilibration phase of the simulations, we used the Berendsen thermostat because it is faster at correcting large temperature differentials but does not produce the correct statistical ensemble [154, 155]. We also note that the field has since moved on to favor the Bussi thermostat, which is an extension of the Berendsen thermostat, as it works well in most contexts [108, 154].

The Nosé-Hoover thermostat is characterised by the use of an extra, massive particle coupled to an external bath. The use of a single bath has been associated with issues with ergodicity and so usually this particle is coupled to a chain of external baths. Usually, the software simulation package GROMACS uses a chain of 10 baths, $M = 10$ [153, 156, 157]. The Hamiltonian for the Molecular Dynamics system coupled to a chain of M external baths is then

$$H_{NH}(\mathbf{x}, \mathbf{p}, \eta_1, \dots, \eta_M, p_{\eta_1}, \dots, p_{\eta_M}) = H_{MM} + \sum_j^M \frac{p_{\eta_j}^2}{2Q_j} + k_B T N_{df} \eta_1 + k_B T \sum_{j=2}^M \eta_j. \quad (2.21)$$

¹Recall that in the ergodic limit, a quantity f calculated from the ensemble average of replicates will be equal to the time average calculated from one replicate in the limit $t \rightarrow \infty$. That is

$$\langle f \rangle_i = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\tau) d\tau$$

Here, usually $N_{df} := 3N$ unless there are constraints placed within the system to freeze atoms. η denotes the 1 dimensional coordinate of the thermostat particle with mass Q , while H_{MM} is the Hamiltonian of the unregulated molecular mechanics system

$$H_{MM}(\mathbf{x}, \mathbf{p}) = \underbrace{\sum_i^N \frac{\mathbf{p}_i^2}{2m_i}}_{E_{kinetic}} + U_{CHARMM}(\mathbf{x}). \quad (2.22)$$

By using Hamilton's equations of motion, H_{NH} evolves by

$$\begin{aligned} \dot{\mathbf{x}}_i &= \mathbf{p}_i/m_i \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i - \mathbf{p}_i \frac{p_{\eta_1}}{Q_1} \\ \dot{\eta}_j &= p_{\eta_j}/Q_j \\ \dot{p}_{\eta_1} &= \left[\sum_i^N \frac{\mathbf{p}_i^2}{m_i} - N_{df}k_B T \right] - p_{\eta_1} \frac{p_{\eta_2}}{Q_2} \\ &\vdots \\ \dot{p}_{\eta_j} &= \left[\frac{p_{\eta_{j-1}}^2}{Q_{j-1}} - k_b T \right] - p_{\eta_j} \frac{p_{\eta_{j+1}}}{Q_{j+1}} \\ &\vdots \\ \dot{p}_{\eta_M} &= \left[\frac{p_{\eta_{M-1}}^2}{Q_{M-1}} - k_b T \right], \end{aligned} \quad (2.23)$$

where \mathbf{F}_i is the force vector on the i th particle. It may be calculated from U_{CHARMM} using Newton's second law.

The parameters Q_j are chosen by the user to control the coupling strength of the baths to each other. We usually choose

$$Q_j = \frac{\tau_{NH}T_0}{4\pi^2} \quad \forall j, \quad (2.24)$$

where τ_{NH} is the time interval between when the thermostat parameters are updated. This means that whenever the simulation is not at a time step that is a multiple of τ_{NH} , we can just evaluate U_{CHARMM} as normal but every interval of τ_{NH} , we rescale the velocities according to the equations of motion in 2.23 to match the correct temperature T .

Remember that we can always calculate the temperature using the instantaneous velocities in the simulation using

$$T = \frac{2E_{kinetic}}{3Nk_B} = \frac{\sum_i m_i v_i^2}{3Nk_B} \quad (2.25)$$

The Nosé-Hoover thermostat, when chained infinitely, allows us to accurately produce what's known as an NVT ensemble, also called the canonical ensemble in the statistical mechanics literature [153]. Where the number of particles in the system (N) remains constant, the volume of the system remains constant (V) and the temperature remains constant (T). In a realistic environment, the pressure remains constant rather than volume, so we need another regulatory mechanism to modulate the volume of the system to regulate the pressure (P) and produce an NPT ensemble.

2.4.2 Under Pressure with the Parrinello-Rahman Barostat

Pressure is critical to the function of living organisms. Membranes burst apart at low pressures [158] and at high pressures cellular function is disrupted [159]. In order to accurately reflect the atmospheric pressure at which living things thrive we have to accurately calculate and modulate it during our simulation.

In order to measure the pressure at the simulation walls, we follow the procedure in [160] by calculating a quantity known as the virial:

$$W(\mathbf{x}) = \sum_i^{N-1} \sum_{j>i}^N \mathbf{r}_{ij} \cdot \mathbf{F}_{ij}, \quad (2.26)$$

where $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ is the Cartesian distance between the i th and j th atoms, while \mathbf{F}_{ij} is the force extorted on atom j by atom i . This is then substituted into the equation

$$P(\mathbf{x}) = \frac{Nk_B T + \langle W \rangle_i}{V}. \quad (2.27)$$

And so using equation 2.27, we can modulate the volume V of the simulation in order to control the pressure throughout the simulation.

For this purpose, we apply the Parrinello-Rahman barostat [161, 162], using a procedure with a similar philosophy to the extended Hamiltonian used in the Nosé-Hoover thermostat. In this case, the system is coupled to an external pressure bath rather than an external temperature bath. First we define that the basis vectors for the periodic simulation box as $\underline{h} := [\mathbf{a}, \mathbf{b}, \mathbf{c}]$. When the box is scaled to change the volume these basis vectors are multiplied by a set of scalars $s_i := (\xi_i, \eta_i, \zeta_i) \in [0, 1]$. We perform a change of coordinates so that the contributions of the particles onto the boundaries is easily calculated from our equations so we express the atomic coordinates as

$$\begin{aligned} \mathbf{x}_i &= \xi_i \mathbf{a} + \eta_i \mathbf{b} + \zeta_i \mathbf{c} \\ &= \underline{h} \mathbf{s}_i. \end{aligned} \quad (2.28)$$

Defining $\underline{G} := \underline{h}^T \underline{h}$, the Lagrangian for the scaling system then becomes

$$L = \frac{1}{2} \sum_i^N m_i \dot{\mathbf{s}}_i^T \underline{G} \dot{\mathbf{s}}_i - \sum_i \sum_{j>i} \phi(\mathbf{r}_{ij}) + \frac{1}{2} M \text{Tr}(\dot{\underline{h}}^T \dot{\underline{h}}) - P_{ext} V, \quad (2.29)$$

where $\phi(\mathbf{r}_{ij})$ is the pairwise potential between two atoms in U_{CHARMM} , while M is a constant of proportionality associated with the kinetic energy derived from the movement the particles undergo as they scale. It has units of mass. P_{ext} is our target, externally applied pressure. This Lagrangian allows us to derive the equations of motion

$$\begin{aligned}\ddot{\mathbf{s}}_i &= - \sum_{j \neq i} \frac{1}{m_i \mathbf{r}_{ij}} \frac{d\phi(\mathbf{r}_{ij})}{dr_{ij}} (\mathbf{s}_i - \mathbf{s}_j) - G^{-1} \dot{G} \dot{\mathbf{s}}_i \\ \ddot{\mathbf{h}} &= \frac{1}{M} (\mathbf{Y} - P_{ext}) \underline{\sigma}.\end{aligned}\tag{2.30}$$

The matrix $\underline{\sigma} := V(\mathbf{h}^T)^{-1} = V[\mathbf{b} \times \mathbf{c}, \mathbf{c} \times \mathbf{a}, \mathbf{a} \times \mathbf{b}]$ contains information about the size and orientation of the simulation box, while

$$\mathbf{Y} = \frac{1}{V} \sum_i m_i (\underline{h} \dot{\mathbf{s}}_i) (\underline{h} \dot{\mathbf{s}}_i)^T + \sum_i \sum_{j > i} \frac{1}{r_{ij}} \frac{d\phi(\mathbf{r}_{ij})}{dr_{ij}} \mathbf{r}_{ij} \mathbf{r}_{ij}^T,\tag{2.31}$$

represents the stress tensor which acts across each of the faces of the unit cell. This system of equations can be solved numerically to control the pressure of the simulation system by modulating the length of the basis vectors \mathbf{a}, \mathbf{b} and \mathbf{c} contained in \underline{h} .

Together with the Nosé-Hoover thermostat, the Parrinello-Rahman barostat produces the NPT ensemble, also called the isothermal-isobaric ensemble in the statistical mechanics literature. The combination of these two methods is thus appropriate for simulating a cellular environment.

2.5 The Process of Preparing an MD Simulation

The process of taking a molecular structure and putting it in a cellular environment to simulate it at physiological temperatures is both an art and a science. It's a science because a biophysicist must be aware of the many tricks that structural biologists use to image a macromolecular complex—but it's an art because accounting for those tricks and making the necessary modifications is rarely straightforward. How do you build a missing loop? What charge state is an amino acid most likely to take in a physiological context? These questions must be carefully answered by analysing the literature about the system of interest. Once the protein structure has been built, the system is immersed in a water solvation bath alongside a concentration of salt ions, usually sodium chloride if the environment is thought at be extra cellular or potassium chloride if the environment is thought to be intracellular. Once the initial conditions have been decided, we need to make a few preparatory steps so the simulation collects reasonable results and doesn't hurtle along some unphysical trajectory. The steps so that the simulation remains realistic are fleshed out in some more detail in [108] but we produce a short summary below.

1. Minimisation: Here the atoms are moved down along U_{CHARMM} to resolve any clashes in the system which would cause LINCS or SHAKE to diverge. This is usually done with simple minimisation algorithms such as steepest descent or

conjugate gradient descent [163]. This performed while the simulation box has a fixed volume.

2. Relaxation: Harmonic restraints are placed on the heavy atoms (non hydrogen) in the system so that large conformational changes do not occur while the macromolecules are heated and settle into their solvation environment. This may be done under the NVT or NPT ensembles depending on the system. Sometimes the system is heated slowly from 0 Kelvin up to the desired thermal temperature in order to avoid large conformational changes which might result from different parts of the system heating at different rates.
3. Equilibration: Often after relaxation more simulation time is run so that the system can settle into local minima further. This process makes sure that the physically relevant local minima are being sampled once we move to production. This process could be run for a few nanoseconds or up to a microsecond. It depends on the system. This is usually done with the NPT ensemble.
4. Production: Here the NPT ensemble is applied and the system is allowed to evolve under Newton's equations of motion while data is collected for analysis.

2.6 Choosing an Appropriate Time Step

The discrete time step, Δt which is used to integrate our equations of motion is one of the most important determinants in the performance of the simulation. We would like Δt to be as large as possible, so that the minimum number of calculations are made to sample the desired time scale. In the case of proteins this usually runs between 10^{-6} and 10^{-3} s [131].

As you can see in table 2.1 the fastest motion in molecular systems is dictated by stretching of covalent bonds. Studies of the resonance of molecules by infrared spectroscopy determined that the O-H type bonds oscillate the fastest, with a resonance peak at 3600 cm^{-1} [164].

Due to Nyquist's theorem the largest Δt parameter we can choose *must* be less than half the speed of the fastest degree of freedom in the system [165]. However, empirically we have found that condensed matter systems require even shorter time steps to maintain their stability [163]. The Verlet leap-frog scheme used in most MD codes requires between 5 and 10 integration steps per period of the fastest harmonic mode in a system, to maintain stability [166, 167]. The choice of too large a time step means that the system will escape local free energy minima, accumulating kinetic energy and eventually "blow-up" [108]. In the case of biomolecular systems we are challenged by the fact that they are so hydrogen-rich. Since hydrogen is so light, its motion is much faster compared to the other molecular motions involving heavier, slower moving atoms. Its correlation time is on the order of 1 femtosecond. In classical simulations we are able to get away with using 2 femtoseconds with the use of specialised integration schemes such as SHAKE [168] and LINCS [169] to constrain the fast motion of hydrogen atoms. This allows us to use $\Delta t = 2\text{ fs}$ in atomistic classical MD simulations.

The use of techniques such as hydrogen mass repartitioning [170], virtual site topologies

Motion	Timescale
Covalent Bond-stretching	$1 - 2 \times 10^{-15}$ s
Covalent Bond-angle bending	$5 - 10 \times 10^{-15}$ s
Sidechain Motions	$10^{-12} - 10^{-6}$ s
Rigid Body Motions	$10^{-9} - 1$ s
Ion Conduction	$10^{-9} - 10^{-6}$ s
Protein Conformational Changes	$10^{-9} - 10^{-3}$ s
Alpha Helix Formation	$10^{-9} - 10^{-6}$ s
Beta Sheet Formation	$10^{-6} - 10^{-3}$ s
Protein Folding	$10^{-6} - 10$ s

Table 2.1: Timescales of Motions in a Molecular System

The time step of a simulation must be small enough to capture the motions in the fastest degree of freedom. In hydrogen-rich biomolecular systems the bottle neck can be found in the fast bond vibrations in lighter atoms. This stands in tension with the phenomena we are interested in on longer timescales such as protein folding. Sources: [68, 163, 164, 167, 172, 173]

[167] and multiple time step schemes [171] have also gained popularity in recent years in order to increase time steps further, up to $\Delta t = 5\text{fs}$.

2.6.1 Verlet Leap-Frog Integration

To produce molecular trajectories we can use the potential U_{CHARMM} which we calculated with equation 2.13 and calculate the forces exerted on the atoms in the system. By Newton's 2nd law we have

$$\mathbf{a}(\mathbf{x})_i = \frac{d^2\mathbf{x}_i(t)}{dt^2} = -\frac{1}{M_i} \nabla_i U_{CHARMM}(\mathbf{x}_i). \quad (2.32)$$

We can use this calculation of acceleration a_i of the i th atom to update the positions and velocities of the atoms in the molecular system with the following triplet of equations known as the leap-frog Verlet method [164]:

$$\begin{aligned} \mathbf{v}_i^{n+1/2} &= \mathbf{v}_i^{n-1/2} + \Delta t \mathbf{a}_i^n \\ \mathbf{x}_i^{n+1} &= \mathbf{x}_i^n + \Delta t \mathbf{v}_i^{n+1/2} \\ \mathbf{v}_i^{n+1} &= \mathbf{v}_i^{n+1/2} + \frac{\Delta t}{2} \mathbf{a}_i^{n+1} \end{aligned} \quad (2.33)$$

Note that $v_i^{n-1/2}$ will have been calculated during the previous time step and a_i^{n+1} may be calculated by the updated positions found by calculating \mathbf{x}_i^{n+1} .

In MD, we are less concerned with the accuracy of a particular trajectory so much as collecting sufficient statistics to calculate macroscopic properties such as free energies or diffusion profiles. This means the choice of 4th order solvers such as the Runge-Kutta method would be inappropriate. Although they may use a large time step, they require 4 evaluations of U_{CHARMM} per iteration and are thus more expensive than any

2nd order method. Hence, we prefer symplectic (energy preserving), 2nd order methods such as Verlet integration so the simulation remains stable after millions of time steps [171].

2.7 Free Energy Calculations and Enhanced Sampling: Making Simulations More Useful

The above work sets out how to perform what is known as unbiased MD simulations. These are powerful tools but as we will discuss in section 2.8.2, if one only relies on unbiased simulations they will quickly exceed the available computer power (Figure 2.6). Imagine there is an event that we know from experimental evidence our system must exhibit, but it occurs slowly. Examples of this include the passage of an ion through a channel and the binding of a drug. We *could* calculate the Gibbs free energy of a given molecular configuration \mathbf{x}_0 using Boltzmann reweighting

$$G(\mathbf{x}_0) = -\frac{1}{k_B T} \ln(P^u(\mathbf{x}_0)). \quad (2.34)$$

Where $P^u(\mathbf{x}_0)$ represents the probability of obtaining state \mathbf{x}_0 , estimated from an unbiased simulation. From here on, a subscript of u indicates a quantity obtained from an unbiased simulation and a subscript b represents a quantity from a biased simulation.

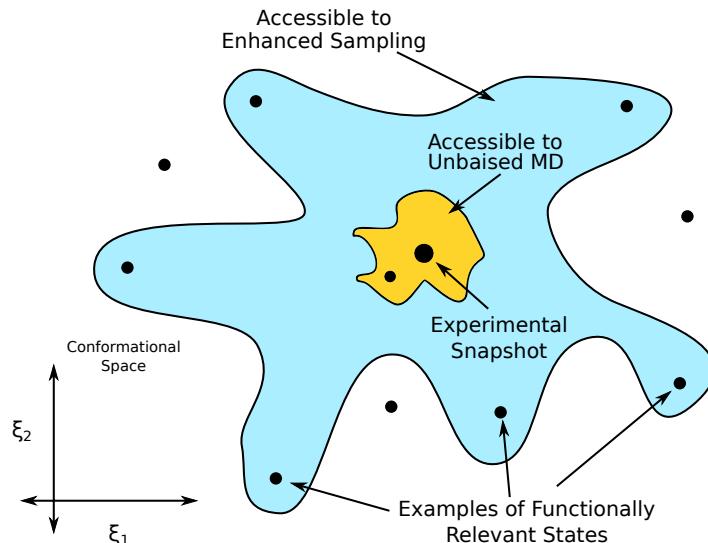


Figure 2.6: Enhanced Sampling Allows Access to More Protein States

The “Resolution Revolution” in structural biology, fuelled by cryo-EM has lead to an era of structural abundance [63]. This means we have many structural snapshots of proteins with which to begin our study of their physical mechanisms. With a given amount of computer time, unbiased MD simulations can explore a small local neighbourhood surrounding these snapshots, shedding limited light on adjacent physiologically relevant functional states. With the same amount of computer time, Enhanced Sampling Techniques and Free energy calculations can access and study many more functional states. The more advanced these statistical techniques become the more functional

states we will be able to access. These states might reveal drug binding pockets or the underlying mechanism behind the protein's function. Advances in computer power and statistical machine learning herald make this an exciting time for to be studying computational biophysics.

Equation 2.34 shows how there is exponentially poor sampling in regions with high U_{CHARMM} . So it is clear that we will not collect sufficient statistics for a good estimate with a reasonable amount of computer power. Therefore, we must be clever in how we direct our available resources. This means intelligently sampling sections of the molecular phase space which are of interest to us physically, but are not reached in our unbiased simulations. A technique that is used extensively throughout this thesis is the addition of a biased potential to the molecular potential U_{CHARMM} which we calculated for the purposes of unbiased MD simulations. This bias will drive the simulation to regions of interest using the formalism

$$U'_{CHARMM} = U_{CHARMM} + U_{bias}(\xi). \quad (2.35)$$

Note how the U_{bias} term is explicitly dependent on a parameter ξ . This parameter is known by many names, an order parameter, a collective variable (CV) or a reaction coordinate (RC). Each of these names has its origin in a different subfield but they all refer to an abstract coordinate which we use to drive the simulation to a given state. This could be a phase transition from a liquid to a gas, the progress of a chemical reaction or more likely in our case, the progress toward a molecular conformation.

The functional form of U_{bias} depends on the free energy calculation being employed. Broadly, there are two sets of mutually exclusive categories of free energy methods. The first two categories are equilibrium and non-equilibrium techniques. Equilibrium techniques perturb the molecular system slowly, as to to always remain close to thermal equilibrium. Conversely, non-equilibrium techniques seek to use theorems from statistical mechanics to account for the thermal energy imparted onto the system due to the addition of the $U_{bias}(\xi)$ function [190–192]. Meanwhile, alchemical techniques may be used to study the change in free energy when the chemical composition of the system is changed [193], while in non-alchemical methods the chemical system remains constant, as in unbiased MD. A list of some of these techniques can be found in table 2.2. A deep theoretical exploration of the most popular techniques may be found in [112] and a frequently updated overview of modern free energy methods can be found in [194]. In the next two sections we will focus on the two most common non-alchemical equilibrium free energy techniques, umbrella sampling and metadynamics.

2.7.1 Umbrella Sampling

Umbrella sampling is possibly the most popular method of calculating the free energy of a system along a reaction coordinate. The conceptual philosophy for the method is demonstrated in figure 2.7. The molecular system is replicated in several "windows" and a harmonic U_{biased} is added to restrain the collective variable ξ at several different points. The fluctuations of ξ are recorded and the overlap between windows allows us

	Equilibrium	Non-Equilibrium
Non-Alchemical	<ul style="list-style-type: none"> • Umbrella Sampling [174, 175] • Replica Exchange MD (REMD) [176, 177] • Metadynamics (MetaD)[178] • String Method with Swarms of Trajectories (SMwST) [179–181] • Gaussian Accelerated Molecular Dynamics (GAMD) [182] • Adaptive Biasing Force (ABF) [183] 	<ul style="list-style-type: none"> • Steered Molecular Dynamics (SMD) [184] • Non-equilibrium MetaD [185]
Alchemical	<ul style="list-style-type: none"> • Free Energy Perturbation (FEP) [186] • Thermodynamic Integration (TI) [112] • Replica Exchange FEP [187] 	<ul style="list-style-type: none"> • Non-equilibrium FEP [188, 189]

Table 2.2: Classification of Common Free Energy Calculation Techniques

Various free energy techniques have been developed over the years. Due to their accuracy and statistical simplicity, the most numerous and most popular methods fall into the category of equilibrium methods. Both equilibrium and non-equilibrium alchemical methods are used in drug discovery to screen novel compounds for affinity to a given target. They may also be used to test the energetics of small changes to a molecular structure, such as from a missense mutation or changes to one moiety on a molecule. Meanwhile, we will focus on non-alchemical methods as they are most useful for the study of the conformational changes of proteins. They are generally cheaper than alchemical methods when considering large conformational changes, and since we can assume that molecular structures remain constant through a simulation, these methods lend themselves well to our purposes.

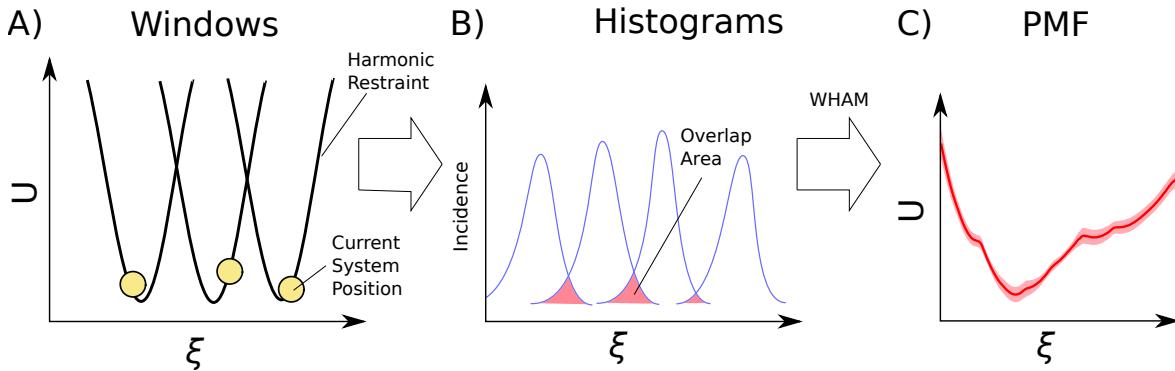


Figure 2.7: Illustration of Umbrella Sampling

A) Several simulations are repeated with only one change. A bias potential is added somewhere along the reaction coordinate ξ . B) The value of ξ is recorded in each of the windows and then graphed as histograms. C) The Overlap in neighbouring histograms is integrated via the WHAM method to calculate the Potential of Mean Force. This gives us the energy landscape. Fluctuations in the overlap in the data can be used to estimate the error for the PMF.

to calculate the potential of mean force (PMF), and thus the energy landscape along the reaction coordinate.

In more technical detail, umbrella sampling involves a functional form where U_{bias} separated into N windows. With the bias function in the n th window has the form:

$$U_b^n = \frac{k_\xi^n}{2}(\xi(\mathbf{x}) - \xi_0^n)^2 \quad (2.36)$$

Where ξ_0^n is the equilibrium position of the restraint. k_ξ^n is the strength of the harmonic restraint in the n th window. Typically this is the same in all windows. The more overlap between adjacent windows the more those windows are attracted to each other and the steeper the gradient of the free energy surface (FES) must be pushing those windows together. Conversely, when there is less overlap between adjacent windows it indicates the presence of a barrier in the energy landscape between those windows.

Umbrella sampling is extremely useful for calculating all sorts of experimental quantities and physiologically relevant properties such as folding energies [195], lipid binding [196], ion conduction [197], and drug binding [198]. However, it is particularly sensitive to the choice of initial configuration and collective variable [196]. The former issue is particular to umbrella sampling because generally short runs are used since so many windows are spawned during the method, the simulations must be at equilibrium *before* the method is attempted, and then sufficient statistics must be collected to average over any conformational changes orthogonal to the collective variable. In these ways, care must be taken when using this method to not introduce systematic error into the calculation [199]. A deep knowledge of the molecular system under investigation can help alleviate some of these issues.

Weighted Histogram Average Method (WHAM)

There are a few candidates for calculating a PMF using the statistics calculated in umbrella sampling. The Weighted Histogram Average Method (WHAM) [200], Umbrella Integration (UI) [201] and the Multistate Bennett acceptance ratio (MBAR) [202] are all used. We will briefly outline the mathematical formulation and estimation of errors of the WHAM method as it is more popular [197]. Our explanation follows [174] which covers the topic in more detail.

The method begins by dividing the sampled region into a set of K histograms with K being greater than the number of biased windows N . The whole PMF can be estimated (poorly) from the i th biased window using

$$P_u^i(\xi) = P_b^i(\xi) \exp(\beta U_b^i(\xi)) \langle \exp(-\beta U_b^i(\xi)) \rangle, \quad (2.37)$$

where the $P(\xi)$ functions represent the probability density function calculated from samples collected at point ξ . The samples collected in each histogram then gives us an estimate of the PMF according to equation 2.34. The estimates from each histogram are then combined in a weighted sum using

$$P_u(\xi) = \sum_i^K p_i(\xi) P_u^i(\xi) \quad (2.38)$$

where the weights, $\sum_{i=0}^N p_j = 1$ are chosen such that the statistical error is minimised across the domain of ξ [203]. This means we solve an optimisation problem for the variance of the unbiased estimates

$$\frac{\partial \text{Var}(P_u)}{\partial p_i} = 0 \quad (2.39)$$

to give the best estimate of the PMF given the samples in each $P_b^i(\xi)$. Essentially we are vertically shifting the estimates obtained in each of the histograms in order to minimise the error across the PMF.

By convention, we estimate the error in the PMF by splitting up the statistics collected for the distributions of $P_b^i(\xi)$ and constructing PMFs from these independent samples. For example, we might have collected 100 ns of data in each window, but we could construct 5 independent estimates for the PMF using 20 ns blocks of trajectories. The Standard Error of the Mean (SEM) can then be used to estimate the error across the surface [109]. By convention, an umbrella sampling calculation is said to have converged when the SEM from these independent samples has fallen below 1 kcal/mol across the PMF. Note that there may be sources of systematic error not captured by this criterion.

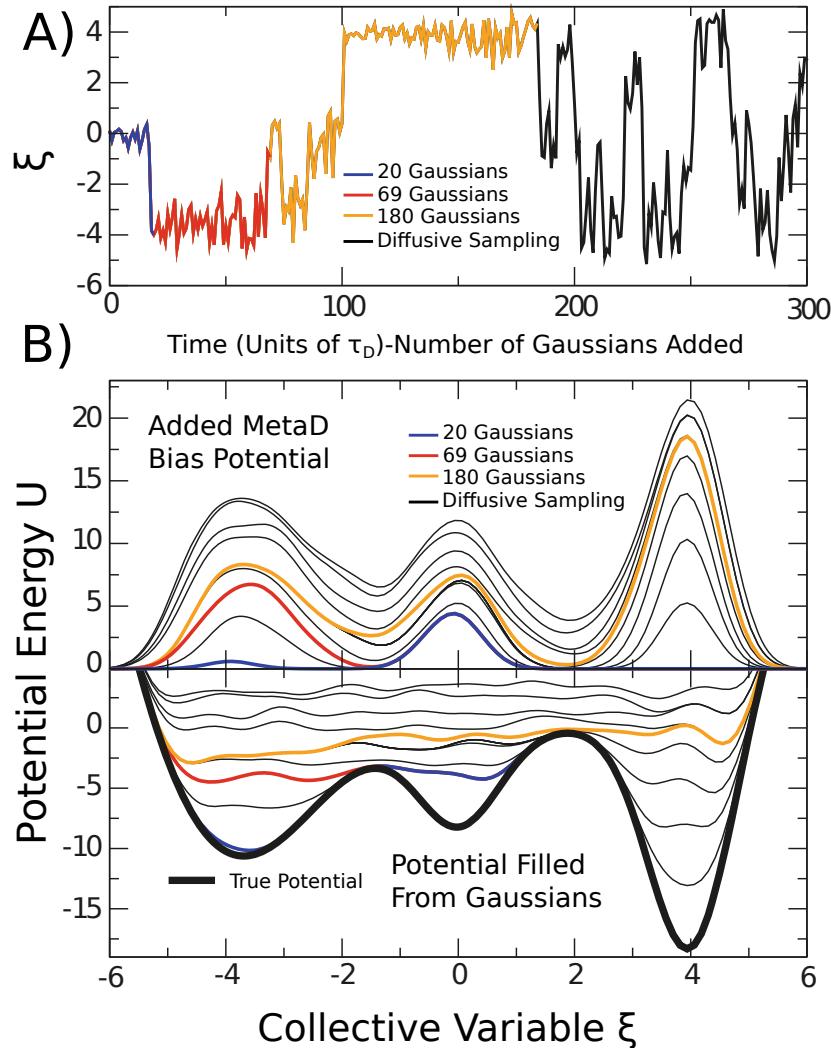


Figure 2.8: Illustration of Metadynamics

A) The trajectory of a collective variable ξ in a metadynamics simulation. The blue region denotes the trajectory up to the time where 20 Gaussians have been deposited. After which the basin at $\xi = 0$ has been filled. The red data points denote the time up until 69 Gaussians have been deposited, after which time the second basin at $\xi = -4$ has been adequately sampled. Finally the Orange data indicates that the 3rd basin at $\xi = 4$ has been adequately sampled after which which time the system begins to sample ξ diffusively and the simulation has converged. B) The upper panel demonstrates the repulsive potentials that have been added to the simulation in order to drive it to new, unexplored regions. Multiplying the values of this function by -1 will be the estimate of the free energy surface². Note how in the lower panel the true potential energy surface is gradually filled by the added Gaussians functions. Source [204].

2.7.2 Metadynamics

Developed in the lab of Michele Parrinello [178], metadynamics has proven to be a popular method in many applications of computational science, not just in molecular dynamics simulations of biomolecular systems [205–207]. A pedagogical review of the method and its variants written by Pratyush Tiwary can be found in [204] or by Giovanni Bussi in [208]. The method relies on a time dependent form of U_{bias} given by

$$U_{bias}(\xi, t) = \sum_{\substack{t'=\tau_D, 2\tau_D, \dots \\ t' < t}} B \exp \left(-\frac{(\xi(t) - \xi(t'))^2}{2\sigma^2} \right) \quad (2.40)$$

This means that at regular intervals of τ_D during the simulation we drop virtual, repulsive Gaussian potentials at positions along ξ in order to encourage the simulation to sample regions of ξ it has not visited already. The process is illustrated in figure 2.8. The thermodynamic assumption of this method is that the deposition is done slow enough that the system remains at equilibrium, so a small Gaussian height should be chosen. Usually, the Gaussian widths σ are chosen to be the size of the variance in the unbiased measurements of ξ .

The FES estimate at time t from this method is simply the sum of all the Gaussians we have added into the system inverted:

$$U_u(\xi, t) = \frac{1}{t_c - t} \int_{t_c}^t U_b(\xi, t') dt \quad (2.41)$$

Formally, convergence is reached when the observed probability density is uniform across ξ . That is:

$$P_b(\xi) = \frac{1}{V_\xi} \quad (2.42)$$

where V_ξ is the volume of the phase space spanned by ξ . However, in practice the function $U_{bias}(\xi)$ is simply inspected at intervals for fluctuations about an average function [209].

There are many flavours of metadynamics. The most popular is well-tempered metadynamics which gradually reduces the Gaussian height B as the simulation progresses [210]. In theory, this guarantees convergence with vanishingly small error. However, this method requires an estimate of how long the simulation will take to converge. There is also infrequent metadynamics which can be used to estimate the diffusion profile along ξ [194, 211–213].

A useful feature of the Metadynamics method is that it can be linearly sped up with the number of parallel simulations. This is known as multiple walker metadynamics [214]. Since we are simply attempting to sample from the same potential U_{CHARMM} up to a desired time, until the criterion in equation 2.42 is met, we can speed this process up by running simulations in parallel to sample different regions of ξ and adding Gaussian to the same U_{bias} function.

Conceptually, metadynamics performs the same role as umbrella sampling and should in theory produce the same results in the same system with the same collective variable ξ . However, it has some specific contexts where it outperforms umbrella sampling. This method is more suited to an exploration of free energy space where it can intelligently explore regions which are poorly sampled, whereas umbrella sampling requires some foreknowledge of the surface being investigated in order to guess which parts of the landscape require more sampling (in umbrella sampling regions with a high gradient in the FES require a higher density of windows). However, metadynamics can be very difficult to converge. A good indicator of such systematic errors is the presence of what is known as hysteresis. Here, unphysically large differences in energy may be calculated between metastable states. This would indicate that there are orthogonal degrees of freedom, not sampled along ξ which will drive the system along a minimum energy pathway. The barrier will eventually come down in the infinite sampling limit but many barrier crossings will need to be observed in order to appropriately sample the intermediate states. Excellent discussions of how to solve these systematic errors, sometimes called hysteresis, be found in articles by Giovanni Bussi [207, 208].

2.8 Short Comings of Classical MD

The short coming of classical molecular dynamics fall into two classes. First is the accuracy of the chemical forcefields outlined in section 2.2.1. Second, there is the inability of modern computers to deliver enough samples of the energy landscape to collect sufficient statistics to reach a rigorous conclusion. We are challenged by the physical nature of molecular systems as outlined in section 2.1.2. The more accurate our forcefields, the more computationally expensive our calculations become, and so the harder it is to deliver a sufficient number of samples. Hence, the solutions to these two problems are diametrically opposed. In this short section we will explore the current efforts to find solutions to both problems.

2.8.1 The Problem with Forcefields

The approximations inherent in equation 2.13 are not without a cost to accuracy. In certain situations, many of which are biologically relevant, it has been shown that polarisation effects not captured by fixed partial charges play an important role in the dynamics of the system. This has been demonstrated in the literature for Gramicidin A, where polarisable forcefields are able to more accurately reproduce the experimental measurements of current [215].

The other context where polarisation is important to consider involve divalent ions such as calcium or magnesium. Here, the highly charged environment near a divalent ion will induce changes in the dipole moment of surrounding atoms. This is not possible in the fixed charge formulation in equation 2.13, making investigations of these biologically important chemical species difficult [216, 217].

However, for most situations, particularly those involving bulk water with a low concentration of solute, classical forcefields are sufficiently accurate to sample conformational motions of biomolecules [218]. Sadly, it should also be kept in mind that classical MD is

not able to simulate any chemistry such as forming and breaking of bonds or a change in the protonation state of an amino acid. Such interactions require considerations of quantum mechanics which are computationally expensive [219].

There are several efforts to address some of the above issues. Some groups are trying to improve the accuracy of classical forcefields using machine learning and Bayesian inference [220, 221]. But there are also attempts to move beyond the functional form of equation 2.13 by explicitly including the effects of polarisation. A popular method at the moment is to add a massless drude oscillator as an extra bead to atoms as in the CHARMM drude forcefields, championed by the laboratory of Alexander Mackerell Jr. and others [222]. An alternative solution is to explicitly calculate the dipole and quadrupole moments of each atom at each time step, as is done in the AMOEABA forcefield [223]. These both substantially increase computational cost but have displayed much better agreement with experiments in biological systems, where classical forcefields have been shown to fail [215, 223, 224].

Ultimately, the functional form in equation 2.13 used by classical forcefields does not have sufficient degrees of freedom to address all possible chemical contexts. Careful consideration must always be given to whether the forcefield is being used in a context which is faithful methods it was formulated to represent. So long as the user is aware of the situations where a given forcefield falls short, classical forcefields can be a powerful tool for the study of molecular systems.

2.8.2 The Problem with Sampling

Collecting sufficient statistics is often computationally infeasible for molecular systems. Even though computers have sped up exponentially for the last 50 years we are still orders of magnitude from being able to brute force the time scales of many biological processes, as displayed in table 2.1.

The slow time step demanded in classical MD due to the fast motions of certain atomic groups such as hydrogen is fundamentally at odds with the time scales of many important biological processes such as drug binding or protein folding which occur on the time scale of milliseconds or seconds.

Methods are now emerging which can intelligently drive a simulation toward regions unexplored in the collective variable space by unbiased simulations. For some time the field has used steered methods or adaptive sampling methods such as Umbrella Sampling or Metadynamics to drive the simulation toward sections of the energy landscape which are under sampled. These methods universally rely on a choice of collective variable ξ which corresponds to a slow degree of freedom. Such a choice is not usually simple. In the case of ion channels one may rationally choose the placement of the ion along the conduction pathway as the collective variable but the choice is less obvious in the case of more global conformational changes.

The success of simulations at the millisecond timescale by D.E Shaw research suggest that we are in reach of an exciting area in biological research [131, 225]. Enhanced sampling methods will be able to routinely reach motions that occur on these time

scales and as software and hardware improve we will be able to push further, to simulate larger systems.

The advances we are seeing at the moment which I find exciting are the use of machine learning methods to tease out these degrees of freedom in order to accelerate them with already established enhanced sampling methods. That is, make a more careful choice of the collective variable ξ . This could be done with a variety of algorithms such as Principal Component Analysis (PCA) [226], Time lagged independent component analysis (TICA) [227, 228] from Frank Noë, a variational approach to conformational dynamics (VACs) applied by members of Michelle Parinello's laboratory [229] or Reweighted autoencoded variational Bayes for enhanced sampling (RAVE) [230] from Pratyush Tiwary's laboratory. These algorithms have the potential to build on the above rigorous physics of simulations and revolutionise our understanding of biomolecular systems. As a small example, in chapter 7 we will discover collective motions using PCA which dilate the CFTR ion channel and then study the passage of ions through this dilated structure using umbrella sampling.

2.9 Conclusion

It is hoped that the preceding chapter can serve as a roadmap for any physicists interested in beginning to study the exciting field of computational biophysics. The information in each section should serve to help the reader understand the foundations of how simulations are performed. The next steps would be to learn more about the molecular biology and biochemistry of macromolecules so they can better understand the chaotic dance occurring inside cells. For recommended reading on this topic there is Schlick [164], Frauenfelder [103] and Phillips [3]. There is a lot to learn and the barrier to entry can seem daunting. The reader is encouraged to join mailing lists and other forums where the thriving computational chemistry and computational biophysics communities communicate. Send cold emails asking for help when you are stuck, remember to consult software manuals as well. Below is a list of software to get you started building and running simulations.

- [Python](#) [231]. A versatile programming language that will help with all parts of a computational workflow.
- [Visualised Molecular Dynamics](#) (VMD) [232]. Molecular visualisation software with a scripting language, great for building simulation systems, viewing trajectories and rendering figures.
- [Alchemy wiki](#). An extensive wiki focussed on pedagogical explanations of alchemical free energy methods.
- [CHARMM-GUI](#) [233]. A web server which offers a great starting point for constructing MD simulation systems and also offers configuration scripts for all the major MD software packages.
- [MDANALYSIS](#) [234, 235]. A python package with a diverse set of tools to analyse molecular dynamics trajectories and structures.

- [Statistical Biophysics Blog](#) A personal blog by respected computational biophysicist Daniel M. Zuckerman. Here you will find some simple theoretical exercises and some helpful personal advice, not just technical. His book “Statistical Physics of Biomolecules: An Introduction” is also a very pedagogical introduction to the subject of molecular biophysics [15].
- [MODELLER](#) [236–238]. A python package which helps build homology models of protein structures or add missing parts to protein models.
- [NAMD](#) [239]. A powerful, scalable user friendly MD simulation package.
- [GROMACS](#) [157]. A less user-friendly MD simulation package than NAMD but offers some different features and (at the time of writing) faster performance.
- [OPENMM](#) [240]. A flexible MD simulation package that is optimised for usage on GPUs. Is controlled with a python API. A good choice for desktops and workstations.
- [PLUMED](#) [241]. A plugin for existing MD simulation packages which lets the user define their own collective variable ξ . Very useful for sophisticated analysis and free energy calculations.
- [Protein Data Bank \(PDB\)](#) [242]. A public database of all published biomolecular structures at atomic resolution. Structures are released under a creative commons license so they can be used for almost any purpose.
- [Uniprot](#) [243]. A database containing a wealth of data, including structures and biochemical annotations for millions of proteins in different organisms. Very handy when starting a new project.
- [ALPHAFOLD2](#) [69]. A highly accurate AI generated database of protein structures. This was considered a watershed moment in the field when it was released. The predicted CFTR human structure from this AI agrees with our predictions in chapter 4.
- [CHARMM forcefield](#) [123]. The parameter files for the CHARMM forcefield formatted to be read by popular MD simulation packages.
- [AMBER forcefield and simulation package](#) [244, 245]. An MD simulation package and toolkit. Here you will also find instructions for how to extract and use the latest AMBER forcefield in other MD simulation packages.
- [PARMED](#) [246]. A python package for converting and manipulating MD file types.
- [CCL.net](#). A job site where computational chemistry jobs are shared.
- [The Biophysical Society](#). The American biophysical society is the largest community of biophysicists. The yearly meetings draw thousands of participants from across the world. At this website you will also find job postings and community news.

No individual who has studied a specific discipline has the skills necessary to pick up biomolecular simulation software and begin using it. Physicists lack the understanding of the biology and the chemistry involved in the biological systems, while chemists and biologists may lack an understanding of the deep mathematics that has gone into producing highly accurate simulations of molecular systems. It will take time to get used to an interdisciplinary way of thinking. The reader is also encouraged to seek out collaborators of wet lab disciplines such as cell biologists and protein biochemists to help answer the important problems in biology.

3

Review of Cystic Fibrosis and The Molecular Basis of its Treatment

Because of what's inside me; Because of my genes.

– -Bob Flanagan “Why?” [247].

3.1 Introduction

Here we aim to elucidate the molecular nature of Cystic Fibrosis (CF)—a disease caused by mutations to a protein called the Cystic Fibrosis Transmembrane conductance Regulator (CFTR). In subsequent chapters we will use Molecular Dynamics (MD) to determine how rare mutations cause this protein to malfunction. These results are presented with the goal of up building a molecular understanding for the treatment of rare forms of CF. Given the importance of the CFTR protein system to this work, in this chapter we will look at its structure and dynamics in some detail. In some sections of this chapter, I have gone into far more detail than necessary in order to compile details which might serve as a reference text for future endeavours into molecular studies of CFTR.

To understand the motivation for this level of detail, we will first review some of the medical literature surrounding the disease itself. CF is a multi-system disease with many complications. The management of symptoms can be a significant burden for patients and their families but recent treatments and advancements in translating basic biological research into the clinic have given many hope of a more normal life.

After this review of clinical outcomes, we will once more work our way upward, as ultimately, CF symptoms are due to the mutation of a single gene. The detailed outline of CFTR structure and function will be our starting point for a physical understanding of CF pathogenesis under various mutations. This will then be followed by an overview of the mechanism behind the restoration of CFTR by small molecule drugs known as CFTR modulators.

In this chapter we will also show how to integrate our molecular perspective from simulations with the collaborators of this thesis who have grown *ex vivo* organoid models from a patient's own cells. Results from these organoid models will be presented in subsequent chapters to prove that small molecule drugs appear to be capable of treating a diverse set of molecular phenotypes. In chapter 8 we will combine these results to argue that this personalised approach to medicine leads us to expect that patients with rare missense mutations will likely benefit from existing treatments, but they are currently excluded from accessing due to an outdated understanding of disease pathogenesis.

The end of this chapter will contain a brief overview of the different experimental techniques which were employed by collaborators of our simulation work to characterise the response of different mutations to different drugs. The language in these sections is kept simple, so the unfamiliar physicist reader can understand some common techniques which they will encounter in the literature surrounding biochemistry and cell biology. The intention here is not replication but comprehension. Hence, an experienced biochemist or a cell biologist would likely find these explanations lacking. For a deeper understanding, the physicist is encouraged to spend more time learning from their experimental colleagues, or even perform some of these experiments themselves if possible.

Another small deviation of this chapter compared to the previous chapters is the chosen structure. In previous chapters we always began with the smallest physically relevant length scale and then worked our way upward. In chapter 2 I showed how we can begin from Schrödinger's wave equation and work to build simulation techniques which let us understand whole biomolecular systems. This was to give an example of the philosophy outlined in chapter 1, where the physicist designs abstract formalisms and integrates them to model macroscopic physical phenomenon. In this chapter I have made a small but conscious deviation from this pattern. Instead of beginning with a discussion of the smallest physically relevant system, the protein which causes disease, we will first discuss some of the clinical realities of living with CF. This deviation from a purely physics inspired structure is made so that we do not lose sight of our goals when studying biophysics.

When one trains to practice medicine they are taught how to conduct themselves ethically, with their patients best interests at heart [248]. Typically, a physicist has no such training, so they may miss the human aspect of medicine by studying disease impersonally [82].

In subsequent chapters we will work with the theoretical and impersonal tools of molecular physics, to understand the cause and treatment of a disease. We do this in the hope that a physical understanding may improve patients' lives. However, when using such techniques we must always remain aware of how such technologies could be misused in living systems. Consequences of the callous pursuit technological advancement can be seen in the participation of physicists in political and moral discourse in the 20th century. Here, the contributions from physicists were varied and not always for the better [249–255]. Should we wish to embark on the enterprise of studying biophysics we have a responsibility to do so ethically. Thankfully, many important questions have been

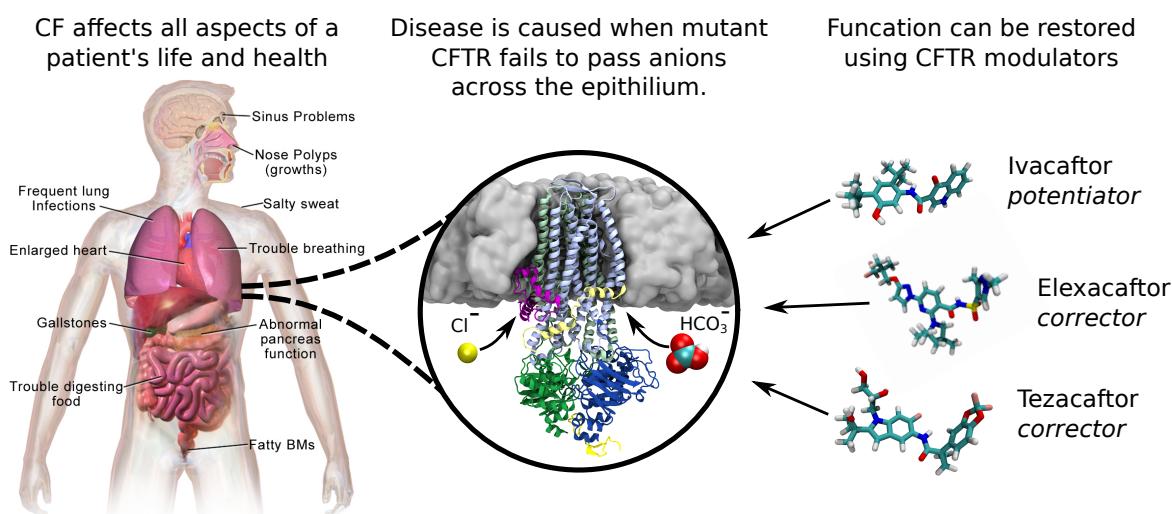


Figure 3.1: Cystic Fibrosis is a Debilitating Disease Whose Cause is Genetic

Cystic Fibrosis causes malfunction in several major organ systems, affecting all aspects of a patient's life and health. The cause of the disease is genetic, when two defective copies of the gene for an anion channel, CFTR, are inherited from each parent, epithelial cells cannot pass chloride or bicarbonate ions, leading to a pathogenic buildup of salts inside epithelial cells. In recent decades, small molecule drugs have been discovered which act directly on CFTR to restore its function. This thesis seeks to understand what kind of defects these drugs are capable of rescuing. We find that they are likely to rescue a wide range of defects.

studied by philosophers, and the reader is encouraged to seek texts on bioethics and also ethics in the use of artificial intelligence [256–261]. These considerations are likely to become increasingly important—as biophysics advances, so too does its capacity for misuse [262, 263].

3.2 Clinical Realities of Living with Cystic Fibrosis

Cystic Fibrosis (CF) is the most common fatal genetic condition in Caucasian populations. Over 162,000 people are estimated to be afflicted globally, with a significant proportion living undiagnosed [264, 265]. Even with decades of research there is no known cure for CF, and patients currently have a life expectancy below 50 [266]. Management of disease symptoms also bears significant financial and emotional costs to patients and their families [267, 268]. From a cellular perspective, the symptoms of the disease are due to the inability of certain cells to regulate their salt content, causing an osmotic imbalance, which leaches moisture away from epithelial membranes (Figure 3.1) [269].

All organs of the body are lined by cells called epithelia. They protect organs from trauma and infection. The most acute symptoms of CF are due to the dehydration of this lining. When dehydrated, fine, motile structures on the epithelium called cilia collapse, rendering them unable to “beat” in order to clear mucus and pathogens [270, 271]. Simultaneously, the mucus which naturally lines the epithelium thickens as moisture is

leached away from it by osmotic pressure.

This thickened mucus has pathogenic effects in several organ systems. The most acute symptoms occur in the lungs, where bacterial colonies grow and form a biofilm. This damages the organ and remains one of the most troublesome chronic complications for CF patients [272, 273]. Frequent infections mean that patients rely on antibiotics and avoid other people with CF in case they expose each other to new or drug resistant pathogens [274, 275].

In addition to these bacterial colonies, the buildup of mucus inhibits the normal function of the lungs, limiting their ability to absorb oxygen. Patients often undergo hours of physical therapy each day to help them clear this mucus [276, 277]. Inhalation of a saline solutions may also help relieve symptoms, as this draws moisture out of the epithelial cells by counteracting the pathogenic osmotic pressure gradient [278]. As Figure 3.2 indicates, much of the increased life expectancy for people with CF has been due to the improved management of this mucus and the populations of bacterium infecting it [266].

In CF care, maintaining lung function remains the primary concern. To track disease progression the most commonly used biomarker is FEV1%. This stands for Forced Expiratory Volume in 1 second. It is a measure of the volume of air a patient is able to forcibly expel in one second, compared to their total lung capacity [271]. As their lungs degrade, FEV1% will over the course of the patients life, meaning respiratory failure is usually the cause of death [279]. Once a patient's FEV1% falls below a certain level (30%-50%) they will usually begin to consider a lung transplant if they are eligible [280].

The adoption of lung transplants has lead to a significant increase to the life expectancy of CF patients but patients may be excluded from receiving a transplant for a variety of reasons [281, 282]. Double lung transplants remain the only option for CF patients as the donor lung would become infected with bacteria from the CF afflicted lung left in their body [266].

CF also causes both acute and chronic issues in the digestive system. Due to the buildup of mucus, the large intestines struggle to absorb nutrients and the pancreas cannot secrete digestive enzymes. These complications can lead to CF related diabetes which afflicts roughly half of adults with CF [283]. To manage and avoid this complication, patients are administered digestive enzymes and consume a high fat diet [284]. A patient whose pancreas does not produce sufficient enzymes to digest nutrients is referred to as “pancreatic insufficient” (PI). This is an important determinant in the severity of disease and a patients’ quality of life [285, 286].

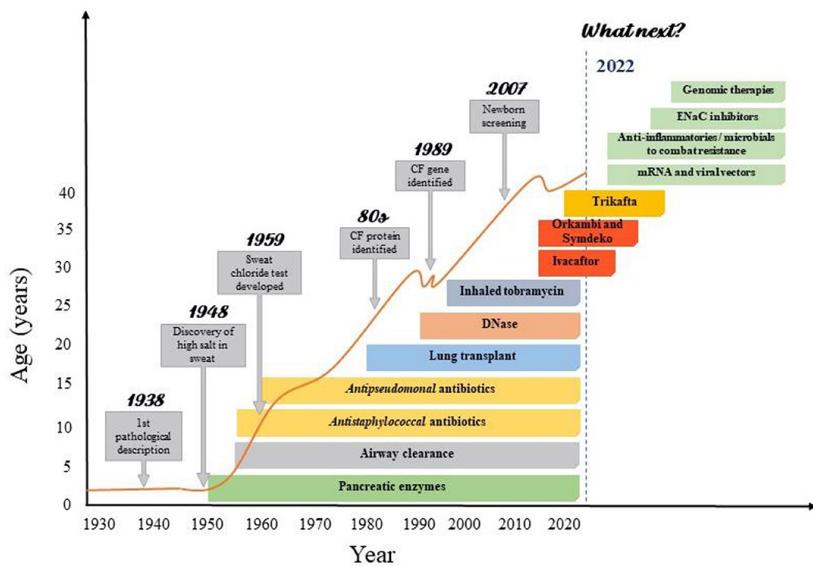


Figure 3.2: CF Clinical Progress

The orange line depicts the life expectancy of CF patients in relation to the development of different treatment. Life expectancy correlates highly with translational research. As CF becomes a chronic condition patients will likely undergo a combination of more and more advanced therapies. Source [287]

Figure 3.2 displays something quite interesting. As patients live longer, CF care must address more disease complications. For example, once pancreatic enzymes were routinely administered, patients began to live beyond two years old [288, 289], which lead to an increasing need to study and manage the role of bacterial infections in the CF lung [290]. With the advent of modulators and refinements to existing therapies, the face of the disease will change. Now that patients are living past 40 we must confront new complications such as bone disorders and infertility [291, 292]. This means that there is still a need to study the cellular and molecular nature of CF, both its root cause and the cause of complications.

3.3 The Misfunction of CFTR Causes Cystic Fibrosis

Now that we have an overview of the clinical aspects of CF, we will now begin our journey upward. As we mentioned in the introduction, we will be collecting molecular details about the ion channel at the core of CF pathogenesis to understand the disease and its treatment. Hence, the next sections will outline in considerable detail, the structure, function and malfunction of the CFTR anion channel.

CFTR is unique. It is a type of protein known as an ABC transporter, designated ABCC7. The ABC super family of proteins bind ATP and use the energy from hydrolysis (breaking) of ATP to pump different substrates across the cell membrane, against a concentration gradient [293]. But CFTR does not act as a transporter like the other members of its family. Rather, it can be thought of as a “leaky pump” since it moves through a gating cycle but is unable to move its substrate against a concentration gradient [294, 295]. It is possible that the evolutionary misappropriation of this protein

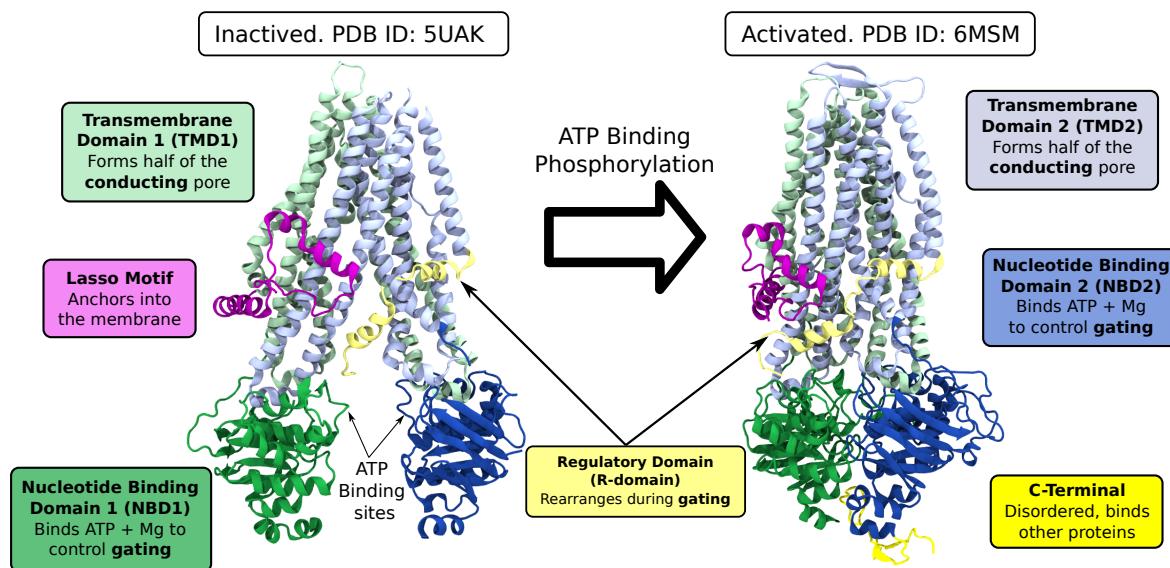


Figure 3.3: CFTR Structure

There are currently two conformations of the human CFTR protein which have been imaged at atomic resolution. These cryo-EM structures have captured hCFTR in an inactivated and an activated conformation. The inactivated state (PDB ID: 5UAK) is neither phosphorylated nor bound to ATP. Observe how the NBDs are far apart and the TMDs are not parallel, forcing a constriction near the top of the protein which does not allow the passage of ions. By contrast, the activated structure is bound to two ATP molecules (PDB ID: 6MSM), demeriting the NBDs and bringing the TMDs into a parallel configuration where they are able to form a pore. Despite having the biochemical characteristics which we would expect from the conducting conformation of this ion channel, the available structure of the activated hCFTR protein exhibits a constriction at the extracellular end of the channel which does not allow chloride ions to pass through. In chapter 7 we will demonstrate enhanced sampling techniques to study a fully conducting CFTR conformation, based on this activated structural snapshot. All simulations presented in this work were based on the ATP-bound phosphorylated human CFTR structure (PDB ID: 6MSM) [297].

from a transporter into an ion channel is the reason for the large number of mutations which cause it to misfunction [296].

3.3.1 CFTR structure and function

As an ATP stimulated ion channel, CFTR sits in an inverted “V” configuration in the cell membrane until it is activated by chemical messengers like ATP and PKA [294, 297, 298]. When it is phosphorylated, the protein changes shape to form a continuous pore across the membrane. This is the conformational change visualised in Figure 3.3.

Chemically speaking, the protein is composed of a single chain of 1480 amino acids. This chain folds into a pseudo-symmetric structure, organising CFTR into 7 domains. In the order of their primary structure they are:

1. The Lasso motif (AA 1-68). The lasso anchors into the membrane and serves as an interaction hub for protein partners such as syntaxin and filamin. These are important in cellular trafficking [299–301].
2. Transmembrane Domain 1 (TMD1, AA 69-376). This domain forms half of the anion conducting pore, and importantly, TM1 and TM6 in this domain form the extracellular end of the pore for anion permeation [302, 303]. Most positive amino acids which directly interact with passing anions are located in this domain [295]. We will look at many of these amino acids in detail in chapters 5 and 7 [304]. There is a pocket in this domain which also serves as a binding site for both corrector class drugs and protein partners such as WNK1, which stimulates the conductance of bicarbonate [305].
3. Nucleotide Binding Domain 1 (NBD1, AA 377-629). This is one of the domains which binds ATP. This domain has a dense concentration of disease causing mutations, including the most common mutation $\Delta F508$ [86]. This domain contains a Walker A motif but lacks a Walker B motif. This causes one of the ATP binding sites to be degenerate, meaning it cannot hydrolyze ATP.
4. Regulatory Domain (R-domain, AA 630-855). This is a disordered domain, containing more than 11 phosphorylation sites [298]. In the inactivated conformation of CFTR, a helical segment of this domain wedges between the TMDs. Binding to PKA encourages this helix to unwedge from between the TMDs, while phosphorylation causes the disordered domain to remain in solution. This process has important consequences for the function of the channel [298, 306–308]. In the open state, the wedge relocates to a position just below the lasso motif [297]. The identity of the wedge fragment will be analysed in detail in chapter 4 [309].
5. Transmembrane Domain 2 (TMD2, AA 856-1168). This domain forms the other half of the anion conducting pore. TM8 in this domain has been shown to bind certain drugs [310, 311]. This helix also plays a critical role in regulating channel gating and anion selectivity [312].
6. Nucleotide Binding Domain 2 (NBD2, AA 1169 - 1450). This is the other domain which helps bind ATP molecules. This domain is home to a conserved Q-loop, which regulates the binding of ATP in ABC transporters. NBD2 contains both a Walker A and a Walker B motif, allowing ATP binding and hydrolysis at what is known as the consensus binding site [313–315].
7. C-terminus (AA 1451 - 1480). This domain is natively disordered but it serves as an interaction hub in WT-CFTR, anchoring CFTR to other proteins in the cell through its PDZ binding domain [316, 317]. Mutations which truncate the C-terminus such as Q1411X cause destabilisation of the protein in the membrane, which can lead to disease [318].

3.3.2 Previous MD Studies of CFTR

Because the first full-length structure of human CFTR was only published in 2017 [319], previous computational studies of CFTR have had to rely on homology models.

More recently these were based on the phosphorylated zebrafish CFTR protein (PDB ID:5W81) [320], or for a longer time, a distantly related bacterial ABC transporter SAV1866 (PDB ID:2HYD) [321–323]. These studies have yielded interesting results but it is often hard to translate their findings to the human protein. Even in the zebrafish CFTR protein, the similarity to the human amino acid sequence is only 55%—and as we will see, CFTR is a highly sensitive protein, its function is poorly conserved under mutations. Note for example, how the action of certain drugs is not conserved in zCFTR [324]. Additionally, the short time scales of the simulations (rarely reaching 500 ns) in these studies opens the possibility that the homology model of the protein remained trapped in local minima, raising questions as to the reliability of their findings.

Despite their short comings, these earlier publications have had some notable findings which we will discuss briefly in the relevant sections of this review. In this section we will review four studies which sought to understand the basic structure and function of CFTR.

Perhaps the most interesting of these studies used path-CV metadynamics to propose an open state of the channel, by steering a zebrafish based homology model toward a fully outward-facing conformation of SAV1866 [322]. The resulting model has several characteristics expected of the open channel, such as the critical R352-D993 salt bridge. However, it fails to create a narrow selectivity filter, which we would expect given results from *in vitro* biophysical experiments [312, 325–329]. Additionally, the free energy landscape in this study does not actually display a minimum at the coordinates which the authors state corresponds to the open state, nor did they validate their open structure with long unbiased simulations (simulations were truncated after just 50 ns). This raises some questions as to whether their proposed open conformation was indeed stable in their model system. However, it was the innovative use of metadynamics in this study which served as the inspiration for the extensive free energy calculations in chapter 7. Hence, it was this ambitious study which motivated us to search and discover a fully open conformation for the human channel.

A more recent study again employed metadynamics with the zebrafish structure, to study the permeation of chloride through the two extracellular portals in zCFTR [330]. This study found a similar conduction pathway, but discovered 10 kcal/mol energy barrier in the selectivity filter to the conduction of chloride. This is is much higher than we would expect from experimental measurements of the single channel [331, 332].

Another study, still under review at the time of writing, attempted to resolve the chloride permeation pathway in the same human, phosphorylated CFTR structure which we have investigated in this thesis (PDB ID: 6MSM) [333]. Here, Zhi Zheng, working in the Reggis Pommés lab, used unbiased MD coupled with an induced 500 mV membrane potential. This strong electric field was used to encourage chloride ions to move through the channel. This study was successful in observing 17 conduction events in 10 μ s of unbiased sampling. Although excellent progress, this is an order of magnitude below the number of conduction events we would expect from this level of sampling—given that the conductivity of the open conformation in WT-CFTR is 6-10 pS [334, 335]. Implications of the results from both these papers will be discussed in

chapter 7.

There is one last article which we will discuss in this section. Here, Corradi et al. looked at the basic function of CFTR, simulating both human and zebrafish models of the channel in the closed and open states respectively [336]. These simulations were used to assess the stability of TM8 in TMD2. These studies were motivated by the fact that all CFTR structures solved by the laboratory of Jue Chen lab exhibit an unusual fold in TM8 [45, 311, 319, 320, 337, 338]. The helix unwinds in the middle of the membrane—a configuration that was initially thought to be energetically unfavourable. This also appeared to conflict with lower resolution structures of chicken CFTR solved by another laboratory [339]. The generated surrounding significant controversy in the literature [293, 296, 310, 312]. However, most recent studies, including the simulation study by Corradi et al., have supported the TM8 model found by the Chen lab [312, 336]. The given reason for distrust of this conformation of TM8 were motivated by the use detergents by the Chen lab in solving the CFTR structure [310]. However, careful work by the Fay lab to reconstitute CFTR in a nano disk confirms that indeed, TM8 takes on a kinked confirmation in its native lipid environment [340]. Hence, these controversies should now be considered resolved and the Chen structure vindicated, while keeping in mind that TM8 appears to exhibit considerable flexibility *in vitro*.

In summary, with some exceptions, it is difficult to draw from the previous *in silico* literature of CFTR because studies have lacked an appropriate structure as a starting point. Previous CFTR homology models have largely failed to capture critical molecular details of the full human CFTR protein. Hence, these studies mostly served as a proof of concept and inspiration for further work, rather than giving results which can be directly applied to the hCFTR protein.

3.3.3 Catalytic and Conformational Changes During the Gating Cycle

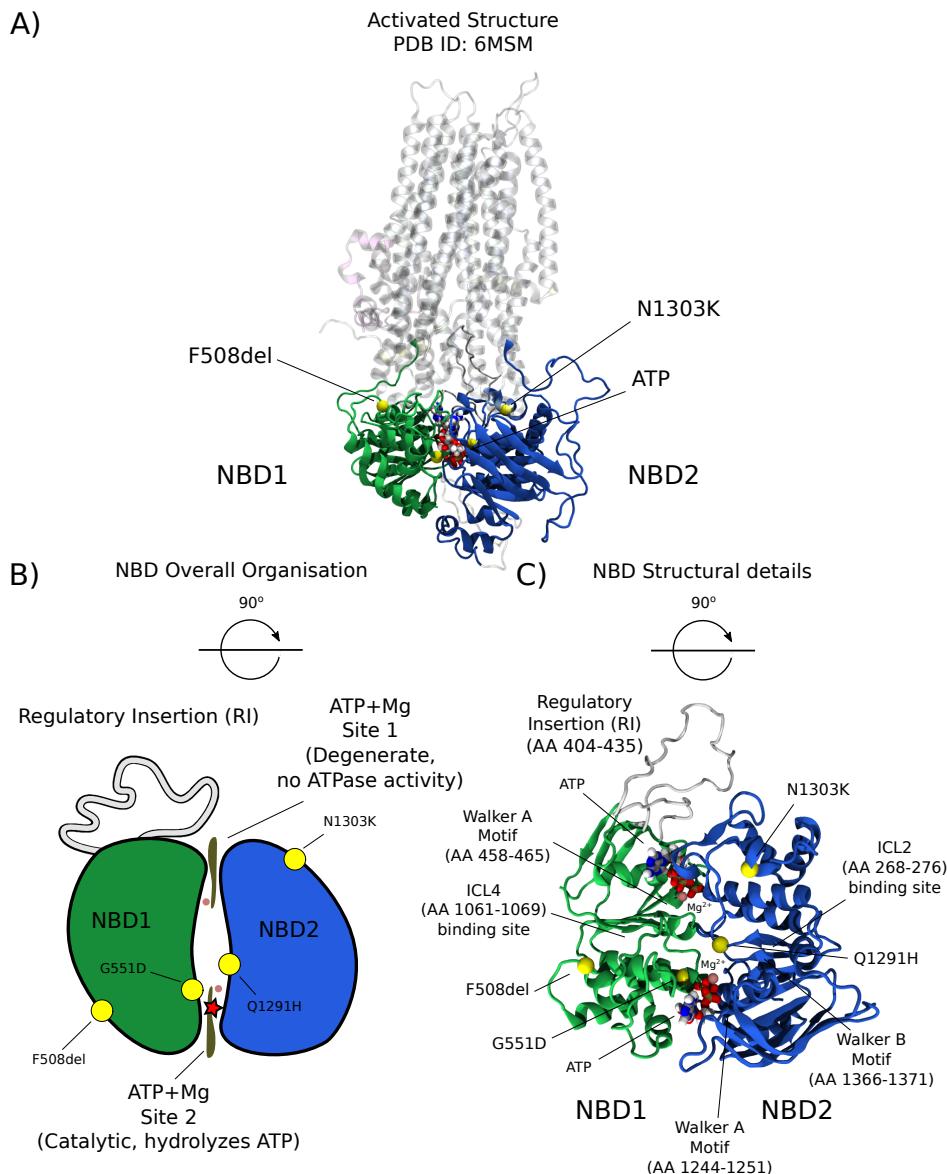


Figure 3.4: NBD Organization

A) The NBDs are critical to the gating cycle of CFTR. They rest below the TMDs where they bind and hydrolyse ATP in order to regulate the gating cycle. B) Each NBD contacts both ATP molecules. This configuration is known as a head to tail dimer and is common amongst ABC transformers. One of the unique aspects of CFTR is the fact that only one of the ATP binding sites is capable of breaking down its ATP molecule through hydrolysis [341]. When ATP is hydrolysed at site 2, the channel is allowed to close [296]. The regulatory insertion is another unique feature of CFTR and has been shown to increase NBD1 stability when deleted. C) Molecular details of how the NBDs bind ATP at the two binding sites. Here we have highlighted several molecular details. The Walker A motifs are on both NBDs and are responsible for stabilising ATP in a conformation which allows hydrolysis to produce ADP and a phosphate group [342].

The Walker B motif on NBD2 is directly responsible for mediating hydrolysis, by allowing the movement of the magnesium cofactor [343, 344]. Mutation of E1371 in Walker B is sometimes engineered to kill catalytic activity and lock the channel in an open state [297, 341]. Note the placement of both G551D, a gating mutation, and Q1291H near site 2. The most common disease causing mutation $\Delta F508$ is located on a helix of NBD1 which buries itself into a hydrophobic pocket. N1303K is located at a roughly symmetric position on NBD2.

The conformational transition from the inactive to active state differs significantly in CFTR compared to other ABC transporters. A transporter acts as a gate to a substrate—there is never a continuous path to allow the substrate to diffuse through the protein. However, since it is a channel CFTR gives rise to a continuous pore [295].

CFTR's gating cycle is dominated by the action of the NBDs and the R-domain. The NBDs are largely similar to other transporters in the ABCC subfamily, so we will discuss them first.

NBD1 and NBD2 each have two binding sites for an ATP+ MG^{2+} complex. This facilitates their dimerization in what is known as a “head to tail” configuration where both NBDs make contact with both bound ATP molecules [345] (Figure 3.4). Usually, in an ABC transporter the NBDs would contain a “Walker A motif” to help them bind ATP *and* a “Walker B” motif to catalyse the hydrolysis. A unique aspect of CFTR is that NBD1 is missing its Walker B motif, meaning one of the ATP binding sites is incapable of hydrolysing ATP. This ATP binding site is sometimes referred as “site 1”, “the degenerate site” or “the non-consensus site” and it plays an important role in stabilising the open state of the channel [87, 346].

The other ATP binding site is referred to as “site 2”, “the catalytic site” or the “consensus site.” The hydrolysis of ATP at this site triggers the closure of the channel. Sometimes, for the purposes of experiments, the catalytic activity of site 2 is intentionally halted through an engineered mutation. For example, E1371Q-CFTR is studied in structural determination experiments as it exhibits a much longer lived open state than the WT channel. This allows more snapshots to be taken of the open structure during cryo-EM experiments [345, 347–349]. Indeed, the protein structure, which was the source for our work, carried this mutation, and it had to be corrected to match the WT amino acid sequence before simulation [297].

The above outlines the generally agreed upon chemistry about the action of the NBDs and gating. Beyond this, there are some details about the gating cycle which are still up for debate. There are two competing models which give different predictions for how tightly coupled the dimerization of the NBDs is to the opening of the channel [346]. The first proposed model is called the “strict coupling model”, which suggests that the NBDs and the TMDs rotate as rigid bodies, to move between the active and inactive states. This model would indicate that ATP binding and NBD dimerization is tightly coupled to the opening of the channel [350].

The second model is called the “energetic coupling” model. This predicts that there are both closed and open states which are accessible while the NBDs are dimerized in the post hydrolytic state, as well as closed and open states while the NBDs are *not*

dimerized [350]. This means we would expect the binding of ATP to be loosely coupled to the opening of the channel. It would seem that, although more complicated, the second model is both more consistent with recent findings in the literature and would also explain some of the other controversies surrounding the structure and function of CFTR, such as the flexibility of TM8 [339]. These considerations are not critical for the work in this thesis but considering them in the future may aid in targeted drug discovery efforts for CFTR, particularly for potentiator class drugs.

Apart from the NBDs, the most important domain for the gating of CFTR is the R-domain. This disordered domain is unique to CFTR. In the inactive state the R-domain wedges between the TMDs, forming a plug and not allowing them to rotate to form a pore. Intermittently, Protein Kinase A (PKA) will bind to one of many segments on the R-domain coaxing the coils of peptide out into solution [298]. PKA may then phosphorylate specific sites on the domain, causing it to remain in an unplugged configuration. It is currently unknown exactly how many phosphorylation sites there are in the R-domain, but more than 11 have been investigated [298].

When the R-domain is phosphorylated it no longer rests between the TMDs. Rather, the amino acids which form the plug relocates to a site below the lasso motif. This was one of the main findings from chapter 4. Further, in chapters 4 and 6 we demonstrate how the destabilisation of this binding site results in a gating defect.

3.3.4 Anion Selectivity

CFTR does not permeate any cations, but there are a set of anions which it may conduct. The conduction pathway of the channel can be split into two regions. The inner vestibule which is wide and covered with basic, positively charged amino acids and the selectivity filter which is formed by the narrowing of the protein in the upper part of the channel. These regions are hotspots for disease causing mutations, with most mutations occurring on TM6 in TMD1 (Figures 3.6 and 3.5) [86]. The selectivity filter can also be split into two sections, a "conductivity selectivity filter", which binds the anions and a "permeability selectivity filter", where dehydration occurs. We will discuss these in more detail in chapter 7 [325].

Compared to cation channels like Gramicidin and KcsA, CFTR is only weakly selective—permeating a large set of anions with varying radii and geometries (Figure 3.5) [335, 351, 352]. CFTR appears to act like other anion channels, where lyotropic (low solvation energy) anions pass more easily than kosmotropic (high solvation energy) anions. This would suggest that partial dehydration of the anions is likely during conductance, indicating a narrow pore [353]. The radius of a dehydrated chloride ion is 1.7Å while a hydrated chloride ion has a radius of 3.2Å [354]. As noted in its source publication, the 6MSM structure which we have studied has a constriction of 1.2Å at its narrowest point. Hence, even if chloride were to be fully dehydrated during conduction, there would have to be some level of conformational change in the protein. This was the motivation for the extensive free energy calculations undertaken in chapter 7.

F337 is the most important amino acid for selectivity, with the F337A mutation leading to a near complete loss in selectivity [355]. Similarly, L102A and T338A mutations lead

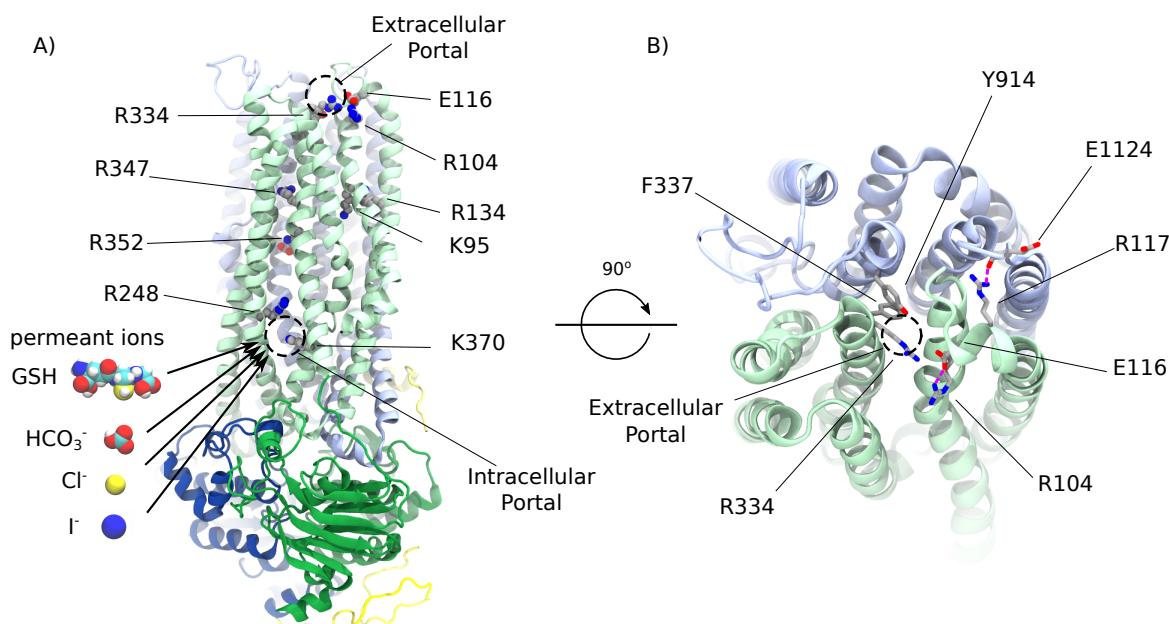


Figure 3.5: Chloride Passage Through CFTR

A) Ions enter through an intracellular gate, which is lined with positive amino acids. The subsequent inner vestibule is also lined with amino acids which chloride loosely associates with as it moves through the channel. B) A top down view of the extracellular portal. We have highlighted some hydrophobic amino acids which play important roles in the selectivity of anions. Pictured is the only available structure of the activated human CFTR channel [297]. Unfortunately, this conformation is not sufficiently open to conduct ions. We will analyse an open conformation in chapter 7, which we predicted through simulations.

to a loss in selectivity but interestingly an increase in conductivity, likely reflecting their position in a narrow section of the pore [356, 357].

Bicarbonate (HCO_3^-) is known to have roughly 25% the permeability of chloride through the CFTR [358]. Some interaction partners such as WNK1 have been shown to influence the selectivity of the channel [305, 359, 360]. The permeation of bicarbonate is very important physiologically. If a CF-causing mutation permeates bicarbonate, there is a high likelihood the patient will be pancreatic sufficient. Increasing bicarbonate secretion through CFTR is being explored as a possible therapeutic strategy [361], but this mechanism has been largely ignored by theoretical explorations of CFTR, with one notable exception [331]. We will attempt to address this gap in the literature in chapter 7.

3.3.5 Different classes of Cystic Fibrosis Phenotype

As the above shows, CFTR is a complex protein, and as we will now see, it can misfunction in a number of ways. Based on the way the mutation presents in *in vitro* assays, disease causing mutations are classified into 6 common classes. Ultimately, the aim of this thesis is to demonstrate that at the atomic level, there is much more nuance to these classes, and a mutation can have features of multiple classes at the same time. This will be demonstrated in chapter 8.

This has important implications for the treatment of the disease. Due to the heterogeneity between patient responses to therapies, there is a growing interest in personalising the treatment of disease, and mapping the molecular effects of mutations can inform this process of therotyping [362, 363]. As patient specific treatment evolves, these classes will likely become less relevant, serving as illustrative tools only to communicate at a higher level what is going wrong with the CFTR protein. The specific choice of treatment will depend on a number of personal factors. However, the canonical classification of mutations is our starting point and it reads as:

- **Class I** No functional protein. Under these mutations no full CFTR polypeptides are translated, due to either problems with the transcription of mRNA or a premature stop codon truncating protein synthesis early. This latter case means the resulting peptide is missing key domains and cannot function.
- **Class II** Folding defect. These mutations cause the translated peptide to misfold into the incorrect tertiary structure. This can inhibit the protein's journey as it is trafficked to the cell membrane.
- **Class III** Impaired Gating. Here the mutation inhibits the ability of the protein to transition from the closed to the open state.
- **Class IV** Decreased Conductance. These mutations cause a barrier in the energy landscape of the CFTR chloride conductance pathway.
- **Class V** Less Expressed Protein. For whatever reason, the cells do not express as many CFTR proteins as they need to.
- **Class VI** Decreased Functional Lifetime. In this case the proteins may fold and

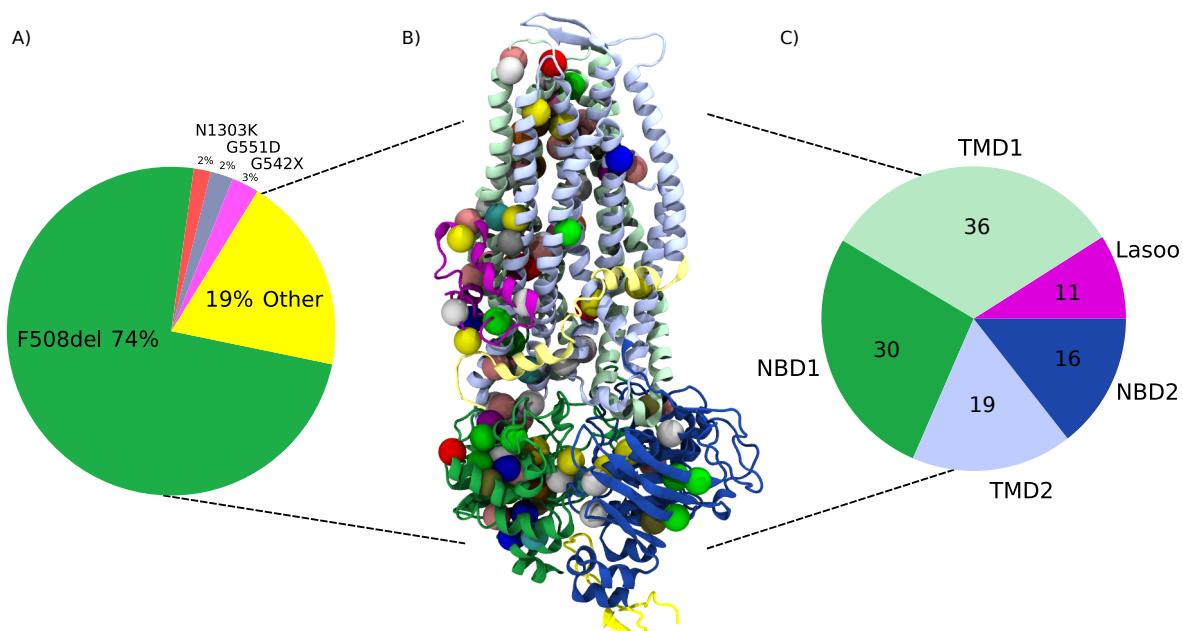


Figure 3.6: Rare Mutations Occur in All parts of the CFTR Protein

A) The reported frequency of all disease causing alleles (not just missense mutations) in the CFTR2 database. At the time of writing there are 402 recorded different disease causing mutations. A patient with CF will carry two of these mutations. 1 in 25 people of Northern European descent are a carrier for CF and are themselves at increased risk for many health problems [364, 365]. B) The location of all 112 known disease causing missense mutations mapped on the CFTR protein itself as spheres. In addition to those visualised, here there are many variants of unknown significance (VUS) which may cause disease but are so rare they have not been clinically studied. C) The incidence of CF causing missense mutations in the different domains of the CFTR protein. TMD1 and NBD1 have the highest concentration of disease causing mutations. Likely due to the former's role in ion permeation and the latter's role in gating [86]. There are no confirmed CF causing missense mutations in the C-terminus or the R-domain. As patient registries are set up across the world, it is likely that more rare CF causing mutations will be discovered in the future.¹

gate correctly but they are not stable in the plasma membrane so they degrade too early.

Although useful from a clinical and experimental standpoint, in reality, these classes do not reflect the molecular nature of a mutation. Mutations may in fact belong to many categories at once, to differing degrees. This has lead to a more granular approach in more recent literature [366]. In particular, through this thesis we will demonstrate that membership of one of these classes can be due to a range of molecular phenotypes. Hence, this classification is becoming increasingly outdated and there is a need to understand the unique, sometimes overlapping, molecular phenotypes for each mutation. Further, through our molecular simulations, we will see that in fact CFTR modulators appear capable of treating many mutations with unique modes of pathogenesis. This is the purpose of chapters 4-7. In each chapter we will analyse a specific mutation in detail and then use *in vitro* techniques or sources from the literature to demonstrate that the molecular defect introduced by the mutation can be rescued by existing CFTR modulators.

A molecular understanding has previously been outlined for certain mutations, though rarely with the advantage of the full hCFTR structure or extensive statistical sampling. For example, [367] performed molecular modelling on a structure SAV1866 and used short unbiased simulations as well as simple free energy estimates in order to demonstrate the deleterious nature of the G551D and $\Delta F508$ mutations.

The findings from this study are remarkably consistent with the findings of more recent studies which have the benefit of more structural information about the protein [368]. This latter study used isolated structures of NBD1 and NBD2 (PDB IDs: 2PZE [369] and 6UK1 [370], respectively) in order to predict the $\Delta\Delta G$ change to the folding energy of CFTR from disease causing mutations. The authors correlated their results with *in silico* prediction functions from Rosetta and find good agreement between *in silico* predictions and experimental measurements. Although this study does not shed light on molecular details of malfunction, it will likely be a critical resource for the grouping of mutations in theratypes.

To date, there are two studies which have used a similar approach to us to study mutations. They also attempted to understand the molecular defects of rare mutations *in silico* and then combining these findings with patient specific platforms to demonstrate their rescue by CFTR modulators. The first of these studies used the zCFTR structure (PDB ID: 5W81), in order to study the ultra-rare W361R mutation [371]. Here, the authors found that a lipid tail binds into a hydrophobic pocket formed by the W361 amino acid. This lead to the expectation that the W361R mutation would cause a class II folding defect. Unfortunately, the authors did not explicitly simulate the effects of the W361R mutation, though the nature of the folding defect they would expect to see is likely not accessible on simulation time scales. It is intriguing that this mutation responds to corrector compounds as the W361R mutation destroys an amino acid that would appear critical to the binding of the corrector compounds [372]. Hence, this publication indicates that, like potentiators there is likely more than one site on CFTR which is responsible for the mechanism of action for correctors.

A second study addressed the intriguing mutation P67L on the lasso motif [373]. This

mutation occurs in a hotspot of CF-causing mutations on the lasso elbow [372, 374]. Here, the authors were able to rely on the structure of activated hCFTR, to model changes in NBD2. However, the simulation methodology in this paper is questionable. The authors simulated hCFTR for less than 200 ns and claim to have discovered allostery between the lasso motif and NBD2. Such an effect would likely take orders of magnitude more simulation time to observe because of the distance between these domains in the simulated structure. The simple statistical test employed by this paper in Figure 7, on a single simulation replicate is also not fit for purpose [150]. However, the pharmacological and biochemical results from this publication are certainly successful in shedding light on the molecular nature of this mutation.

Although the authors of these works have seen the deleterious effects of mutations through simulations, and demonstrated their rescue with pharmaceuticals, these prior studies are not as rigorous as the work we have undertaken for this thesis. None of these studies simulated CFTR for more than 1 microsecond and none of them have employed free energy calculations. However, these studies validate the approach we will use for subsequent chapters and further, allow us to demonstrate the power of using MD and free energy calculations as computational microscopes to investigate CFTR malfunction.

The main findings of this thesis are in keeping with the theme of the studies we have outlined above. The molecular defects observe in each mutation we studied was largely:

- **I37R** is a mutation which occurs on the lasso motif. In chapter 4 we explored how this novel mutation causes a class III gating defect by giving rise to a pathogenic interaction with a fragment of the R-domain. This fragment was poorly resolved in the cryo-EM structure of CFTR, and we first had to identify it using MD. Our structural predictions were then confirmed upon the release of AlphaFold2 [69].
- **R352Q** neutralizes a positive charge on TM6, along the chloride conduction pathway in the inner vestibule. In chapter 5 we used umbrella sampling and unbiased MD to demonstrate the misregulation of the conduction pathway in this mutation, showing that it is a class IV conductance defect.
- **S945L** occurs in TM8 on TMD2. In chapter 6 we discovered a stable network of hydrogen bonds which links S945 to Y852 on the R-domain. We hypothesised that the replacement of the polar serine sidechain with hydrophobic leucine might cause a perturbation to this section of the R-domain. Due to the kinetic barrier from the surrounding lipids, we again employed umbrella sampling to explore how the perturbations this mutation causes to the R-domain results in both a folding and a gating defect.
- **R334W** occurs at the extracellular pore, at the end of the selectivity filter on TM6 in TMD1. In chapter 7 we used multiple walker OPES-metadynamics in order to dilate the existing structure of human CFTR (PDB ID: 6MSM) and study the effects of the R334W mutation on chloride conductance.

Through our MD simulations, we found that each of these rare mutations has a unique molecular fingerprint. They appear to cause CFTR to malfunction in unrelated ways. Further, in each of these cases, we have either presented *in vitro* patient specific assays

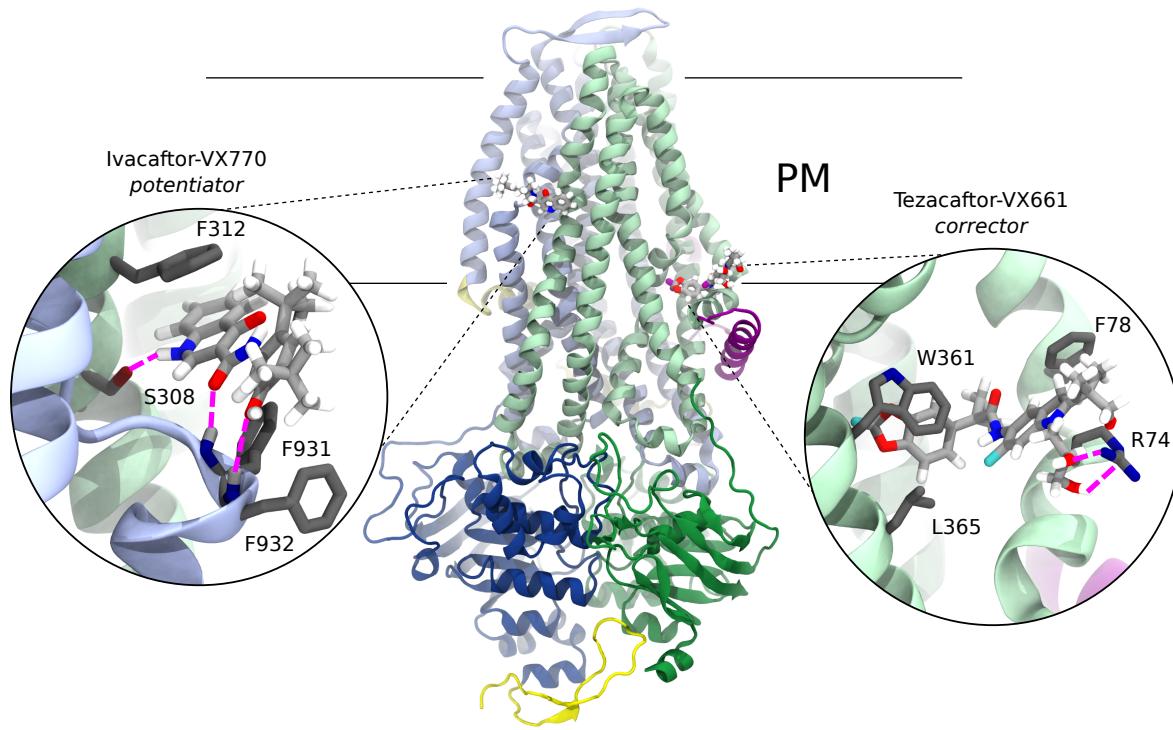


Figure 3.7: Modulators Bound to CFTR

The structurally imaged binding sites of potentiators and type I correctors [311, 372]. Potentiators, in particular, are proposed to bind to other sites as well [311, 375]. Hydrogen bonds are labelled with purple dotted lines. Note how both visualised binding sites require substantial contributions from the surrounding lipid bilayer. The presence of these lipids plays an important role in the binding kinetics and energetics of these drugs [376]. Future molecular studies should account for these nuances.

or surveyed the literature to demonstrate that pharmacological rescue of these rare mutations is possible using existing CFTR modulators. It is then remarkable that these unique molecular modes of pathogenesis all to respond to CFTR modulators. It is these results which inform the conclusions in chapter 8—where we propose a model for the action of CFTR modulators. This model will direct the recommendations in chapter 9 for further studies of CFTR, and argue that many patients who are currently excluded from access to modulator therapy are likely to benefit from these life changing drugs.

3.4 Modulators Act Directly on CFTR to Restore its Function

Since CFTR is at the heart of CF pathogenesis, it has been pursued for decades as target for therapeutics. Through high throughput *in vitro* screening, four compounds which act directly on CFTR have been clinically approved for the treatment of CF.

These drugs are known as CFTR modulators. The drugs currently on the market fall into two categories. Correctors, which aid CFTR to fold into its native state and potentiators which help the channel reach the fully open, conducting state once the protein has folded correctly and integrated into the membrane. Since their discovery, several *in vitro* biophysical experiments have investigated the mechanism of action and binding site of these compounds [372, 375–379].

Due to the excitement surrounding these compounds, there have been a few simple *in silico* studies of CFTR modulators [380–383]. When not performed carefully, *in silico* docking studies for small molecules may produce very crude and inaccurate findings. This was particularly apparent when reviewing the large number of docking studies performed during the coronavirus pandemic of 2020–2022 [384–386]. It is then fortunate that previous MD and docking studies of CFTR modulators have been largely consistent with subsequent structural and biochemical experiments [311, 372, 375, 378, 387]. Even though these studies are limited in the impact of their findings, their success can be attributed to the close integration of these *in silico* techniques with biochemical assays. This would indicate some promise for the approach of this thesis and future *in silico* studies of CFTR—as future studies have access to even more computational power and structural information.

An added benefit of modulators is that they have demonstrated promise in improving not just the most acute symptoms of CF (lung function), but they may be able to treat or even relieve deleterious CF associated complications. In some cases, modulators have been found to relieve pancreatic insufficiency and CF related diabetes [388–390]—while there is ongoing study into the possible benefits of modulators on bacterial infections [391]. This highlights the importance of treating the root cause of a disease, when we are presented the opportunity, and not just its symptoms.

3.4.1 Correctors

Corrector class modulators are able to treat Class II folding defects. This means that there will be more CFTR channels at the cell membrane, conducting chloride and relieving the buildup of osmotic pressure. The most common type of correctors are known as type I correctors, which help the CFTR protein fold into the correct structure by binding to a hydrophobic pocket in TMD1 formed by TMs 1, 3, 4 and 6. This binding mode has been directly visualised by cryo-EM imaging with atomic resolution [372]. This binding mode was predicted earlier by circular dichromism [392] and fluorescence experiments found that an isolated construct of TMH3 and TMH4 were more likely to fold correctly in a model membrane in the presence of corrector compounds [379].

In combination, these studies give us a strong base to understand the mechanism of action for corrector compounds. The correctors tezacaftor (VX-661) and elexacaftor (VX-445) form two components of the triple therapy Trikafta [393, 394]. Tezacaftor is a type I corrector while elexacaftor has been found to have some potentiation capacity on CFTR, leading it to be classified as a type III corrector. A review of all types of correctors can be found in [395]. Further work will aid in the creation of new compounds to refine our exploitation of current correction strategies.

3.4.2 Potentiators

Potentiator class drugs bind to CFTR in order to increase the number of channels which are conducting chloride in the cell at any one time. This helps to balance osmotic pressure over the epithelium and relieve symptoms [378, 396–398]. However, there is some uncertainty surrounding the molecular details of their mechanism of action.

There is consensus that potentiators bind directly to CFTR, and that they increase the likelihood that the channel is in a conducting state. Indeed, some potentiators such as VX-770 (ivacaftor) rescue CFTR with picomolar affinity [376].

Controversy arises about precisely *where* these drugs bind in order to rescue CFTR. Different studies have found different binding sites for the same drugs, using mutagenesis and other methods [311, 375, 399].

Careful investigation of different potentiator molecules give hints for solution to these problems. For example, GLPG1837 has not been approved in a clinical setting but through *in vitro* experiments it has been demonstrated that it is more efficacious at potentiating CFTR even though it has lower affinity for the CFTR protein itself [400]. This would indicate that the highest affinity binding pocket does not produce the greatest modulation. A careful investigation of different binding pockets can be found in [375], while the binding kinetics are characterised in [376] and [378].

More work is needed to resolve the molecular mechanism which results in the highest clinical effectiveness of these drugs. Some of these issues will be discussed in chapter 9 where we will suggest some *in silico* strategies to address them.

The competitive nature of certain potentiator and corrector drugs opens the possibility that specific genetic defects may be optimally rescued by specific combinations and doses of both corrector and potentiator compounds [376]. This means there may be much to gain from the patient specific approach like the one employed for this thesis.

The elucidation of the binding site for potentiators has been attempted with unbiased MD using the hCFTR structure (6MSM) [399]. This study used an octanol slab as well as a model POPC membrane in order to accelerate the diffusion of VX-770 in the transmembrane portions of CFTR. With 10 μ s of unbiased MD sampling, the authors discover 7 binding poses for the potentiator drug. The authors combine their results with a photo-label and mass spectrometry results which will likely play a critical role in future development of potentiators [377]. They use these innovative tools to conclude that both a site on ICL4 and the site visualised by cryo-EM experiments play a role in potentiating CFTR activity. This is consistent with other *in vitro* experiments [376].

Unfortunately, for the *in silico* component of this study, none of the binding sites sampled in MD corresponded to the binding mode visualised by cryo-EM experiments [311]. The authors speculate that this may be because the cryo-EM site is in fact low affinity. However, it is likely that even though 10 μ s represents a considerable investment of computational resources, it is simply not enough sampling to sample ligand binding and unbinding. The kinetics of a membrane facing drug binding site occur on the order of milliseconds [401]. Hence, free energy calculations would be needed for such an endeavour and we will propose such these studies in chapter 9.

Read Through Compounds

The above two categories are most effective for class II, III and IV defects. However, if a full length CFTR protein is not synthesised by the patient, these compounds cannot help. This has given rise to interest in another class of drug known as read through compounds. Roughly 10% of known disease causing mutations result in a premature stop codon in the mRNA of the CFTR gene. This means protein synthesis is stopped early, and the resulting protein will be missing critical domains.

Since there is no full length CFTR protein synthesised by cells carrying these mutations, potentiators and correctors are unable to directly assist. This motivated the development of read through compounds which allow the continued synthesis of the protein past the premature stop codon [402]. At the time of writing no read through compounds have been clinically approved but efficacy has been demonstrated in pre-clinical models [403]. Since correctors and potentiators also modulate the behaviour of WT-CFTR it is likely that correctors and potentiators would be given in combination with read through compounds and other therapies as a complimentary treatment.

Alternative and Complimentary Therapeutic Strategies

As the root cause of disease, CFTR will continue to be an important therapeutic target. However, there will also be increasing value in the therapeutic benefit derived from manipulating other protein systems—both as a complimentary strategy to the modulation of CFTR and to the treatment of complications. This is especially pertinent when we consider that some patients do not respond to CFTR modulators, even when they share the genotype of patients who do. These patients would benefit greatly from complimentary therapeutics [404–409].

As mentioned in section 3.2, one of the most troublesome complications of CF is the chronic infection experienced by patients. Unfortunately, modulators have not been able to reliably alleviate this complication [262]. This is because these bacterial colonies form a protective biofilm and thus become very difficult to eradicate with traditional antibiotics. To this end, bacteriophage therapies are being pursued as a complimentary strategy to those above [410].

There is also exciting work on gene and cell based therapies to directly deliver properly functioning CFTR proteins into the bodies of CF patients [411]. This is being investigated through the use of different advanced vectors, such as the retro virus lentivirus. With these vectors, WT-CFTR is usually added *ex vivo* and then the treated cells are transplanted back into the patient. However, it may soon be possible to edit the genome of a CF-afflicted embryo and cure the patient before they are even born using a system like Crispr-Cas9 [412]. This latter technique has been misused in the past [262], and the reader is again referred to section 3.1 for texts on bioethics.

3.5 Patients with Rare Mutations Struggle to Access Modulator Therapy

CFTR modulators have proven to be a breakthrough in the treatment of CF. However, there is strong motivation behind our focus on rare mutations in this thesis. Roughly 50% of Cystic Fibrosis cases are caused by a homozygous mutation, $\Delta F508$, and 90% of patients carry at least one copy of this mutation. At the time of writing, many jurisdictions such as Australia and the EU have approved a drug combination therapy known as Trikafta for patients carrying one copy of $\Delta F508$. Given that these drugs act directly on CFTR, we would expect that other mutations would also respond to certain modulators.

Thus, the exclusion by regulators of patients with rare mutations leaves a significant part of the CF afflicted population without access to this life saving medication, with many more excluded due to the extreme cost of the drugs themselves [393, 394, 413, 414].

Traditional clinical trials are also not possible for rare mutations, as there may not be enough patients carrying one mutation to study them in the same trial [415]. Hence, we aim to demonstrate that a larger section of the CF population is likely to respond to modulators, particularly those carrying missense mutations. Eventually, this work will contribute to a clinical method known as therotyping, where a granular understanding of CF pathogenesis will allow a personalised choice of treatment.

The treatment of rare mutations has particular significance in CF. Not only would patients be otherwise left behind, but by studying rare mutations, outcomes can be improved for all people with CF. For example, potentiator class drugs were discovered by high throughput molecular screening against the less common class III gating mutation G551D [416]. By studying rare mutations we can gain a more complete understanding of the nature of the disease and so improve therapies for everybody. This approach will become more important in the future as CF is uncovered in regions where it is rarer. In these populations there is also a broader spectrum of CFTR mutations, meaning patients in these regions are more likely to have a rare form of disease [417–419]. As patient registries are launched in these populations, we will discover more rare forms of CF. This expected trajectory of CF epidemiology makes the approach of this thesis critical, such that all people with CF can benefit from the advances being made by translational research [287, 418].

The earlier patients are able to benefit from modulator therapy, the better their prognosis [420, 421]. It is also well documented that patients with the same genotype may respond differently to the same modulators [404]. Hence, by validating a patient specific approach to rare mutations, we can build a platform to choose the right modulators for all patients.

These considerations also show that it is critical for more modulators and complementary therapies to be developed—in order to give all patients more options and better outcomes [404].

3.6 Patient Derived Organoids: A Pre-Clinical Model to Assess Modulators

In chapter 2 we constructed a model based on interacting atoms to study biomolecular systems. Similar to this philosophy, medical researchers also seek model systems to understand the function of organs. In CF research, this has lead to the development of techniques where samples of epithelial stem cells are taken from patients and grown into structures which mimic the function of either the lung or the gut [422, 423]. These are known as patient derived organoids.

This is possible due to a population of adult stem cells which live in the epithelium. These cells maintain their ability to differentiate into a variety of cell types (a property known as pluripotency) [424].

To grow organoids, epithelial stem cells are scraped from a patient's nasal passages or taken during a rectal biopsy. These samples can be used to grow either lung or intestinal organoids respectively [425, 426]. The morphology of the organoids can be confirmed through immunofluorescence microscopy [425, 427]. Organoids are excellent model systems, but like all models, they are incomplete. They often lack specialized cell types and miss much of the complexity which might arise from interactions with other organs [428].

Nevertheless, these models have the potential to be a powerful tool which can overcome the need for clinical trials in the case of rare forms of CF [429]. These trials currently stand as a complex, bureaucratic and time consuming barrier to patient care.

Adult stem cells in the epithelium are preferable to other sources of stem cells. This is because other sources, which might be easier to harvest such as iPSCs from the skin or white blood cells, would require complex, time consuming protocols to grow into organoids [430]. Already, the differentiation and expansion of epithelial samples into the organoids studied here take one month [426].

In the case of CF, this technology allows the construction of a scalable, patient specific platform where the response of a patient's own tissues can be tested to determine their response to a given treatment [426, 431].

It is hoped that these pre-clinical models will allow more patients with rare CF causing mutations to access modulators. This has given rise to an exciting prospect of a practice known as therotyping, enabling clinicians to make a personalised prediction of which therapies will best serve a patient [304, 309, 432, 433]. This thesis seeks to demonstrate that integration of *in silico* simulations into the process of therotyping can further increase the predictive power gained these *ex vivo* models. This is achieved by demonstrating that these *ex vivo* models are capable of resolving CF pathogenesis and treatment for a diverse set of CFTR defects.

Subsequent chapters will make use of these organoids by testing them with a set of *in vitro* assays in order to characterise a patient's response to modulator therapy. For the physicist reader, a brief explanation of these assays is given below.

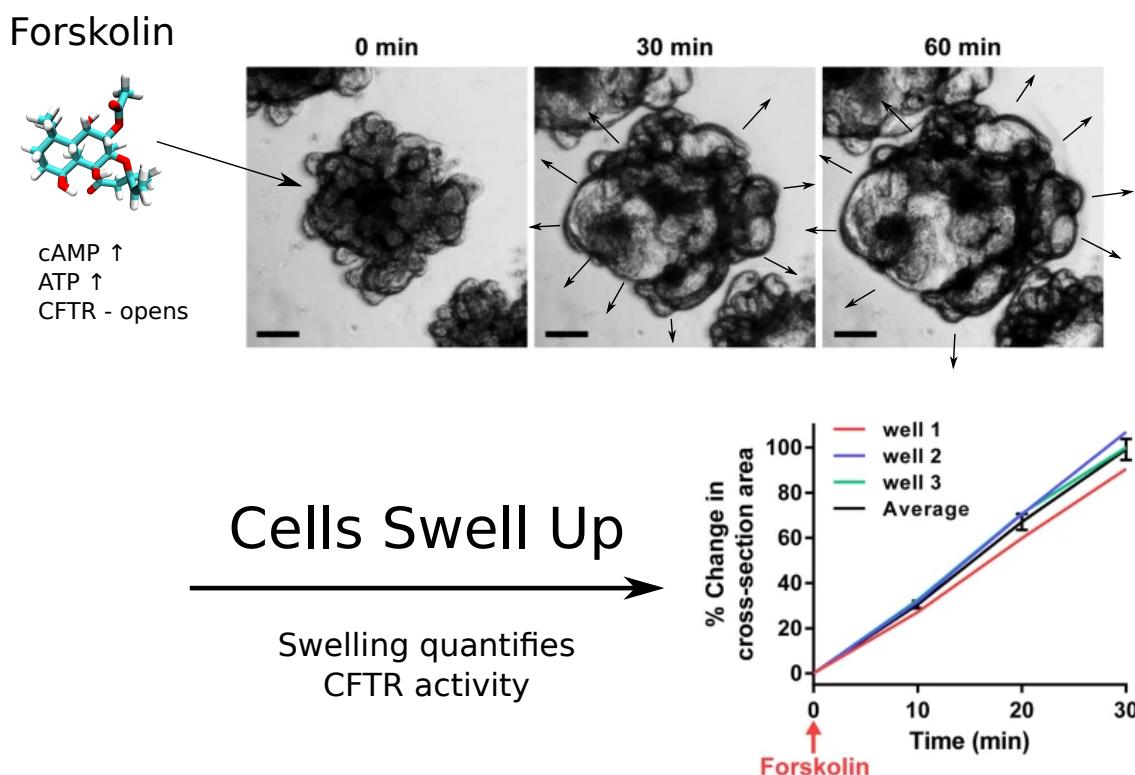


Figure 3.8: Forskolin Induced Swelling

Upon exposure to Forskolin, epithelial cells begin to produce cyclic Adenosine Mono Phosphate (cAMP). This causes an increase in intracellular ATP which will trigger functional CFTR channels to open. This will allow an influx of water into the cell, causing them to swell. The amount of swelling can then be used to quantify the amount of properly functioning CFTR channels in the cell.

3.6.1 Forskolin Induced Swelling to Assess CFTR Function

Forskolin Induced Swelling (FIS) assays can be used to easily characterise the amount of functioning CFTR channels in a patients cells [434]. When epithelial cells are exposed to a chemical known as forskolin they begin to rapidly produce cyclic AMP (cAMP, a precursor to ATP in a cell) [435]. The presence of ATP activates the CFTR ion channels, causing the cells themselves to swell due to the intake of moisture. When performed under a microscope with patient derived organoids, this swelling allows cell biologists to easily quantify the activity of CFTR within a cell in a variety of conditions, such as the presence of drugs.

Hence, by comparing the amount of swelling of organoids in the presence or absence of modulators, we can quantitatively assess the response of a patient's cells to modulator therapy (Figure 3.8).

3.6.2 Electrophysiology Directly Measures CFTR Gating and Conduction

The study of electrical activity in biological systems is known as electrophysiology. This field has been an important way for us to develop models for the activity of ion channels and the cells which host them [36]. Since CFTR is an ion channel, measuring its electrical activity is a direct way to assess its function and dynamics. Studies using this technique have thus allowed us to gain a molecular understanding of the disease. A comprehensive review of techniques, specifically developed to investigate CFTR can be found in [436].

In order to study the action of a single ion channel one needs to employ an experimental technique known as patch clamp electrophysiology [27]. Findings with this using this methodology were vital for the conclusions of chapters 5 and 7. For a comprehensive view of modelling from patch clamp derived data see [437]. On the other hand, to assess the competency of a whole membrane or tissue, to understand the effect of a CFTR in an organ, one instead uses what is known as an Ussing chamber [438]. This was a technique used extensively by collaborators of our MD modelling. Our collaborators grew epithelia from a patient's own cells and then placed them placed in an Ussing chamber so their competency to diffuse chloride could be measured.

Patch Clamp Electrophysiology

Patch Clamp Electrophysiology has been critical to the development of molecular biophysics in ion channels. Through the use of a micro pipette filled with electrolyte solution, we can measure the activity of a single ion channel. As shown in figure 3.9, the tip of a micro pipette is placed directly on a cell membrane in order to measure the current flowing through a single ion channel. A single ion channel exhibits a current on the order of 1 pA (for reference, CFTR exhibits 0.7 pA at a bias voltage of 100 mV), so the formation of a tight seal with the membrane is critical to filter out noise. To tighten this seal, a small amount of suction is usually applied to the membrane through the pipette [36].

In the inside-out variant of this technique, the ion channel along with a patch of surrounding membrane is excised, exposing the system to the surrounding bath. This means that the response of the ion channel to different chemical conditions can be tested by changing the composition of the bath. Patch clamp electrophysiology experiments are not presented in this thesis but they are considered closely when writing chapter 7, so an outline of them has been given here.

Ussing Chambers

An Ussing chamber is a versatile apparatus which allows the measurement of current across an epithelial membrane. In chapters 4, 5 and 6, they were used extensively. Here, a mono-layer was derived from an organoid and used as a patch in the Ussing chamber. By blocking other ion channels such as ENaC during measurements, a clear picture of CFTR function can be measured in order to create an understanding of the nature of disease in each patient. The monolayer could also be exposed to different

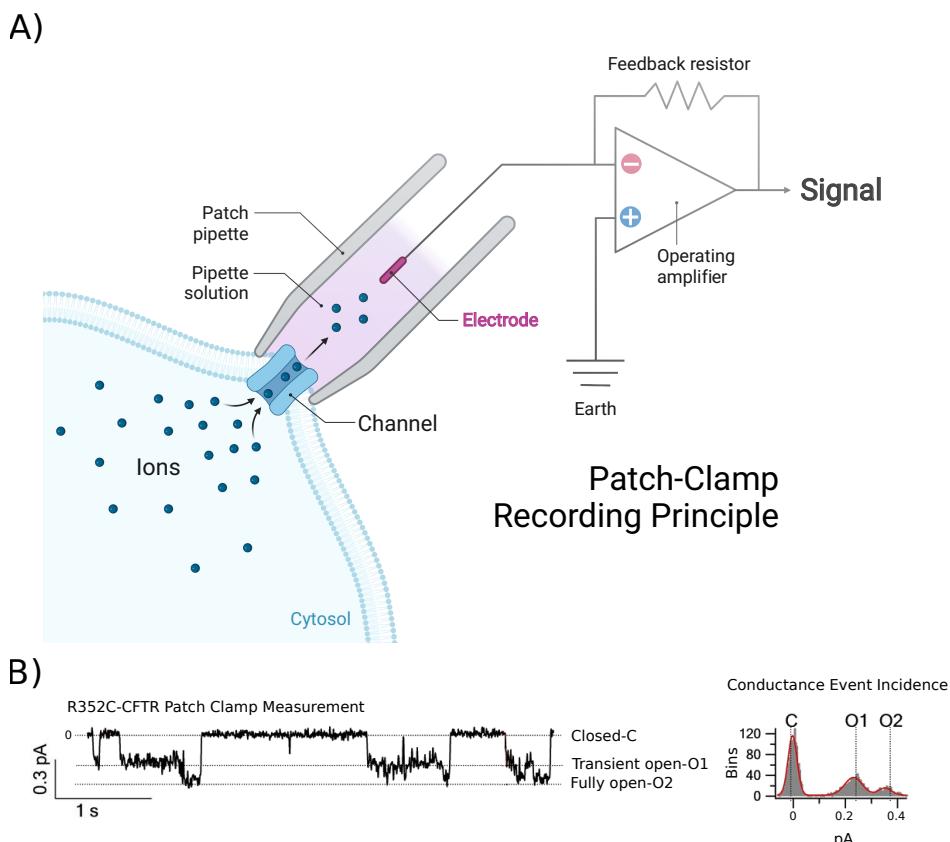


Figure 3.9: Patch Clamp Apparatus

A) In patch clamp electrophysiology, a micropipette is placed against the membrane to create a tight seal [439]. B) The current trace from the measurement of a mutant CFTR channel. The R352C mutant was chosen for visualisation as it shows a clearly shows two open states: a transient state with lower conductance, known as the O1 state and a fully open state known as the O2 state [350]. The accidental discovery of this mutant has been used extensively in studying the gating cycle of CFTR, giving an interesting example an example of how useful electrophysiology can be for molecular biophysics. Source data from [350]. Note that this mutant has roughly half the conductance of WT-CFTR.

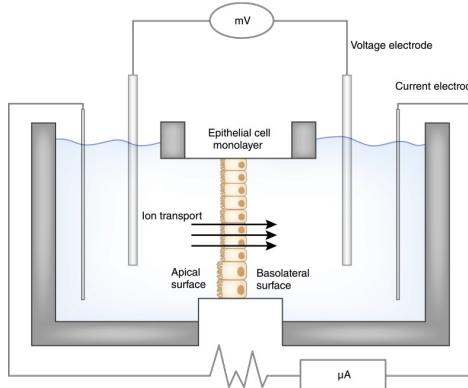


Figure 3.10: Ussing Chamber Setup

Ussing chambers can be used to characterise the function of an epithelial monolayer. The membrane can be exposed to different cellular conditions and its conductance can be measured under these conditions to test different drugs. For our purposes, an Ussing chamber may be used to measure the ion transport of an epithelium grown from a patient derived organoid. Source [438]

regimens of modulators in order to assess the ability of those modulators to rescue the function of the organoid. The higher the current, the more CFTR proteins were functioning and the better the rescue of the mutation.

In Ussing chamber measurements, current is usually expressed as Amps per unit area, which normalises the measurement to the size of the patch. It is also important that the patch of epithelial membrane is continuous, as holes would allow current to pass unrestricted, giving errantly high readings. This highlights how great care must be taken when growing organoids.

3.6.3 Western Blotting Assesses CFTR Trafficking

The above methods focus on measuring the electrical activity of the ion channels once they have been recruited to the cell membrane, they will only detect functional channels. In order to actually test for the *presence* CFTR ion channels within the epithelium, functional or otherwise, a technique called western blotting is employed (Figure 3.11).

Because of its ubiquity in biological research, a simplified explanation of the most common blotting technique is given below but there are many variations

1. The proteins in the sample of cells are denatured, sometimes by boiling. This makes them unfold and stretch out into long peptides.
2. The proteins in the sample are injected into an agar gel.
3. The gel is immersed in a buffer solution and a voltage is applied to it in a process known as electrophoresis. This causes proteins to migrate through the gel according to their charge to mass ratio. The heavier the protein, the less it will move.

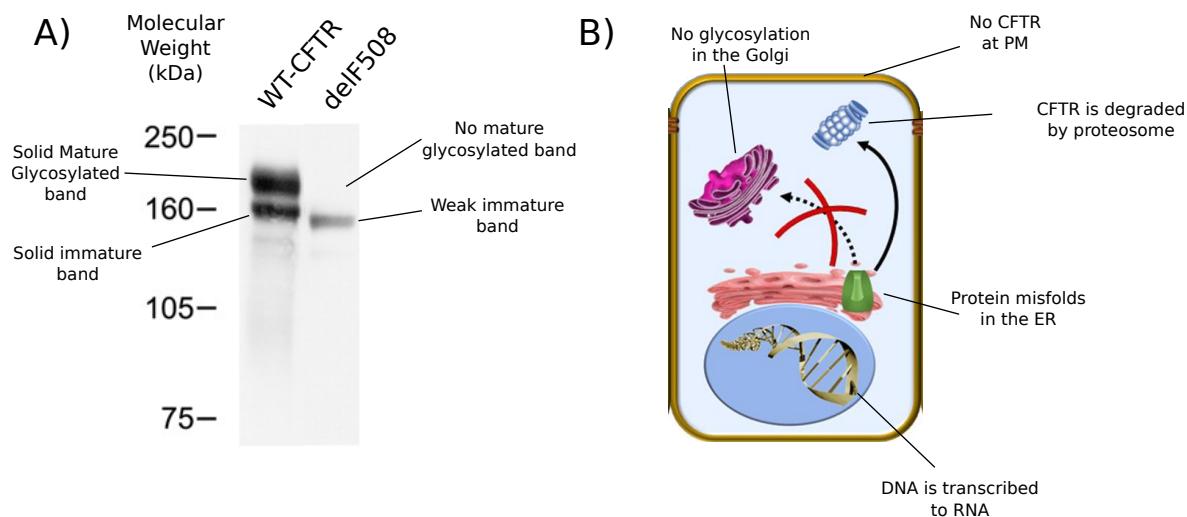


Figure 3.11: Western Blot Explanation

A) Output of a western blot experiment. WT-CFTR has much more abundant protein than $\Delta F508$ [440]. B) The cell biology behind why the $\Delta F508$ mutation is pathogenic. When WT-CFTR folds correctly in the Endoplasmic reticulum it is trafficked into the Golgi where it is glycosylated, before it is then implanted into the cell membrane. In the case of $\Delta F508$, the protein misfolds as it is synthesised. This means that instead of being trafficked as normal, $\Delta F508$ is prematurely degraded by the proteasome [441, 442].

4. The presence of proteins is visualised through immunostaining. First, the gel is washed with a primary antibody. This antibody is specific to protein we wish to detect (in our case, CFTR).
5. The excess primary antibody is washed off the gel, but some of the antibodies will stay bound to the proteins we are interested in.
6. A secondary antibody is washed over the gel. This is engineered to bind to the primary antibody. What is important about the secondary antibody is that it is attached to a special enzyme tag, which we can visualise.
7. The excess secondary antibody is washed off the gel, but some of the secondary antibodies will remain bound to the primary antibody.
8. The tag on the secondary antibody is visualised through the use of chemiluminescence. The tag is typically an enzyme, which emits light when it breaks down a substrate. Hence, for the visualisation step, a photographic film is placed over the gel and it is washed a final time, in the substrate of the enzyme which the secondary antibody is attached to. This produces bands in the photographic film, the darker the band is at position, and the more of the target is present there.

By washing the gel in different primary antibodies, one can detect multiple proteins in one experiment. This is often used to normalise the quantity of protein measured between samples. Hence, one will often see structural proteins such as tubulin or actin in figures showing the results from a western blot. These are used as controls. The level of structural proteins will not usually change between experiments, so the

intensity of these bands is used normalise measurements of the proteins of interest. Western blotting is thus a simple, widely used technique to quantify the amount of specific proteins in a sample.

3.7 Conclusion

Note, throughout the forthcoming chapters how we have relied on results from careful biochemical and biophysical experiments, which elucidated the molecular details of the function and malfunction of CFTR. Close reading of these studies was necessary in order to form hypotheses about the function of CFTR which we could test *in silico*.

In biophysics, we can only formulate our theoretical models through the collection of *in vitro* data. So, we are grateful for the painstaking work of Paul Linsdell, Christine E. Bear, Robert Ford, László Csányi, Tzyh-Chang Hwang, Naels McCarty, Isabelle Callebaut, John Paul Mornon, David Gadsby, Jue Chen, Ina Urbatsch, John Hunt, Julie D. Foreman Kay, and all the members of their laboratories, for their detailed studies of the CFTR protein. This is not to discount the work which went into the identification of the CFTR gene [443], nor the outstanding work of the clinicians and physicians who work with patients themselves [288].

CFTR is just one protein. It's a complicated one to be sure but it's far from the most complicated in a mammalian cell [444–446]. I hope that this chapter has given the physicist reader an appreciation for the amount of detail in biological systems. Remember that there are an estimated 42 million individual protein molecules in each cell, in more than 20 thousand variants [16, 447].

The decades of study into CFTR were motivated by the terrible disease it is responsible for. But this work also gives us the opportunity to demonstrate the insight which can be gained from understanding the function of proteins. In this vein, the next 4 chapters will give examples of how to combine the simulations we looked at in chapter 2 with results from patient specific assays, to understand more about CF pathogenesis. These results will be tied together in chapter 8 to help us understand *how* drugs are rescuing CFTR function, and in chapter 9 we will identify the direction this conclusion would suggest for future research into treating the root cause of Cystic Fibrosis.

4

Molecular Dynamics and Functional Characterization of I37R-CFTR Lasso Mutation Provide Insights into Channel Gating Activity

My name is Benjamin John Goodwin

– Benjamin John Goodwin (personal communication)

Preamble on Formatting

The following chapter has been published in *Cell iScience* [309]. It is typeset here to match the style of the rest of the thesis. To save space, some supplemental information has not been reprinted. It can be accessed at the source publication. Only supplemental figures and tables pertaining to *in silico* studies have been reprinted and these share the numbering of the source publication. For example Figure S8 in the source publication will appear in this chapter as Figure 4.S8. On the other hand, supplementary figures and tables pertaining to experimental results, which are not explicitly included in this chapter are denoted by a *. For example Figure S2 from the source publication is denoted in this chapter as Figure 4.S2*.

Abstract

Characterization of I37R, a mutation located in the lasso motif of the CFTR chloride channel, was conducted by therotyping several CFTR modulators from both potentiator and corrector classes. Intestinal current measurements in rectal biopsies, forskolin-induced swelling (FIS) in intestinal organoids, and short circuit current measurements in organoid-derived monolayers from an individual with I37R/F508del CFTR genotype demonstrated that the I37R-CFTR results in a residual function defect amenable to

treatment with potentiators and type III, but not type I, correctors. Molecular dynamics of I37R using an extended model of the phosphorylated, ATP-bound human CFTR identified an altered lasso motif conformation which results in an unfavorable strengthening of the interactions between the lasso motif, the regulatory (R) domain, and the transmembrane domain 2 (TMD2). Structural and functional characterization of the I37R-CFTR mutation increases understanding of CFTR channel regulation and provides a potential pathway to expand drug access to CF patients with ultra-rare genotypes.

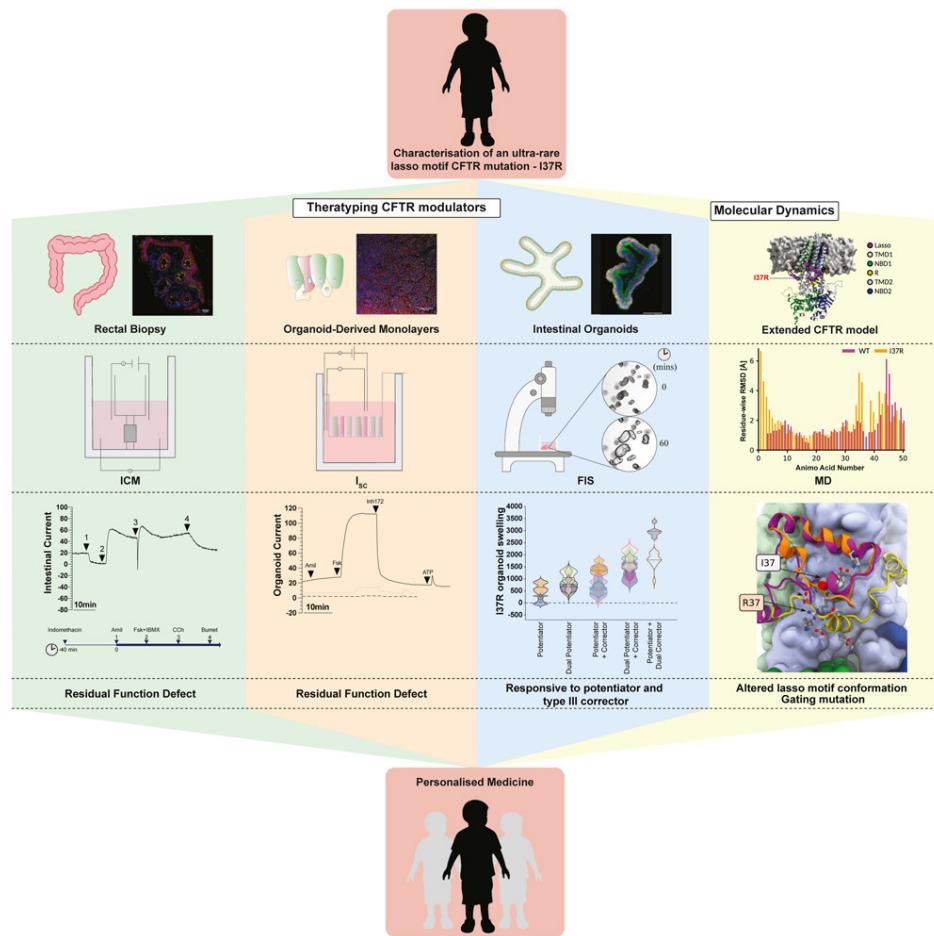


Figure 4.1: Graphical Abstract. Integration of In Silico and In Vitro Experiments for Personalised Medicine

4.1 Introduction

Cystic fibrosis (CF) is a life-limiting genetic disease resulting from mutations in the CF transmembrane conductance regulator (CFTR) gene [448]. CFTR—the only member of the ABC transporter family known to be an ion channel—consists of two transmembrane domains (TMD1 and TMD2) which form an anion-selective pore, two highly conserved nucleotide-binding domains (NBD1 and NBD2) with ATP-binding pockets and

a newly described N-terminal lasso motif [338, 349]. In addition, CFTR has a unique, disordered regulatory (R) domain which contains protein kinase A (PKA) phosphorylation sites. For the CFTR channel to open and close (gate), cAMP-dependent PKA phosphorylation of the R domain first activates the CFTR ([449]). Then, ATP-binding induces the dimerization of the two NBDs which opens the channel pore and ATP hydrolysis closes the pore.

The lasso motif (amino acids (aa) M1-L69), which is partially embedded in the bilayer and interacts with the R domain, was recently resolved following advancements in cryo-electron microscopy (cryo-EM) of the CFTR structure [45, 319]. The first 40 amino acids of the lasso motif, which include lasso helix 1 (Lh1, aa V11–R29), form a circular “noose” structure [322]. The noose structure wraps around the transmembrane helices (TM2, TM6 of TMD1 and TM10, TM11 of TMD2) and is held in place by hydrophobic interactions with L15, F16, F17, T20, L24, and Y28. The C-terminal end of the lasso, which includes the lasso helix 2 (Lh2, aa A46–L61), is tucked under the elbow helix (aa I70–R75) [322]. Variable disease severity and heterogeneous clinical presentation have been reported for the 78 CFTR variants identified so far in the lasso motif (CFTR1 and CFTR2 databases, Table 4.S1*). Evidently, the lasso motif has a multifunctional role in CFTR regulation with variants impacting folding, gating, and stability of the CFTR protein [301, 373, 374, 450, 451].

CFTR modulators, small molecules which directly target CFTR dysfunction, are now available to certain individuals with CF. Currently, two classes are approved; (1) potentiators, which open the channel pore such as ivacaftor (VX-770) and (2) correctors, which assist CFTR protein folding and delivery to the cell membrane. Type I correctors (lumacaftor/VX-809, tezacaftor/VX-661) stabilize the NBD1-TMD1 and/or NBD1-TMD2 interface by binding directly to TMD1 [452, 453] or NBD1 which improves the interaction between NBD1 and the intracellular loops [454, 455]. Type II correctors (C4) stabilize NBD2 and its interface with other CFTR domains while type III correctors (elexacaftor/VX-445) directly stabilize NBD1 [456]. Combination therapies of corrector(s) and a potentiator (Orkambi, Symdeko/Symkevi, Trikafta/Kaftrio) have been approved for CF individuals with F508del, the most common CFTR mutation, as well as several specific residual function mutations. Most recently, Trikafta/Kaftrio has been approved for patients with a single F508del mutation in combination with a minimal function mutation, broadening the population of patients with CF eligible for treatment with CFTR modulator therapy.

Mounting evidence has shown that in vitro functional studies in patient-derived cell models successfully predict clinical benefit of available CFTR modulators for individuals bearing ultra-rare mutations [457–459]. In individuals with CF, adult stem cells are usually collected by taking either airway brushings or rectal biopsies. Single Lgr5+ stem cells, derived from crypts within a patient’s intestinal epithelium, can be expanded in culture medium and differentiated into organized multicellular structures complete with the donor patient’s genetic mutation(s), thus representing the individual patient [460]. Stem cell models can be used for personalized drug screening to theratype and characterize rare CFTR mutations [457, 461, 462]. Determining the functional response of rare, uncharacterized CFTR mutations to modulator agents with known CFTR correction mechanisms enables characterization of CFTR structural defects and enhances

our understanding of CFTR function.

I37R-CFTR is a novel missense mutation in the lasso motif, detected in an Australian male child diagnosed through newborn screening with elevated immunoreactive trypsinogen, raised sweat chloride (>60 mmol/L), and CFTR Sanger sequencing identifying c.1521-1523del (F508del) and c.110C $>$ T (I37R) mutations (Table 4.S2*). We used functional studies and molecular dynamics (MD) simulations to characterize the functional and structural defects of I37R-CFTR. CFTR function was assessed using intestinal current measurements (ICM) in rectal biopsies, forskolin-induced swelling (FIS) assays in intestinal organoids, and short circuit current measurements (Isc) in I37R/F508del organoid-derived monolayers, respectively. The potentiators VX-770 (approved), GLPG1837 (phase II clinical trials), and genistein (a natural food component with potentiator activity [463]) were tested as monotherapies, dual potentiator therapies, or in combination with correctors (VX-809, VX-661, and VX-445). We compared this to our laboratory reference intestinal organoids. For MD simulations, we modeled and examined the structural defect of the I37R mutation on an extended cryo-EM structure of ATP-bound, phosphorylated human CFTR (PDB ID code 6MSM) [45].

4.2 Results

4.2.1 I37R-CFTR Baseline Activity in Patient-Derived Rectal Biopsies and Intestinal Organoids

Intestinal current measurements (ICM) were performed on I37R/F508del and reference CF (F508del/F508del, G551D/F508del) and non-CF (wild-type: WT/WT) rectal biopsies using a standard protocol [464, 465] (Figure 4.2A). Following stimulation with a forskolin (fsk) and IBMX cocktail, rectal biopsies from the I37R/F508del CF participant elicited cAMP-dependent currents of $45.8 \pm 3.8 \mu\text{A}/\text{cm}^2$ —an appreciable 50% of WT-CFTR activity ($p < 0.05$; Figure 4.2A, Table 4.S3*). This response was at least 4-fold higher than those of the reference CF biopsies, although statistical significance was not reached.

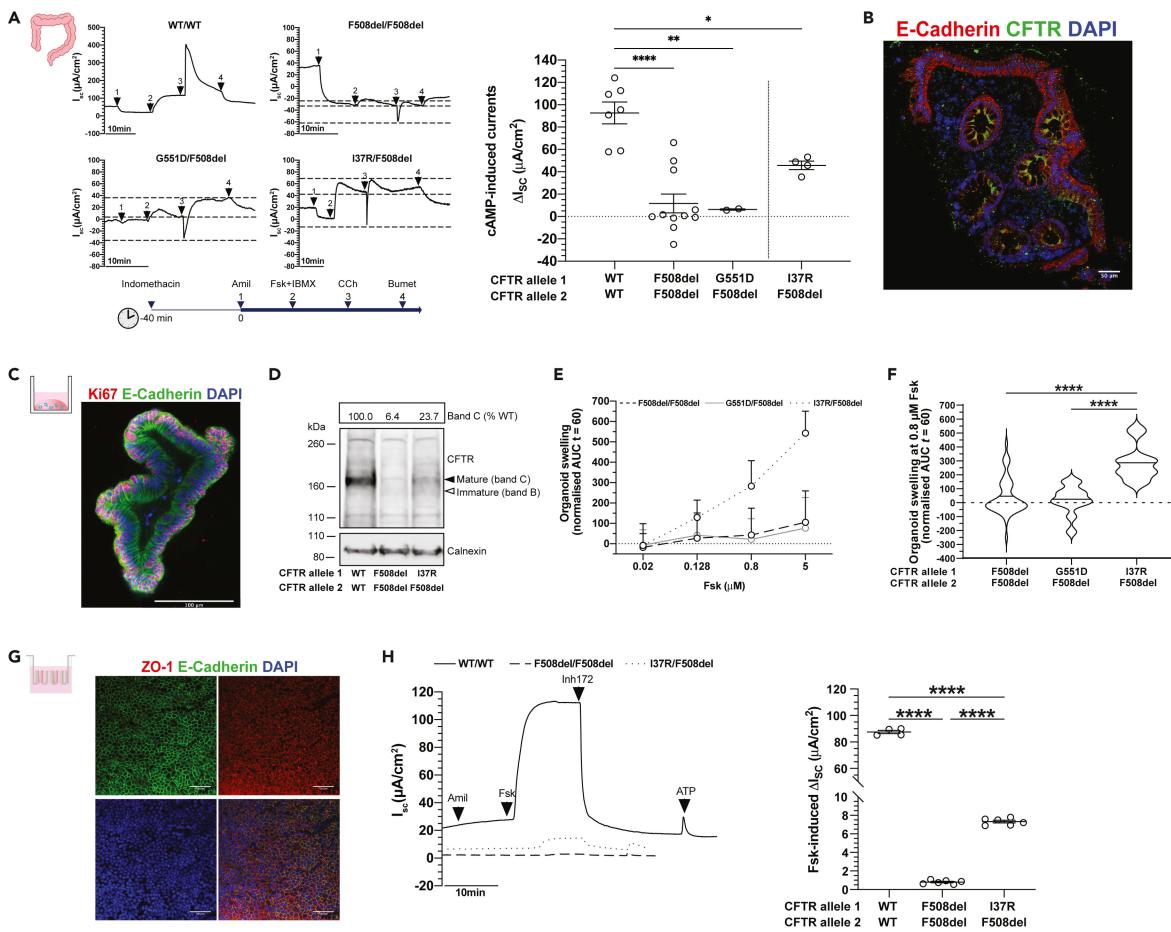


Figure 4.2: Characterization of I37R-CFTR residual function in rectal biopsies and intestinal organoids

(A) Representative Ussing chamber recordings of intestinal current measurements (ICM) in rectal biopsies from WT-CFTR control participants and participants with CF. Dot plots of cAMP-induced current (ΔI_{sc} -Fsk + IBMX) in participants with WT/WT ($n = 2$), F508del/F508del ($n = 3$), G551D/F508del ($n = 1$), and I37R/F508del ($n = 1$) CFTR genotypes. Experiments were performed in the presence of 10 μM indomethacin. Arrows indicate the addition of compounds: 100 μM apical amiloride (1. Amil), apical and basal addition of 10 μM forskolin +100 μM IBMX cocktail (2.Fsk + IBMX), 100 μM basal carbachol (3.CCh), and 100 μM basal bumetanide (4.Bumet). The I_{sc} at the time CCh was added (middle horizontal dotted line), and the maximum (top dotted lines) and minimum (bottom dotted lines) I_{sc} induced are indicated. Each dot represents an individual replicate.

(B) Immunofluorescence staining of CFTR (green), e-cadherin (red), and DAPI (blue) in a rectal biopsy derived from an I37R/F508del participant. 63x/1.4 oil immersion objective. Scale bar = 50 μm .

(C) Immunofluorescence staining of e-cadherin (green), Ki67 (red), and DAPI (blue) in intestinal organoids derived from an I37R/F508del participant. 20x/0.75 dry objective. Scale bar = 100 μm .

(D) Western blot in WT/WT, F508del/F508del, and I37R/F508del intestinal organoids. CFTR maturation was calculated by measuring the level of mature mutant CFTR (Band C) as a percentage of mature CFTR from WT organoids (% normal CFTR). All data were normalized to the calnexin loading control. B and C represents the mature, complex-glycosylated CFTR. B and B represents the immature, core-glycosylated CFTR. See Figure 4.S9* for uncropped Western blot images.

(E and F) Forskolin-induced swelling (FIS) assay in organoids from participants with F508del/F508del ($n = 5$), G551D/F508del ($n = 2$), and I37R/F508del ($n = 1$) CFTR genotypes. Organoids were stimulated with forskolin (fsk) concentrations ranging from 0.02 to 5 μM . (E) FIS expressed as the means \pm standard deviation (SD) of the area under the curve (AUC) calculated from $t = 0$ (baseline) to $t = 60$. (F) FIS of organoids at 0.8 μM fsk at baseline represent residual CFTR function. Data represented as violin plots with mean to show distribution.

(G) Immunofluorescence staining of e-cadherin (green), ZO-1 (red), and DAPI (blue) in organoid-derived monolayers from a CF participant. 20x/0.75 dry objective. Scale bars = 50 μm .

(H) Representative Ussing chamber recordings of short circuit current in organoid-derived monolayers from a WT-CFTR control participant and participants with CF. Dot plots of fsk-induced current ($\Delta\text{I}_{\text{sc}}\text{-Fsk}$) in participants with WT/WT ($n = 1$), F508del/F508del ($n = 1$), and I37R/F508del ($n = 1$) CFTR genotypes. Experiments were performed in the presence of 10 μM indomethacin. Arrows indicate the addition of compounds: 100 μM apical amiloride, 5 μM basal fsk, 30 μM apical CFTR inhibitor CFTRinh-172, and 100 μM apical ATP. Each dot represents an individual replicate. Data in (A) and (H) represented as mean \pm standard error of the mean (SEM). One-way analysis of variance (ANOVA) was used to determine statistical differences. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$.

Co-activation with carbachol (CCh) resulted in a biphasic response in the I37R/F508del biopsies, characteristic of residual CFTR chloride channel function in the CF colon [465, 466]. The initial negative Isc peak indicates apical potassium secretion reached $9.4 \pm 2.5 \mu\text{A}/\text{cm}^2$. Following this, the CCh-induced positive Isc indicates the increase of apical chloride secretion reached $15.78 \pm 2.07 \mu\text{A}/\text{cm}^2$. This biphasic response was similarly observed in the G551D/F508del biopsies ($25.77 \pm 2.16 \mu\text{A}/\text{cm}^2$) but was diminished in the F508del/F508del biopsies ($-2.28 \pm 1.65 \mu\text{A}/\text{cm}^2$). These findings are in accordance with the localization of CFTR protein at the plasma membrane (mature complex-glycosylated CFTR) of the I37R/F508del rectal biopsies, as demonstrated by immunofluorescence staining (green; Figure 4.2B).

Next, CFTR protein expression and maturation was assessed in I37R/F508del, reference F508del/F508del, and WT/WT organoids using Western blot (Figures 4.2C-D). The expression of complex-glycosylated C band in I37R/F508del organoids was 23.7% that of the WT/WT organoids, considerably higher than the 6.4% detected from F508del/F508del organoids (Figure 4.2D). CFTR activity was then evaluated

in I37R/F508del and CF reference intestinal organoids using a fsk-induced swelling (FIS) assay at four fsk concentrations between 0.02 and 5 μM (Figure 4.2E). FIS of I37R/F508del intestinal organoids at 0.8 μM fsk—the optimal concentration for baseline assessment of CFTR activity [467]—was 282.9 ± 36.0 (Figures 4.2E-F). This exceeded the baseline FIS of the reference intestinal organoids by at least 7-fold (F508del/F508del: AUC = 42.8 ± 19.4 ; G551D/F508del: AUC = 21.3 ± 29.4).

The morphological difference between WT (pre-swollen) and CF organoids [468], means comparing CFTR activity between CF and healthy CFTR function by FIS assay cannot be achieved [467, 469]. In order to compare I37R/F508del to wild-type CFTR activity, organoid-derived monolayers were created (Figure 4.2G) and CFTR ion transport was performed [470]. Fsk-stimulated CFTR-dependent currents were 9-fold higher in I37R/F508del monolayers than those of reference F508del/F508del monolayers (7.3 ± 0.2 vs $0.8 \pm 0.1 \mu\text{A}/\text{cm}^2$; $p < 0.0001$), but 12-fold lower than WT/WT monolayers ($87.5 \pm 1.3 \mu\text{A}/\text{cm}^2$; $p < 0.0001$) (Figure 4.2H). This is consistent with the FIS assay results demonstrating high baseline CFTR activity in I37R/F508del intestinal organoids.

4.2.2 I37R-CFTR Functional Response to CFTR Modulator Monotherapy in Intestinal Organoids

We investigated the functional response of I37R/F508del organoids to single potentiators: VX-770, GLPG1837 (G1837), and genistein (Gen). Treatment with VX-770 minimally increased FIS of I37R/F508del organoids by AUC of 59.7 above baseline at 0.128 μM fsk (Figures 4.3A–C and 4.S1*)—the optimal concentration for in vitro assessment of CFTR modulator response to predict clinical effect [467]. G1837 and Gen both significantly increased FIS, albeit with different efficacies (655.8 and 256.8, respectively; Figures 4.3A–C and 4.S1*). None of the potentiator treatments increased FIS in F508del/F508del organoids, indicating no improvement in CFTR activity in response to potentiator therapy (Figure 4.3C). Only G1837 significantly increased FIS in the G551D/F508del organoids (210.4 ± 57.5 ; $p < 0.01$). In comparison to G551D/F508del organoids, G1837 was 3-fold more efficacious in the I37R/F508del organoids ($p < 0.0001$).

testing

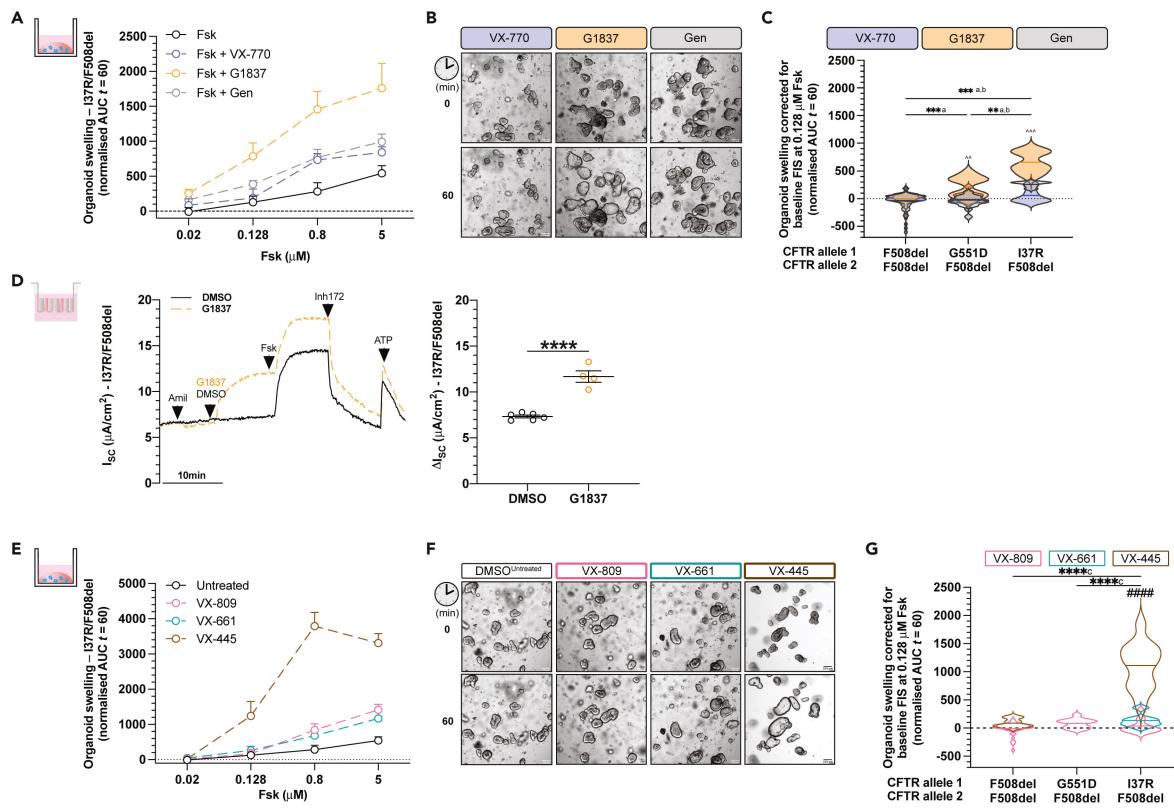


Figure 4.3: Characterization of I37R-CFTR functional response to corrector or potentiator monotherapy in intestinal organoids

Forskolin-induced swelling (FIS) assay in organoids from participants with F508del/F508del ($n = 5$), G551D/F508del ($n = 2$), and I37R/F508del ($n = 1$) CFTR genotypes. Organoids were incubated overnight with 0.03% DMSO (untreated) or 3 μ M VX-809 or 3 μ M VX-661 or 3 μ M VX-445. After 24 h, organoids were stimulated with fsk concentrations ranging from 0.02 to 5 μ M, either alone or in combination with potentiator monotherapy (3 μ M VX-770 or 3 μ M G1837 or 50 μ M Gen).

(A) FIS of I37R/F508del organoids stimulated with VX-770, GLPG1837 (G1837), or genistein (Gen) monotherapy, expressed as the means \pm standard deviation (SD) of the area under the curve (AUC) calculated from $t = 0$ (baseline) to $t = 60$ min.

(B) Representative brightfield images of I37R/F508del organoids at baseline ($t = 0$) and after 1 h of stimulation ($t = 60$) at 0.128 μ M fsk. Scale bars = 100 μ m.

(C) FIS of organoids at 0.128 μ M fsk following stimulation with VX-770, GLPG1837 (G1837), or genistein (Gen) monotherapy. Data corrected for baseline FIS and represented as violin plots with mean to show distribution.

(D) Representative Ussing chamber recordings of short circuit current in I37R/F508del organoid-derived monolayers. Dot plots of total currents stimulated by DMSO or G1837 plus fsk. Experiments were performed in the presence of 10 M indomethacin. Arrows indicate the addition of compounds: 100 M apical amiloride, apical addition

of either vehicle control 0.01% DMSO or 10 μ M G1837, 5 μ M basal fsk, 30 μ M apical CFTR inhibitor CFTRinh-172, and 100 μ M apical ATP. Each dot represents an individual replicate. Data represented as mean \pm standard error of the mean (SEM).

(E) FIS of I37R/F508del organoids pre-incubated with corrector (VX-809 or VX-661 or VX-445) for 24 h, expressed as the means \pm standard deviation (SD) of the area under the curve (AUC) calculated from $t = 0$ (baseline) to $t = 60$ min.

(F) Representative brightfield images of I37R/F508del organoids at baseline ($t = 0$) and after 1 h of stimulation ($t = 60$) at 0.128 μ M fsk. Scale bars = 100 μ m.

(G) FIS of organoids at 0.128 μ M fsk following incubation with corrector (VX-809 or VX-661 or VX-445) for 24 h. Data corrected for baseline FIS and represented as violin plots with mean to show distribution. One-way analysis of variance (ANOVA) was used to determine statistical differences except in (D) where unpaired t test was used. **p < 0.01, ***p < 0.001, and ****p < 0.0001. aP for G1837, bP for Gen and cP for VX-445 of I37R/F508del, \wedge P for G1837 vs VX-770, or Gen and #P for VX-445 vs VX-809 or VX-661.

Because G1837 demonstrated the greatest restoration of CFTR activity in I37R/F508del organoids, we evaluated G1837 treatment of I37R/F508del organoid-derived monolayers. G1837 led to a significant 1.5-fold increase in fsk-stimulated currents (Δ Isc: 4.4 μ A/cm²; p < 0.0001) (Figure 4.3D). This is consistent with the FIS of I37R/F508del organoids, indicating that I37R-CFTR responds to potentiator agents.

Given the I37R/F508del high residual CFTR activity and its localization at the epithelial cell surface, we hypothesized that the I37R-CFTR mutation has minimal impact on CFTR protein folding or maturation. Treatment of I37R/F508del organoids with type I corrector agents (VX-809 or VX-661) did not significantly increase FIS above baseline (Figures 4.3E–G and 4.S1*). In contrast, treatment of I37R/F508del organoids with a type III corrector agent (VX-445) significantly increased FIS by AUC of 1112.5 above baseline, greater than those in the F508del/F508del organoids (42.5). VX-445 has been shown to act as both a corrector and potentiator for certain CFTR mutations [471–473]. Acute treatment of I37R/F508del organoids with VX-445 did not improve potentiation of CFTR (Figure 4.S1*). This supports the observation that VX-445-stimulated rescue of CFTR in I37R/F508del organoids acts by a correction mechanism improving I37R mild folding and processing defects. I37R-CFTR functional response to CFTR modulator co-therapies in intestinal organoids.

Combination treatments of CFTR modulators are used to treat patients bearing CFTR mutations with multiple functional defects such as F508del and patients who are heterozygous for CFTR mutations. We investigated the effect of combinations of potentiators. Dual potentiator combinations increased FIS of I37R/F508del organoids to a greater extent than the respective single potentiators (Figure 4.4A) and had a synergistic effect, where the FIS was greater than the sum of the respective single potentiators (Table 4.S4*). Despite G1837 + Gen having greater efficacy than the other dual potentiator combinations, the magnitude of response was not statistically different between

the different combinations of dual potentiators (Figure 4.4A).

Co-therapy with a corrector (VX-809 or VX-661) and dual potentiators significantly ($p < 0.01$) increased FIS of I37R/F508del organoids compared to co-therapy of a corrector with VX-770 or Gen, but not G1837 (Figure 4.4B). VX-809/G1837 + Gen co-therapy had the greatest efficacy, increasing FIS 1904.0 above baseline. In contrast, corrector/VX-770 + Gen co-therapy had the least efficacy. This trend was consistent with that of the dual potentiators synergistic effect.

Dual correctors (VX-445+VX-661) increased FIS in I37R/F508del organoids by AUC of 1856.6 above baseline, which corresponds with the level of rescue achieved by the most effective corrector/dual potentiator co-therapy (VX-809/G1837 + Gen). The triple combination therapy with dual correctors and a potentiator further increased FIS in I37R/F508del organoids by AUC of 3101.6 above baseline. It is therefore the most effective modulator combination tested in this study.

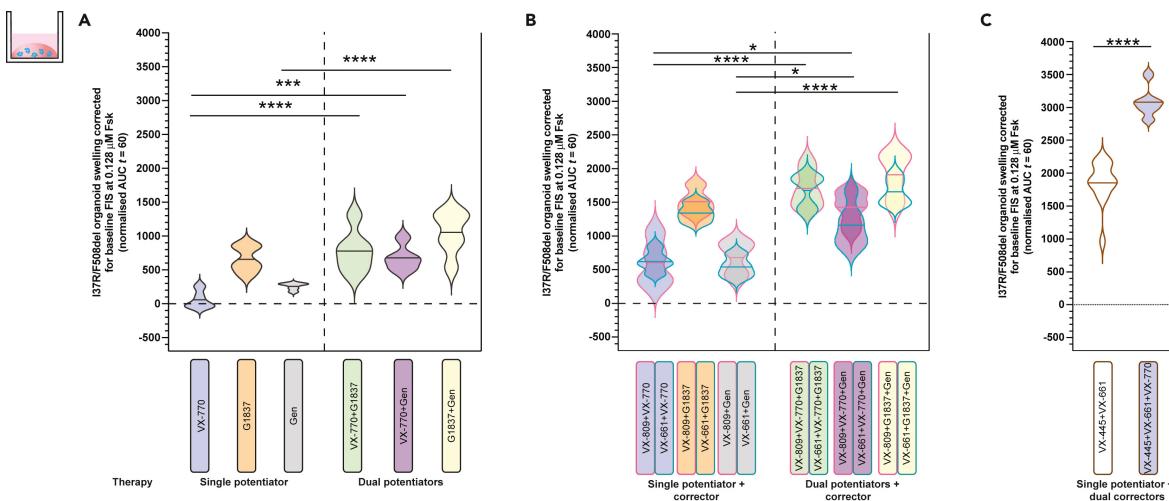


Figure 4.4: Characterization of I37R-CFTR functional response to dual potentiator or corrector therapy or corrector(s)-potentiator(s) co-therapy in intestinal organoids

Forskolin-induced swelling (FIS) assay in organoids from participants with F508del/F508del ($n = 5$), G551D/F508del ($n = 2$), and I37R/F508del ($n = 1$) CFTR genotypes.

(A and B) Organoids were incubated overnight with 0.03% DMSO (untreated) or 3 μ M VX-809 or 3 μ M VX-661 or 3 μ M VX-445 + 18 μ M VX-661. After 24 h, organoids were stimulated with fsk ranging in concentration from 0.02 to 5 μ M, either alone or in combination with a single potentiator (3 μ M VX-770 or 3 μ M G1837 or 50 μ M Gen) or dual potentiators (VX-770 + G1837 or VX-770 + Gen or G1837 + Gen). FIS of organoids at 0.128 μ M fsk stimulated with VX-770, GLPG1837 (G1837), or genistein (Gen) or their combinations, following (A) 24 h pre-incubation with DMSO (untreated) or (B) corrector (VX-809 or VX-661), respectively.

(C) FIS of organoids at 0.128 μ M fsk stimulated without or with VX-770, following 24 h pre-incubation with dual correctors (VX-445+VX-661). Data corrected for baseline

FIS and represented as violin plots with mean to show distribution. One-way analysis of variance (ANOVA) was used to determine statistical differences except in (C) where unpaired t test was used. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$

4.2.3 I37R-CFTR Perturbs the Noose Structure of the Lasso Motif

We next characterized the structural defect of I37R-CFTR using MD simulations. The primary structure of the lasso motif (M1-L69) is conserved across 230 vertebrate species (Figure 4.S2, Table 4.S5*). The lasso motif formed a noose structure that rested against TMD2 (Figure 4.5A). Amino acids V12-R29 were embedded in the plasma membrane while the rest of the lasso motif resided in the cytosol. The noose structure was maintained by a salt bridge formed between K26 and D36 (Figure 4.5B). I37 was positioned in the center of this noose, within a hydrophobic pocket formed by amino acids from the lasso, TMD2, and the poorly resolved R domain in the cytosol (Figure 4.5C).

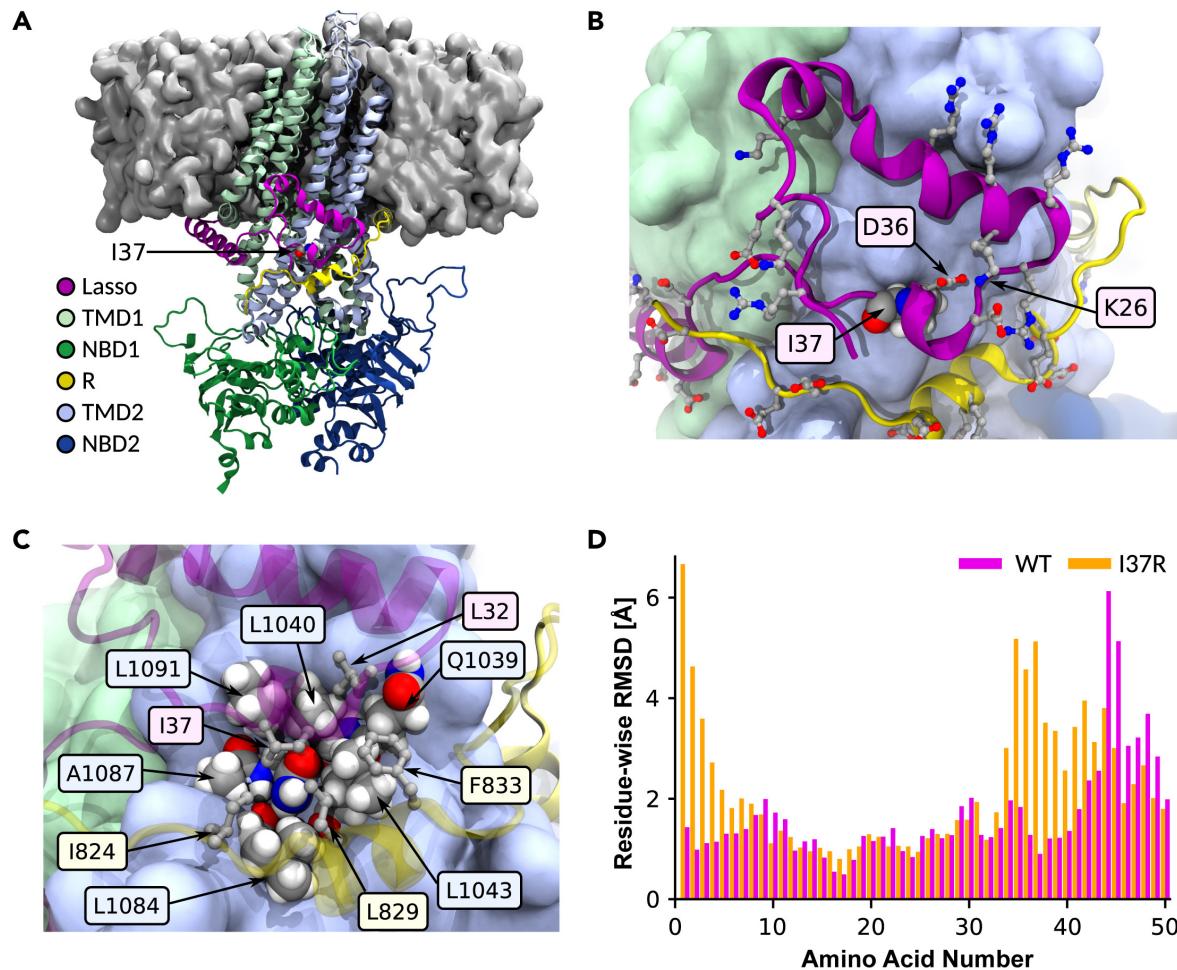


Figure 4.5: Placement of I37R within the lasso motif and the resulting changes to its conformation

(A) Ribbon structure of human CFTR, partially embedded within the plasma membrane (gray surface). I37 rendered as spheres. TMD: transmembrane domain;

NBD: nucleotide-binding domain; R: regulatory domain (R domain).

(B) K26-D36 salt bridge stabilizing the noose structure of the lasso motif. Charged amino acids depicted as balls and sticks. I37 rendered as spheres, and color-coded by element (gray: C; white: H; red: O; blue: N).

(C) I37 positioned within a hydrophobic pocket formed by amino acids from the lasso motif, TMD2, and poorly resolved R domain. Relevant amino acids labeled and depicted as spheres in TMD2, and as balls and sticks in lasso motif and R domain.

(D) Residue-wise root-mean-square deviation (RMSD) to the C-alpha atoms of the WT-CFTR 6MSM model, measuring the conformational change of the WT (pink) and I37R mutant (orange) lasso after 2 μ s simulations. Values are means sampled over the last μ s of simulations.

Mutation of the evolutionarily conserved, non-polar and uncharged isoleucine (I) of I37 to a positively charged arginine (R) introduced an unstable lone charge into the hydrophobic pocket within the lasso motif noose. We hypothesized that this likely results in the rotation of the R37 side chain out of the hydrophobic pocket, and possible coordination with negative charges in the nearby R domain.

To identify a reasonable conformation of the mutant lasso motif, the WT 6MSM model was mutated to R37 and three 2 μ s simulations were performed at physiological temperature (310 K). The R37 side chain rotated out of the hydrophobic pocket in only one of the three simulations. The difference between the root-mean-square deviation (RMSD) of the noose structure of I37R-CFTR compared to the WT was on average 2.8 Å at the amino acids M1-L6, and 1.8 Å at L34-S50 (Figure 4.5D). To confirm this observation, repeat simulations were performed at 350 K (40°C above physiological temperature), a temperature shown to accelerate the potential conformational transitions of proteins [474]. In these higher temperature simulations, the root-mean-square fluctuation (RMSF) of the region around amino acid 37 doubled in two out of three simulations, compared to WT-CFTR at 310K (Figure 4.S3). This confirmed the destabilization of the lasso motif by I37R-CFTR. All WT-CFTR domains and the surrounding bilayer remained stable at the elevated temperature (Figures 4.S4 and 4.S5).

4.2.4 I37R Mutation Strengthens Lasso Motif Interaction with the R-domain

In the 6MSM structure, the R domain is largely unresolved with two exceptions: the first (Q637) and last (T845) amino acids that adjoin neighboring domains, and the backbone atoms of a 17 amino acid segment. This latter segment consists of an eight amino acid disordered coil followed by a nine amino acid alpha-helix [45]. The alpha-helix was separated by approximately 10 Å (1 nm) minimum C-alpha distance to I37 in the lasso motif. This suggested a likely interaction between this segment of the R domain and I37, which necessitated partial modeling of the R domain (Figure 4.6A).

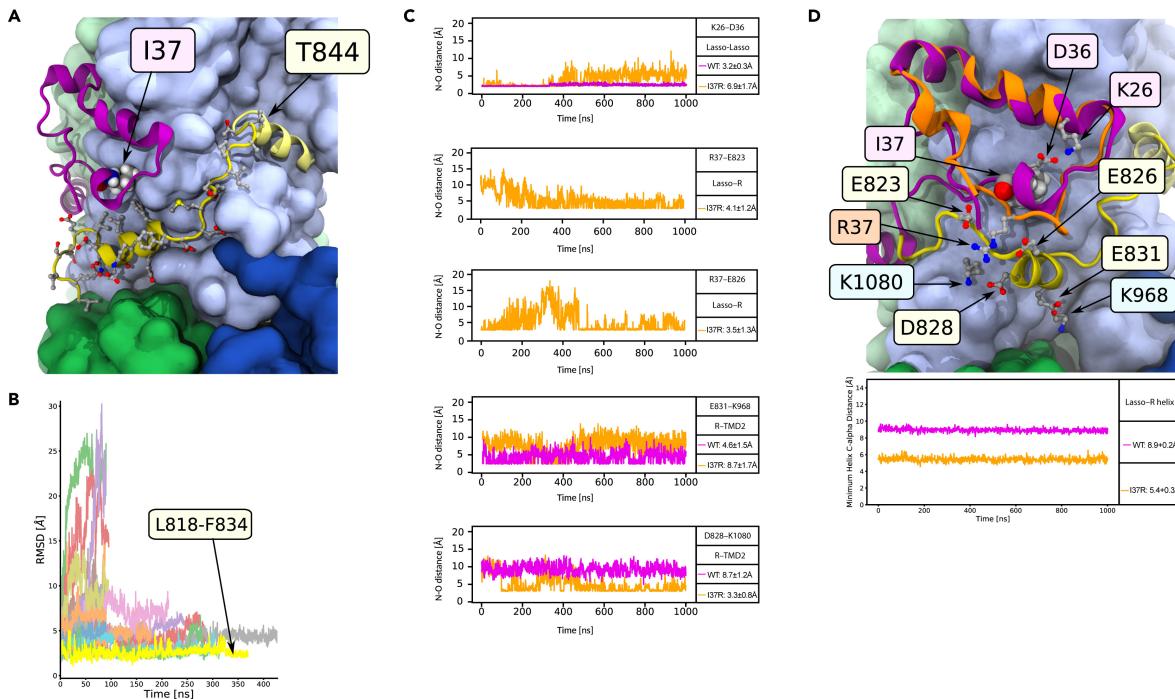


Figure 4.6: I37R interacts with a previously unresolved section of the R domain
(A) The reconstructed R domain amino acids (yellow), depicting the assignment of L818-F834 to the 17 amino acids with only the backbone resolved in the 6MSM structure, and the linking residues to T845 in TMD2. Lasso motif in purple. Side chains depicted as balls and sticks.

(B) The stabilities of all 24 modeled R domain assignments, quantified by RMSD to the 6MSM structure. The most stable alignment of the 17 unidentified amino acids, L818-F834, is highlighted yellow. The full list of tested assignments is shown in Supplementary material 9*.

(C) The minimum N-O distance between newly formed and disrupted salt bridges in the I37R mutant. Distance less than 4 Å indicates direct contact. Values are means \pm standard deviations (SD), sampled over the last 500 ns of simulations.

(D) Conformational changes in the I37R mutant (orange) compared to WT (purple) lasso motif, which brings it closer to the R domain (yellow). Minimum C-alpha atom distance between amino acid 37 and the R domain helix (E826-F834) in I37R and WT.

Modeling of these 17 unidentified amino acids was performed by creating 24 different in silico models of this segment based on the 6MSM structure. In each model, a unique 17 amino acid sequence was determined with a sliding window of one amino acid, starting backwards from amino acid T842 due to the alpha-helix's 20 Å proximity to T845. The 17 amino acids were then connected to T845 with the missing linking amino acids. The structural stability of all 24 modeled segments was tested by performing up to 300 ns simulations for each model and comparing the backbone RMSD measurements against 6MSM (Figures 4.6B and 4.S6). The model with the lowest RMSD (3 Å) and thus

the highest stability was attained when L818-F834 was assigned to the unidentified 17 amino acids, of which the alpha-helix maps to E826-F834 (Figures 4.6B and 4.S6). This assignment was corroborated by NMR measurements of the isolated R domain in solution, where the same segment retained partial helicity [308]. Predictions of the structure of human CFTR by AlphaFold2 also aligned with this assignment of primary structure to the unidentified amino acids (Figure 4.S7) [69]. Several favorable interactions between this R domain model and other parts of the CFTR protein further supported this assignment (Figures 4.5C and 4.6D). Two hydrophobic amino acids (L829 and F833) contributed to the hydrophobic pocket that stabilized the lasso motif around I37. The negatively charged E831 formed a salt bridge with positively charged K968 in TMD2. Together, these interactions secured the R domain alpha-helix into position throughout an extended 2 μ s simulation, resulting in a smaller minimum C-alpha distance to the lasso motif of 8.9 ± 0.2 Å compared to the 10 Å in the 6MSM cryo-EM structure.

The reoriented R in position 37 in the I37R mutant protein, which pointed out of the hydrophobic pocket, rearranged the salt bridge network supporting the lasso motif by breaking the evolutionarily conserved salt bridge K26-D36. Two new salt bridges were formed, one with the negatively charged E823 and another with E826 of the R domain (Figure 4.6C). Furthermore, the E831-K968 salt bridge between the R and TMD2 domains in the WT was exchanged for a D828-K1080 salt bridge in I37R-CFTR (Figure 4.6C). The backbone motions required to accommodate these new charge interactions also perturbed parts of the lasso motif (Figure 4.S3) and R domain. The lasso N-terminus shifted its position towards the R domain and reduced the minimum C-alpha distance between them by 3.5 Å (Figure 4.6D). The overall result was a tighter coupling between the lasso and the R domain which is anticipated to inhibit the R domain movements required for channel gating.

4.3 Discussion

We have described the functional and structural defects of I37R, a novel CF-causing mutation in the segment of the CFTR lasso motif which interacts with the R domain. These were compared to reference CFTR mutations which have known functional defects, either a CFTR folding/maturation (F508del/F508del) or a gating (G551D/F508del) defect. First, ICM performed in I37R/F508del rectal biopsies identified I37R confers high residual activity (50% of WT-CFTR activity). High baseline CFTR activity was similarly observed in FIS of I37R/F508del intestinal organoids and Isc measurements in organoid-derived monolayers. Given we and others showed that F508del is a severe mutation which contributes little functional CFTR [475], this suggests that I37R mutation produces CFTR protein which localizes to the epithelial cell surface. These observations are consistent with the patient's mild CF clinical phenotypes (pancreatic sufficient with faecal elastase > 500 μ g/g, FEV1 z-score -0.11, 99% predicted).

We also characterized the response of I37R-CFTR to modulators (potentiators and correctors) in I37R/F508del intestinal organoids and organoid-derived monolayers. I37R was responsive to potentiaters which improve CFTR gating function and a newly approved corrector (VX-445). Among the three potentiator agents tested, the response to

VX-770 was minimal. The reason for the lack of efficacy of VX-770 is not known, because molecular modeling studies propose that VX-770 shares the same mechanism of action and binding sites with G1837 [311, 375]. Both VX-770 and G1837 are proposed to potentiate CFTR by increasing channel open probability (P_o) through stabilization of the open-pore conformation, independent of NBD dimerization and ATP hydrolysis which normally controls channel gating [378, 416]. However, the differing potentiator efficacies are not a new observation. G1837 was previously shown to be more potent and effective than VX-770 in human bronchial epithelial cells from a G551D/F508del and a R334W/F508del CF participant [400, 476]. Similar observations were reported in heterologous HEK293 cells expressing Class III (G551D, G178R, and S549N) and Class IV (R117H) CFTR mutants [400, 476]. We conclude that perhaps G1837 has additional binding sites or actions distinct from VX-770, which in the case of I37R-CFTR, results in significant potentiation of the CFTR channel.

We further showed that dual potentiator combinations exerted synergistic restoration of CFTR activity in I37R/F508del organoids. This synergistic restoration is not exclusive to I37R-CFTR, because similar findings have been reported for other CFTR mutations responsive to potentiators [477–480]. Synergism is commonly achieved when potentiators have distinct binding sites and mechanisms of actions. One potentiator could induce allosteric interactions that favor the activity of the other potentiator [481]. The potentiator synergy observed in our dual potentiator combinations supports our hypothesis that G1837 may have additional binding sites or mechanisms of action to VX-770. While VX-770 has been shown to provide clinical benefit to patients with responsive mutations [482–484], it does not restore the P_o of gating defect mutants (G551D-CFTR) to full WT-CFTR activity [416]. This opens the possibility that using another potentiator with a different mechanism of action could complement VX-770 activity and increase CFTR activity beyond that of VX-770 monotherapy. While VX-770 and G1837 act independently of NBD dimerization and ATP hydrolysis [378, 416], genistein promotes ATP-dependent gating of CFTR by binding to the NBD1/2 interface and inhibiting ATP hydrolysis [485]. Genistein has been demonstrated to increase VX-770-potentiated CFTR activity in intestinal organoids, even when VX-770 was used at near-saturating concentrations [477]. Our observations reiterate and expand on these findings to suggest that potentiators with different mechanisms of action could provide synergistic restoration of CFTR activity to responsive CFTR mutations compared to potentiator monotherapy.

Chronic treatment with type III corrector VX-445 rescued CFTR activity in I37R/F508del organoids, while neither type I correctors (VX-809 or VX-661) rescued activity. This response is attributed to the I37R and not the F508del mutation in the I37R/F508del organoids, because VX-445 did not restore CFTR activity in F508del/F508del organoids. While VX-445 has been shown to have partial potentiator activity [471–473], VX-445 did not potentiate CFTR activity in I37R/F508del organoids when administered acutely. This is the first study to interrogate the potentiator action of VX-445 in intestinal organoids; however, previous studies have been performed in donor-derived bronchial and nasal epithelial cells and immortalized cell lines. The higher correction efficacy of VX-445 when compared with VX-809/VX-661 has previously been shown, although this is likely to be dependent on the CFTR variant [486–488]. For instance,

direct binding of VX-445 to NBD1 to stabilize and prevent the domain unfolding may make it more effective in correcting CFTR mutations that impact NBD1 function (such as F508del located in NBD1).

The lack of I37R-CFTR correction by VX-809 or VX-661 could be attributed to the dependency of these modulators binding to and stabilizing the TMD1. TMD1 function is modulated by interaction with lasso helix 2 (Lh2, aa A46–L61) as deletion of Lh2 from the WT CFTR was shown to completely abrogate VX-809-mediated CFTR maturation [373]. MD studies showed that VX-809 occupancy at the TMD1 binding site causes the Lh2 to move, such that the network of salt bridges in Lh2 holds TMD1 (CL1) and TMD2 (CL4) in the correct orientation [382, 456]. This then allows for allosteric coupling between NBD1 and TMD1 or 2, which is important for cooperative domain folding of CFTR. In support of this, mutation of critical amino acids at the binding pocket of VX-809 on CFTR, or those involved in the architecture of this site, were shown to diminish the sensitivity to VX-809 correction. L53V and F87L mutations, which are located in the vicinity of the VX-809 binding site in the TMD1, were shown to prevent VX-809 correction in F508del HEK283 cells [382]. Considering the above and because I37 is only a few amino acids away from the Lh2, it is plausible that the local conformational changes associated with the I37R mutation which we have identified in our study (Figure 4.5D) may disrupt the allosteric coupling between NBD1 and TMD1 or 2, preventing correction with type I correctors.

CFTR missense mutations in the lasso motif are not well characterized. This is because most of these mutations are rare, with an allele frequency of less than 0.01% in the CF population (Table 4.S1*). The only characterized missense mutations in the region of the lasso motif where I37 resides—between Lh1 (amino acid 19–29) and Lh2 (amino acid 46–61)—are R31C and R31L [86, 451]. Experimental studies in heterologous COS-7 cells showed both mutations cause a mild processing defect and accelerated CFTR internalization. Individuals heterozygous for these CFTR mutations are reported to have a mild disease phenotype with pancreatic sufficiency [451]. One individual with the R31C/F508del CFTR genotype was reported to have a normal sweat chloride level (25 mmol/L) and nasal potential difference [489]. CFTR2 classifies R31C as a non-CF disease causing mutation. Notably, mild disease phenotypes (mild pulmonary symptoms, pancreatic sufficiency) are reported for several other lasso motif missense mutations including P5L, E56K, and P67L (Table 4.S1*), as was found for the I37R/F508del participant in this study. This suggests that perhaps lasso motif mutations do not significantly impact the overall CFTR structure and function given its short length (69 of 1480 amino acids, 4.7%). It is also plausible that the role of the lasso motif could be compensated for by other CFTR domains.

To better understand the functional defect of I37R-CFTR, we used MD simulations to model the structural features of I37R and how they are altered relative to WT-CFTR. The amino acids 34–39 were shown to interact with the R domain in the phosphorylated, ATP-bound CFTR structure [45]. This interaction was absent in the closed conformation of CFTR [338], suggesting that the short region of amino acids 34–39 interacts with the R domain to regulate CFTR channel gating. We found that the disruption of the evolutionarily conserved K26-D36 salt bridge in I37R-CFTR brings the lasso motif closer to the R domain. We also found that the I37R side chain

rotates out of its hydrophobic pocket to form interactions with negatively charged E823 and E826 on the R domain. We speculate that R37 clamps the lasso motif to the R domain, preventing the dynamic movement of the two domains necessary for a normal CFTR opening and closing cycle, thus causing a gating defect. This supports our functional observations, wherein I37R-CFTR demonstrated significant responsiveness to potentiator agents which are known to increase channel opening time. Furthermore, in the I37R-CFTR model, conformational changes in the lasso motif were also evident but were limited to short regions (M1-L6, L34-S50), indicating that the overall architecture of the CFTR protein remains largely intact. Additionally, our simulations did not show any change to the pore architecture of CFTR (Figure 4.S8).

The simulated structure in this work is of CFTR in its active state [45]. Because of this, we believe the pathogenic interactions discovered in this study have a significant contribution to the deleterious effects of the I37R mutation. However, the enhanced lasso motif-R domain interactions should be interpreted in the context of the μ s timescales reachable by unbiased simulations. The lasso domain is known to exhibit conformational flexibility during both folding and functional stages of CFTR [490], which take place on timescales longer than is currently feasible to study in atomistic simulations. Therefore, there may be pathogenic interactions in I37R-CFTR in addition to the ones captured by the simulation of this particular CFTR structure.

The I37R/F508del participant in this study will only meet the Therapeutic Goods Administration (Australia) requirements for treatment with Trikafta/Kaftrio triple combination therapy once he turns 12 years old given the single copy of the F508del mutation. He is not eligible for single potentiator therapy or corrector/potentiator combinations of lumacaftor/ivacaftor or tezacaftor/ivacaftor. This emphasizes the importance of characterizing the structural and functional defects of ultra-rare CFTR mutations together with the assessment of in vitro response to modulator drugs in patient-derived cell models to build the case for access to treatment with available modulators through precision medicine health technology assessment pathways. Furthermore, when multiple CFTR modulators are available to patients with CF, determining the best modulator for patients with a rare mutation not investigated in a clinical trial may be supported using in vitro personalized cell models.

Limitations of the study

Organoids often lack specialized cell types and fail to recapitulate the complexity of native organs [428]. For example, mesenchymal, endothelial, and microbiome are absent from intestinal organoids. Integration of such features remains technically challenging and their absence may impact drug response. Another important drawback of organoid systems is the heterogeneity in their size when seeded for FIS assay. As the size of organoids increases, diffusion-dependent drug supply becomes less efficient. This may in turn impact the accuracy of outcome of drug assay. Reducing this variability will be essential to fully capitalize on the potential of organoids in drug screening. Another limitation of the organoid systems is the variability in the magnitude of FIS response in intestinal organoids across different CF laboratories. This is due to the dependence of organoids on media that is developed in-house with many locally produced media

factors [459, 467]. This limitation can be resolved by the creation of reference donor organoids which are made available and used internationally between CF laboratories.

4.4 Supplementary Information

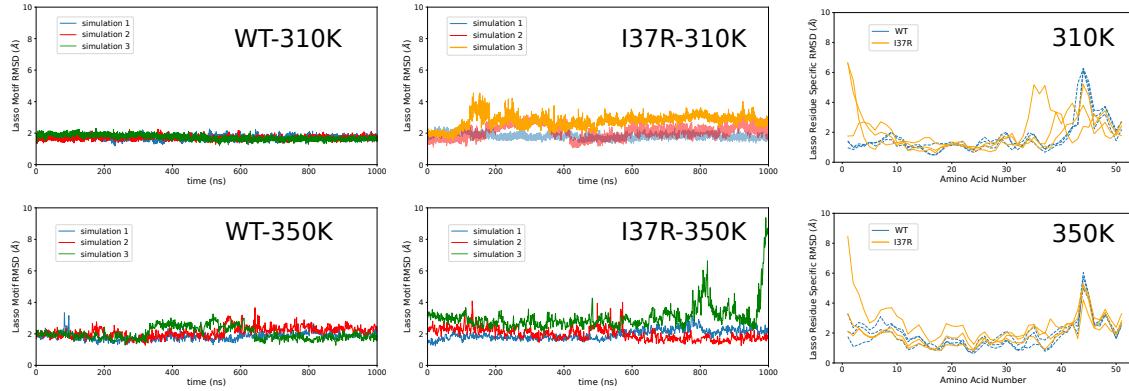


Figure 4.S3: Comparison between the stability of the lasso motif in I37R-CFTR and WT-CFTR at 310K and 350K. Related to 4.5D.

Root-mean-square deviation (RMSD) after molecular dynamics simulations of 24 models for the 17 unresolved residues of the R domain after the whole system was aligned to the transmembrane domains (TMDs) of the 6MSM structure.

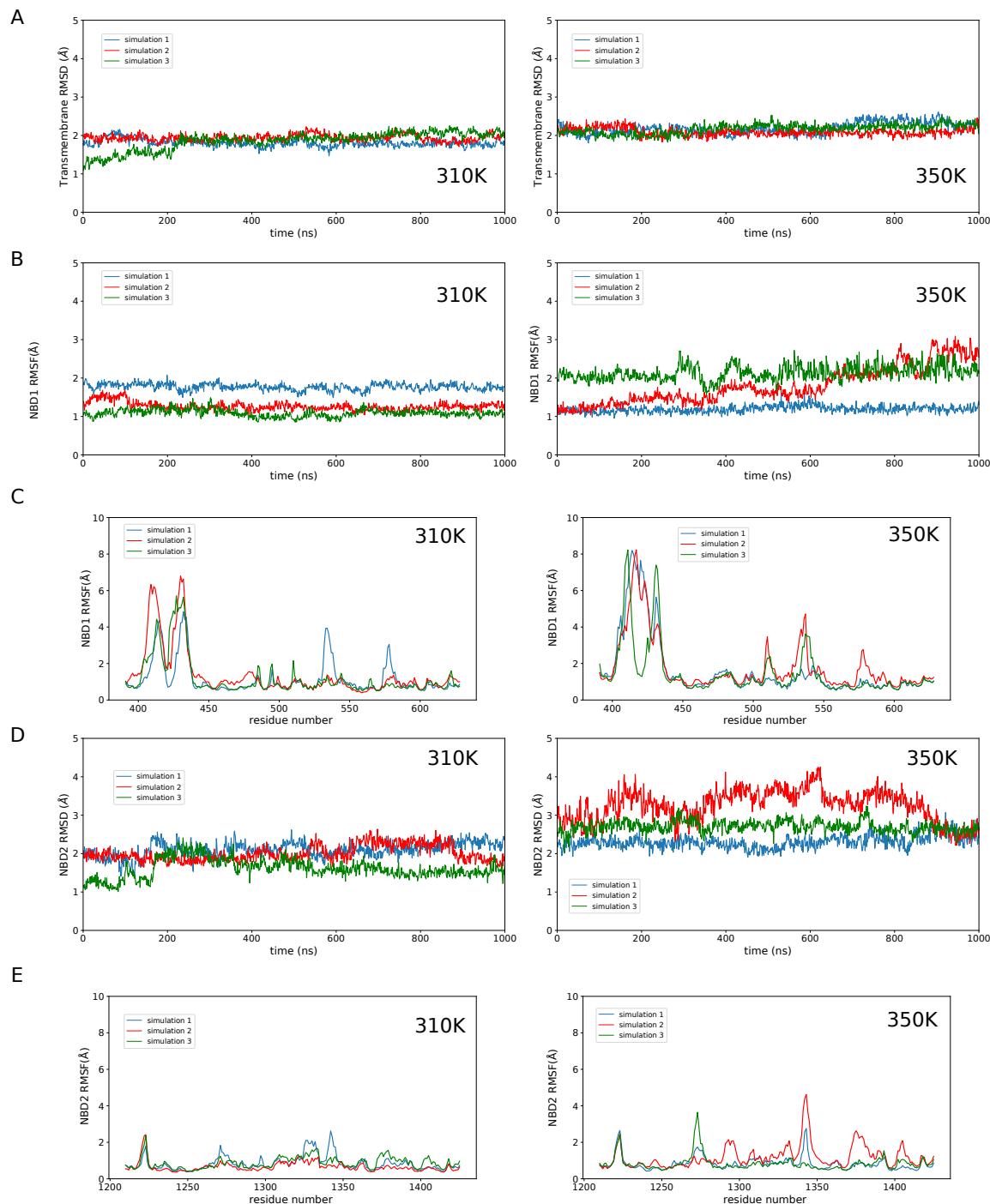


Figure 4.S4: Comparison of the stability of different WT-CFTR domains at 310K and 350K. Related to 4.5D.

(A) The transmembrane domains of WT-CFTR did not show significant conformational changes at the elevated temperature of 350K. (B) Small conformational changes in NBD1 were observed at 350K. The root-mean-square deviation (RMSD) for NBD1 was calculated using amino acids 391-628, excluding 401-440 as these amino acids compose the disordered Regulatory Insertion (RI). (C) The root-mean-square fluctuation (RMSF) profile for NBD1. The increase in RMSD at 350K was attributed to the already flexible sections. (D) NBD2 showed small conformational changes at 350K. The

RMSD for NBD2 was calculated using amino acids 1226-1423. (E) The RMSF profile of the amino acids in NBD2 indicated that already flexible sections of NBD2 changed conformation at the elevated temperature while the rest of the domain remained stable.

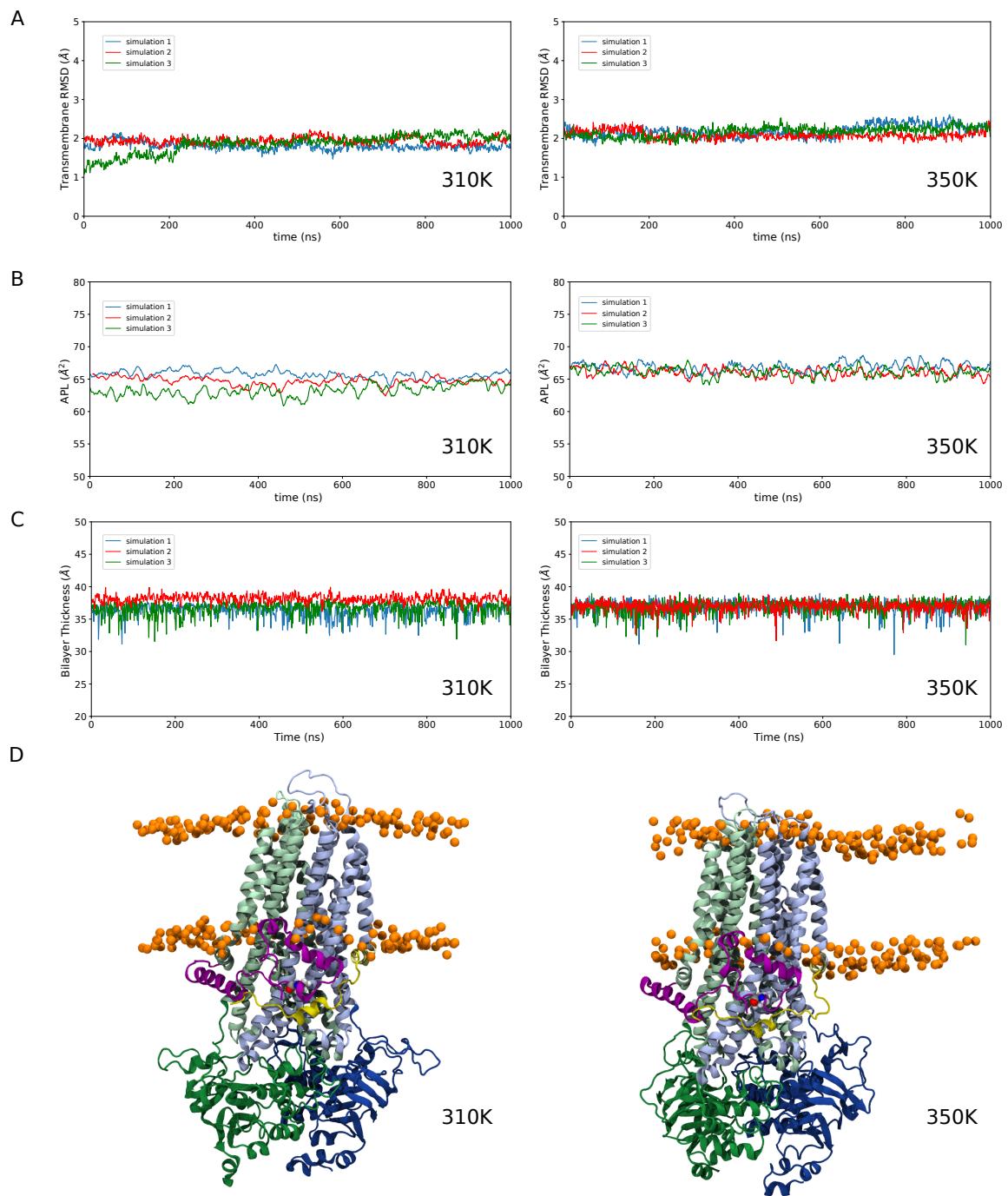


Figure 4.S5: Comparison of the stability of the bilayer surrounding WT-CFTR at 310K and 350K. Related to Fig. 4.5D

(A) The area per lipid (APL) for the modelled 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine

(POPC) bilayer was within expected tolerances at the elevated temperature of 350K. APL was plotted using a 10 ns moving average. (B) The thickness of the bilayer did not change significantly when simulated at 350K compared to 310K. (C) Visual inspection confirmed the stability of the bilayer at 350K (phosphate headgroups are depicted as orange spheres).

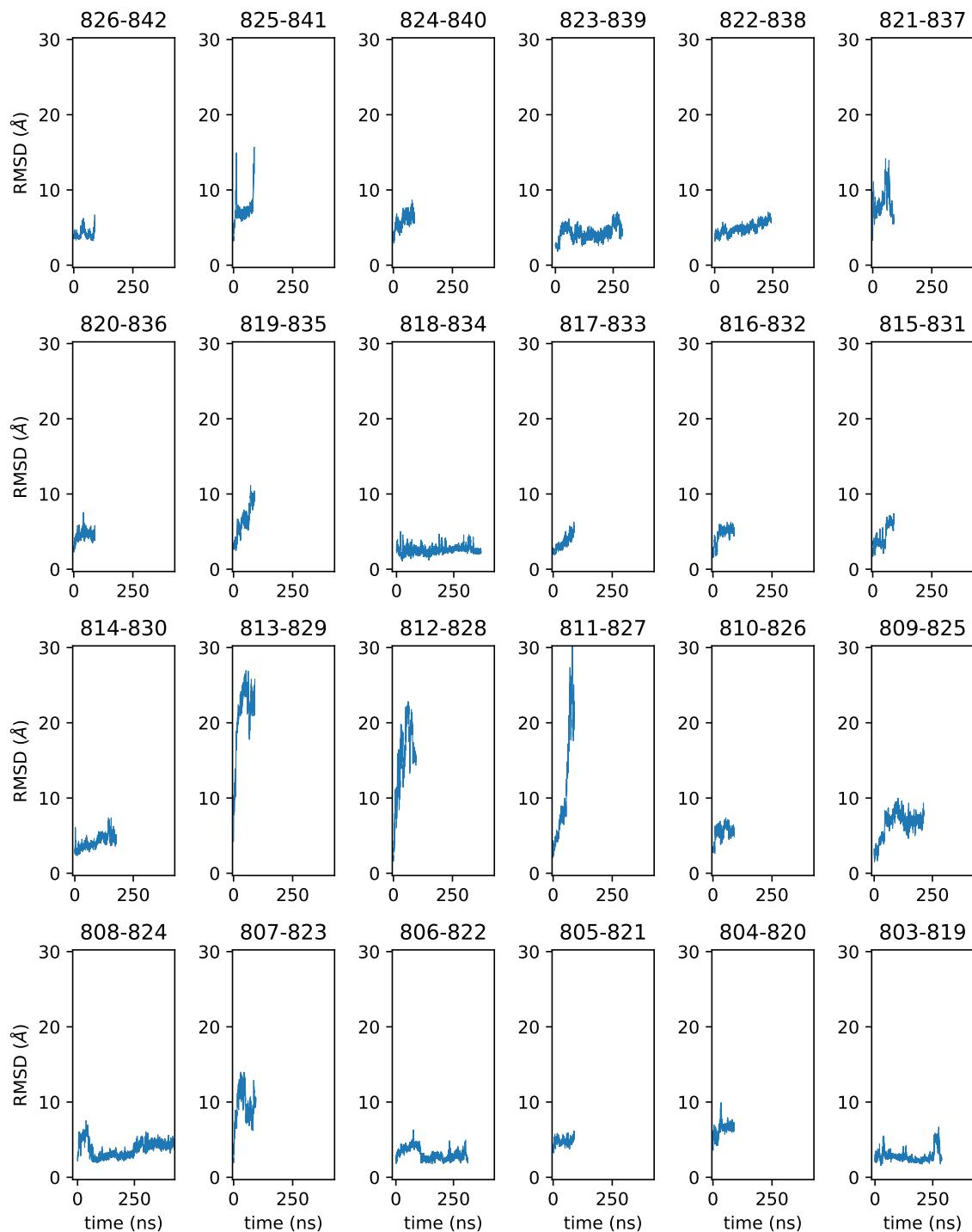


Figure 4.S6: Results from All Tested R-domain Models

Root-mean-square deviation (RMSD) after molecular dynamics simulations of 24 mod-

els for the 17 unresolved residues of the R domain after the whole system was aligned to the transmembrane domains (TMDs) of the 6MSM structure.

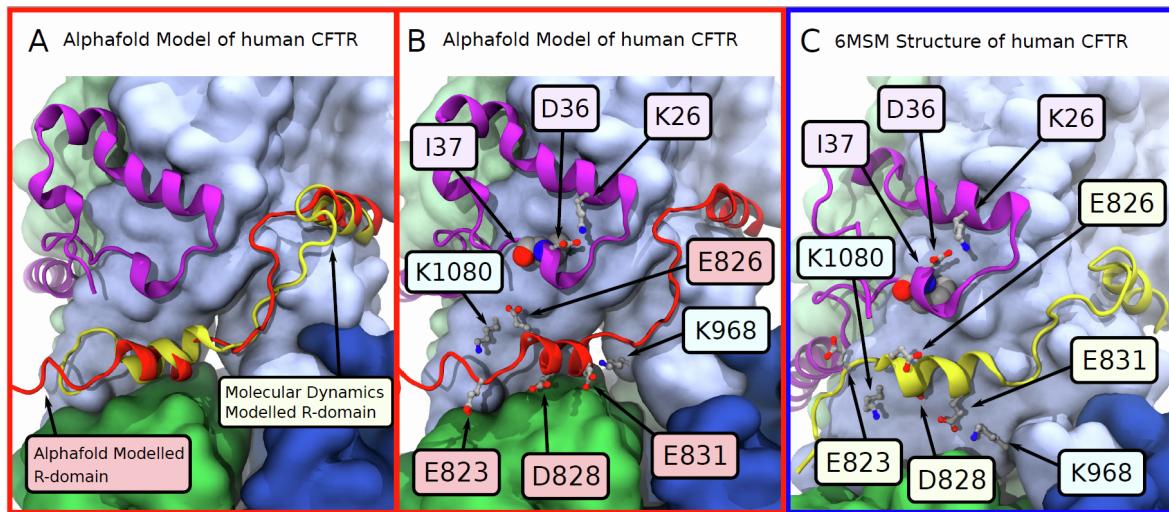


Figure 4.S7: A comparison between the prediction of the structure of human CFTR by Alphafold2 and the modelled segment of the R domain

Panels with red border indicate the background model in that panel is the Alphafold2 prediction, while the panel with blue border indicates the background model is that of 6MSM, the experimentally derived cryo-EM structure of human CFTR. The MD derived model of the R domain is coloured yellow while the Alphafold2 derived R domain is coloured red. (A) A closeup visualisation of the alignment between the Alphafold2 predicted R domain and the MD based prediction. This figure showed the agreement between the prediction of Alphafold2 and the MD model. The mouse, rat, human and zebrafish CFTR structures in the Alphafold2 database display the same tertiary structure, linking the helical R domain segment closest to the lasso motif, to the beginning of TMD2. This is consistent with the assignment of L818-F834 to the unknown segment of the 6MSM model using MD. (B) The placement of important charged sidechains in the structure predicted by Alphafold2. (hCFTR Alphafold2 model: <https://alphaFold.ebi.ac.uk/entry/P13569>) (C) The placement of important charged sidechains in the 6MSM model, including the modelled helix.

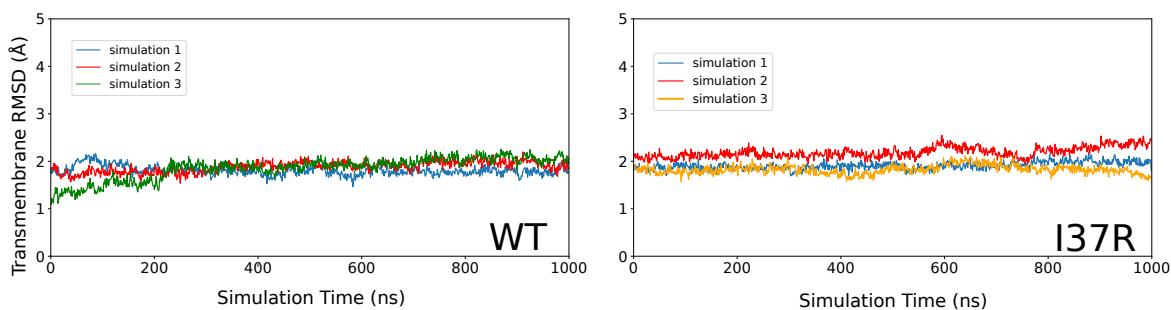


Figure 4.S8: Comparison between the stability of the transmembrane domains in I37R-CFTR and WT-CFTR at 310K. Related to Fig 4.5D

Throughout the microsecond simulations, the transmembrane domain of I37R-CFTR did not deviate significantly from the WT-CFTR indicating no allostery in the transmembrane domains of the I37R mutation. The data coloured orange denotes the sim-

ulation described in detail in the results section.

4.5 Method Details

Intestinal Current Measurement Superficial rectal mucosa samples (2 – 4 per donor) were freshly obtained using biopsy forceps (CK Surgitech NBF53-11023230) and placed in cold RPMI1640 media (Sigma R5886) with 5% FBS. Intestinal current measurements were performed under voltage-clamp conditions using VCC MC8 Ussing chambers (Physiologic Instruments, San Diego, CA) [491–493]. Biopsy tissues were bathed in Ringer solution containing (mM) 145 NaCl, 3.3 K₂HPO₄, 0.4 K₂HPO₄, 10 D-Glucose, 10 NaHCO₃, 1.2 MgCl₂ and 1.2 CaCl₂. Ringer solutions were continuously gassed with 95% O₂-5% CO₂ and maintained at 37°C. 10 μM indomethacin was added to both apical and basal chambers, and tissues were stabilised for 40 min. Tissues were then treated with pharmacological compounds (in order): 100 μM amiloride (apical) to inhibit epithelial sodium channel (ENaC)-mediated Na⁺ flux, 10 μM forskolin + 100 μM IBMX cocktail (apical and basal) to induce cAMP activation of CFTR, 100 μM carbachol (basal) to increase intracellular Ca²⁺ levels and activate basolateral Ca²⁺-dependent K⁺ channels and 100 μM bumetanide (basal) to inhibit basolateral Na⁺/K⁺/2Cl⁻ (NKCC) co-transporter.

Forskolin-Induced Swelling Assay Passage 3-15 organoids were seeded in 96-well plates, in 4 μl 70% matrigel droplet per well containing 25–30 organoids. The next day, organoids were incubated with 1.84 μM calcein green (Thermo Fisher Scientific C3100MP) for at least 30 min prior to addition of fsk at 0.02, 0.128, 0.8 or 5 μM concentrations, to determine cell viability. For CFTR potentiation, a single potentiator (3 μM VX-770 or 3 μM G1837 or 50 μM Gen) or dual potentiators (VX-770+G1837 or VX-770+Gen or G1837+Gen) was added together with fsk. Time-lapse images of organoid swelling were acquired at 10-min intervals for 60 min at 37°C using Zeiss Axio Observer Z.1 inverted microscope (Carl Zeiss, Jena, Germany), on an EC Plan-Neofluar 5x/0.16 M27 dry objective. Organoids were pre-incubated with 3 μM VX-809 or 3 μM VX-661 or 3 μM VX-445 or 3 μM VX-445+18 μM VX-661 for 24 h prior to FIS for CFTR correction where indicated. Three wells were used per condition and each participant’s FIS experiment was repeated 3 to 4 times.

Quantification of Forskolin-Induced Swelling Organoid swelling was quantified using a custom-built script. A segmentation strategy implemented using ImageJ/Fiji was performed on brightfield images. The raw image was processed with a gaussian blur (s=1.3) to reduce noise. After the directionality and magnitude of the local gradient was identified, pixels were classified as either ‘Background’, ‘Ridge’, ‘Valley’, ‘Rising’ or ‘Falling’ dependent on their neighbouring pixels along the previously calculated local directionality. Clean-up filters were applied that remove noise and small objects, such as ridges that only touched background pixels, and erosions to decrease rising and falling edges to better approximate object boundaries (‘Peaks’). A size exclusion was applied that would discriminate debris in the sample preparation from organoids of interest. This segmentation strategy was used to identify area covered by organoid at

each time point. The total surface area of organoid at 10-min intervals over 60 min post-fsk stimulation were calculated and normalized against t=0 to render the relative amount of swelling from t=0. The area under the curve, AUC (calculated increase in organoid surface area from t=0 to t=60; baseline=100%) was then calculated using GraphPad Prism software.

Quantification of CFTR-Mediated Ion Transport in Organoid-Derived Monolayers Short circuit current (I_{sc}) measurements were performed under voltage-clamp conditions using VCC MC8 Ussing chambers (Physiologic Instruments, San Diego, CA). Cells were bathed in 20 mM HEPES buffered-Ringer solution containing (mM): 120 NaCl, 0.8 K₂HPO₄, 5 D-Glucose, 1.2 MgCl₂ and 1.2 CaCl₂. Ringer solutions were continuously gassed with 95% O₂-5% CO₂ and maintained at 37°C. 10 μM indomethacin was added to both apical and basal chambers and cells were stabilised for 15 min. Cells were then treated with pharmacological compounds (in order): 100 μM amiloride (apical) to inhibit epithelial sodium channel (ENaC)-mediated Na⁺ flux, vehicle control 0.01% DMSO or 10 μM G1837 (apical) to potentiate cAMP-activated currents, 5 μM forskolin (basal) to induce cAMP activation of CFTR, 30 μM CFTRinh-172 (apical) to inhibit CFTR-specific currents and 100 μM ATP (apical) to activate calcium-activated chloride currents. I_{sc} in response to forskolin was considered as baseline activity (Δ I_{sc}-Fsk) and I_{sc} in response to forskolin and potentiator (Δ I_{sc}-Fsk+Pot) was used as the measure of modulator response.

Immunofluorescence A rectal biopsy from a I37R/F508del participant was embedded in Tissue-Tek Optimal Cutting Temperature (OCT) compound (Sakura Finetek, CA) and snap frozen prior to storage at -80°C. The frozen biopsy was cut into 4 μm slice sections, and the sections were fixed in ice-cold methanol for 15 min. Intestinal organoids cultured from the I37R/F508del participant and organoid-derived monolayers cultured from a F508del/F508del participant were fixed in 4% paraformaldehyde and ice-cold methanol respectively for 15 min. Fixed samples were blocked using IF buffer (0.1% BSA, 0.2% Triton and 0.05% Tween 20 in PBS) with 10% normal goat serum (Sigma G9023) for 1 h at room temperature before incubation in primary antibodies overnight at 4°C. The biopsy section was stained with CFTR (1:50, Abcam ab2784) and E-cadherin (1:100, Cell Signalling 3195) antibodies. Intestinal organoids were stained with Ki67 (1:250, Abcam ab15580) and E-cadherin (1:250, Life Technologies 13-1700) antibodies. Organoid-derived monolayers were stained with ZO-1 (1:250, Life Technologies 61-7300) and E-cadherin (1:250, Life Technologies 13-1700) antibodies. On the following day, samples were washed with IF buffer 3 times, 5 min each and incubated with Alexa Fluor conjugated secondary antibodies (1:500, Life Technologies A-11029, A-21329) for 1 h at room temperature. Samples were mounted with Vectashield hardset antifade mounting medium containing DAPI (Vector Laboratories H-1500). Images were acquired using Leica TCS SP8 DLS confocal microscope (Leica Microsystems, Wetzlar, Germany), either on a 63x/1.4 or a 20x/0.75 objective. Images were processed using ImageJ (National Institutes of Health, Bethesda, MD).

Western Blotting Intestinal organoids were lysed with TNI lysis buffer (0.5% gelatin, 50 mM Tris pH 7.5, 250 mM NaCl, 1 mM EDTA) [494] containing protease

inhibitor cocktail (Roche 04693159001) on ice for 30 min. Lysates were then sonicated using the Bioruptor Pico (Diagenode, Liège, Belgium) at 4°C for 20 cycles of 30 sec on and 30 sec off. Lysates were spun down at 14,000 rpm at 4°C for 20 min and protein concentrations were determined using the BCA Protein Assay Kit (Thermo Fisher Scientific 23225). Lysates (100 µg per sample) were separated using NuPAGE 3 – 8% Tris-Acetate gels (Thermo Fisher Scientific EA0375BOX) at 100 V for 30 min, followed by 150 V until separation was complete. Proteins were transferred onto a nitrocellulose membrane using wet transfer at 20 V for 1 h at RT. The membrane was then incubated in 5% non-fat dry milk in phosphate-buffered saline containing 0.1% Tween (PBST) for 1 h at RT. CFTR bands were detected using anti-CFTR antibody 596 (1:500; University of North Carolina, Chapel Hill and Cystic Fibrosis Foundation) incubated at 4°C overnight. Protein bands were visualised using ECL Select detection reagent (Cytiva RPN2235) on the ImageQuant LAS 4000 (GE Healthcare, Chicago, IL). Calnexin was used as the loading control, detected using anti-calnexin antibody (1:1000; Cell Signalling Technology 2679). Protein band densitometry was performed using ImageJ (National Institutes of Health, Bethesda, MD). CFTR maturation in I37R/F508del and F508del/F508del organoids were estimated by measuring the level of mature mutant CFTR (band C) as a percentage of mature CFTR from WT organoids (% normal CFTR) [495].

In Silico System Composition A 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayer was generated using the VMD membrane builder plugin (Humphrey et al., 1996) in which a model based on the phosphorylated human CFTR channel (PDB ID: 6MSM) was embedded [45]. The system was solvated with TIP3P water and neutralised with 0.15 M of potassium chloride ions [496]. The WT-CFTR system included 236 POPC molecules, 128 potassium ions, 140 chloride ions and 44503 water molecules.

Extended 6MSM Structure: Modelling the Unidentified Section of the R Domain The 6MSM structure was extended in order to resolve a previously unassigned section in the R domain. The R domain is 227 residues long (F630-H856) and is largely disordered [307]. In the 6MSM structure, the sidechains of 17 residues of this domain are labelled “UNKNOWN”, due to inadequate electron density in the region. The first 8 residues are unstructured while the next 9 residues form an alpha helix. The distance between the end of the helix and the first visible residue in TMD2 (T845) is 20 Å [45]. Using VMD’s autopsf plugin [232] we populated the side chains of the unknown section. Modeller 9.19 was then used to link the R domain to TMD2 at T845 [236]. 24 possible primary structure alignments of this region were simulated. The Root Mean Squared Deviation (RMSD) of the backbone alpha carbon atoms of the extended section with respect to the 6MSM structure was calculated over 300 ns of MD simulations. The most stable alignment was chosen from the lowest RMSD compared to the 6MSM structure. The most stable configuration was capped with the neutral forms of the C and N termini and incorporated into our CFTR model. Four other missing loops namely residues 410-434, 890-899, 1174-1201, 1452-1480 were reconstructed using Modeller 9.19, based on visual analysis and the lowest discrete optimised protein energy (DOPE) score [237]. The N and C termini of the CFTR model were capped

with the physiological, charged termini.

Molecular Dynamics Simulation Protocols The 6MSM structure carries an engineered mutation to avoid the hydrolysis of the bound ATP, giving it a longer lifetime in the open conformation (E1371Q). This mutation was corrected to match the WT-CFTR sequence using the mutator plugin of VMD. The I37R missense mutation was constructed in the same way. GROMACS v2019.3 with the CHARMM36m forcefield was used for all MD simulations [123, 157]. Minimisation via a steepest descent algorithm was performed until all forces were below 24 kcal/mol/Å. This was followed by relaxation simulations of all heavy atoms in the system starting with a restraint of 10 kcal/mol/Å² and then halving this restraint every 200 ps in 15 iterations. Relaxation and production were run with 1 and 2 fs time steps, respectively. Relaxation was followed by 5 ns of equilibration. During relaxation, a Berendsen thermostat and barostat were applied, and for production a Nosé-Hoover and Parrinello-Rahman thermostat and barostat were applied respectively [155, 162, 497]. To maintain the area per lipid (APL) properties of the POPC membrane at experimental values during production runs, pressure coupling was applied in the z-direction normal to the membrane bilayer while the x-y dimensions of the cubic simulation volume was fixed [498]. While semi-isotropic pressure coupling better replicates membrane environments [499], this constant area approach was adopted to circumvent an issue with GROMACS 2019.3reb (<https://gitlab.com/gromacs/gromacs/-/issues/2867>). Production runs were extended up to 2 μs at 310 K with three replicates for all simple MD simulations. The last 1 μs of the longest simulations for each system were selected for further analysis. This was the longest time feasible to simulate with available computational resources. All RMSDs were calculated using the positions of alpha carbons with reference to the 6MSM experimental structure [45]. Analysis scripts were written in python using the MDAnalysis library [234, 235]. Bilayer thickness and area per lipid were calculated with the FATSLIM software package [500].

Mathematical Formulae

Root Mean Square Deviation (RMSD)

$$RMSD(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_i(t) - x_i^{ref}|^2} \quad (4.1)$$

Root Mean Square Fluctuation (RMSF)

$$RMSF_i = \sqrt{\left\langle |x_i(t) - x_i^{ref}|^2 \right\rangle_t} \quad (4.2)$$

Quantification and statistical analysis Data for Figures 4.2A, 4.2G and 4.3D are presented as dot plots with mean ± standard error of the mean (SEM). Data for Figures 4.2D, 4.3A and 4.3E are presented as line graphs with mean ± standard deviation (SD). Data in Figures 4.2E, 4.3C, 4.3G, 4.4A, and 4.4B are presented as violin plots with mean. One-way analysis of variance (ANOVA) or unpaired t-test was used to determine

statistical differences as indicated. Statistical analysis was performed with GraphPad Prism software v9.0.1. A p value of less than 0.05 was considered to be statistically significant.

4.6 Acknowledgements

We thank the study participants and their families for their contributions. We also thank Sydney Children's Hospital's (SCH) Randwick respiratory department especially Leanne Plush, Amanda Thompson, Roxanne Strachan, and Rhonda Bell for the organization and collection of participant biospecimens. SAW is supported by an Australian National Health and Medical Research Council grant NHMRC_APP1188987. MA acknowledges support of a top-up scholarship from Cystic Fibrosis in Australia. MA and KA are supported by Australian Government Research Training Program Scholarship. We acknowledge for the generous provision of the L-Wnt3A cell line. Computations were performed on the Gadi HPC at the National Computational Infrastructure Center in Canberra and Artemis at the Sydney Informatics Hub in The University of Sydney. We thank Dr John R. Riordan (University of North Carolina-Chapel Hill) and Cystic Fibrosis Foundation for providing anti-CFTR antibody #596. Support statement: This work was supported in part by an Australian National Health and Medical Research Council grant (NHMRC_APP1188987), a Rebecca L. Cooper Foundation project grant, a Cystic Fibrosis Australia-The David Millar Giles Innovation Grant, Sydney Children Hospital Network Foundation, and Luminesce Alliance Research grants.

4.7 Author Contributions

Conception and design: SAW and AJ. Recruitment and consent: LF and SAW. Collection of rectal biopsies: CYO and LF. Ion transport assay: NTA. Culturing of organoids: NTA, SLW, and SAW. FIS microscopy: IS, KA, and SLW. FIS scripts: MC and RW. FIS analysis: NTA and SLW. Immunofluorescence microscopy: SLW. Western blot: SLW. Molecular Dynamics: MA, PC, RG, and SK. CFTR sequence alignment: AC. Figure preparation: SLW, MA, NTA, AC, KA, and SAW. Writing – original draft: SLW, MA, and SAW. Review and editing: SAW, KA, RG, SK, and LF with intellectual input from all other authors. Supervision: SAW and SK.

4.8 Delcaration of Interests

SAW is the recipient of a Vertex Innovation Grant (2018) and a TSANZ/Vertex Research Award (2020). Both are unrelated and outside of the submitted manuscript. AJ has received consulting fees from Vertex on projects unrelated to this study. CYO has acted as consultant and is on advisory boards for Vertex pharmaceuticals. These works are unrelated to this project and manuscript. All other authors declare no conflict of interest.

5

Molecular Dynamics and Theratyping in Airway and Gut Organoids Reveal R352Q-CFTR Conductance Defect

Cells have a mind of their own

– Shafagh A. Waters (personal communication)

Preamble on Formatting

The following chapter has been published in the *American Journal of Respiratory Cell and Molecular Biology* [304]. It is typeset here to match the style of the rest of the thesis. To save space, some supplemental information has not been reprinted. It can be accessed at the source publication. Only supplementary figures and tables pertaining to *in silico* studies have been reprinted and these share the numbering of the source publication. For example Figure E4 in the source publication will appear in this chapter as Figure 5.E4. On the other hand, supplementary figures and tables pertaining to experimental results, which are not explicitly included in this chapter are denoted by the symbol *. For example Figure E2 from the source publication is denoted in this chapter as Figure 5.E2*.

Abstract

A significant challenge to making targeted CFTR modulator therapies accessible to all individuals with cystic fibrosis (CF) are many mutations in the CFTR gene that can cause CF, most of which remain uncharacterized. Here, we characterized the structural and functional defects of the rare CFTR mutation R352Q – with potential role contributing to intrapore chloride ion permeation – in patient-derived cell models of the airway and gut. CFTR function in differentiated nasal epithelial cultures and matched intestinal organoids was assessed using ion transport assay and forskolin-induced swelling (FIS) assay respectively. CFTR potentiators (VX-770, GLPG1837 and VX-445) and

correctors (VX-809, VX-445 +/- VX-661) were tested. Data from R352Q-CFTR were compared to that of twenty participants with mutations with known impact on CFTR function. R352Q-CFTR has residual CFTR function which was restored to functional CFTR activity by CFTR potentiators but not the corrector. Molecular dynamics (MD) simulations of R352Q-CFTR were carried out which indicated the presence of a chloride conductance defect, with little evidence supporting a gating defect. The combination approach of in vitro patient-derived cell models and in silico MD simulations to characterize rare CFTR mutations can improve the specificity and sensitivity of modulator response predictions and aid in their translational use for CF precision medicine.

5.1 Introduction

Cystic fibrosis (CF), a rare, life-limiting disorder affecting ~90,000 individuals worldwide [501], results from a malfunction in the cystic fibrosis transmembrane conductance regulator (CFTR) protein. CFTR functions as a cAMP-stimulated anion channel [502]. Chloride ions diffuse through the CFTR channel pore, causing a charge imbalance that is corrected by the flux of sodium ions through sodium channels [502]. The resultant ion imbalance causes osmosis of water into the extracellular space. More than 2,000 variants in the CFTR gene are known. CFTR mutations are grouped into six major classes based on their functional defects in protein synthesis (I); intracellular maturation, processing, or folding (II); channel gating (III) or conductance activity (IV); diminished quantity (V); and diminished stability (VI) [503].

CFTR modulators can restore mutant CFTR protein function [432]. Two subclasses of modulators are in current clinical use, though approval varies between countries. Potentiators such as ivacaftor (VX-770|Kalydeco; Vertex Pharmaceuticals) increase CFTR channel gate opening. Correctors such as lumacaftor (VX-809), tezacaftor (VX-661), and elexacaftor (VX-445) increase delivery of misfolded CFTR to the cell surface. VX-809 and VX-661 are type I correctors that work by stabilizing the NBD1-TMD1 and/or NBD1-TMD2 interface by binding directly to TMD1 [452, 453]. VX-445 is a type III corrector that directly stabilizes NBD1 and has been shown to have copotentiator activity [456, 471, 472, 488]. Combination therapies of potentiator and corrector(s) (Orkambi, Symdeko, Trikafta; Vertex Pharmaceuticals) are approved for use in patients with CF homozygous for F508del, the most common CFTR mutation [504–506]. Trikafta is also approved for patients with CF who are heterozygous for F508del [504]. The remaining patients with CF are either compound heterozygous or homozygous for mutations other than F508del, only some of which are approved for treatment with modulator therapy [507]. A large number of CFTR mutations are not approved for modulator therapy, because their exact mechanism of CFTR dysfunction and/or responsiveness to modulator therapy is often unknown. Traditional randomized clinical trials of modulators to identify modulator-responsive patients with CF with rare CFTR mutations are costly, time-consuming, and impractical [508].

A major breakthrough in the field of CF has been the development of two preclinical patient- and organ-specific cell models [477, 509, 510]. Both models are used in functional assays that allow rapid quantitative measurements of CFTR function. Differentiated airway epithelial cells are used in transepithelial ion transport assays (short

circuit current [I_{sc}]) [511]. Intestinal organoids are used in forskolin (Fsk)-induced swelling (FIS) assays [512]. Each serves as a personalized functional model of the patient’s CFTR mutation and response to modulators. However, how cell models of the airway compare with those of the gut created from the same patient is not well understood. Mean changes in CFTR function in vivo correlate with the CFTR rescue assessed via the I_{sc} and FIS assays in cell models of subjects with the same mutations [477, 511]. In addition, cell model responses in subjects with ultrarare mutations have successfully predicted clinical benefit [457–459]. Results so far support the use of patient-derived cell models to guide personalized treatment [429].

We established human nasal epithelial cell (hNEC) cultures from 11 participants with wild-type functioning CFTR (WT/WT) and 10 participants with CF. The participants with CF included five participants with known and characterized CFTR folding/maturation defects (F508del/F508del), three with a CFTR synthesis defect (G542X/F508del), and one with a CFTR gating defect (G551D/F508del). The results from these CFTR mutations with a known impact on CFTR function were used as a reference and were compared with the results from one participant with the R352Q/F508del CFTR genotype. We also created matched intestinal organoids from seven participants with CF and one WT-CFTR control participant (see Table 5.E1* in the data supplement). R352Q, a rare CFTR mutation in the CFTR channel pore, was chosen because it has been characterized in heterologous systems but is yet to be tested in a patient-derived model. Single-channel and whole-cell patch-clamp studies have demonstrated R352 to be a molecular determinant of anion selectivity and permeability in the CFTR channel pore [513]. Our aims were to assess CFTR baseline activity and CFTR modulator response in I_{sc} and FIS assays. Potentiator agents VX-770 and GLPG1837 (a potentiator in clinical trials; hereafter G1837) were tested individually and in combination with a CFTR protein-folding corrector VX-809. The effect of VX-445 as a potentiator (acute treatment) and corrector (chronic treatment) monotherapy, as well as part of triple combination therapy VX-445+VX-661+VX-770 (Trikafta) were assessed in FIS assays. In addition, to gain a complementary and better understanding of the effect of R352Q mutation on CFTR function, *in silico* structural atomistic modeling and molecular dynamics simulations were performed. Some of the results of these studies have been reported previously in the form of a preprint (<https://doi.org/10.1101/2021.08.11.456003>).

5.2 Results

5.2.1 R352Q-CFTR Baseline Activity and Response to CFTR Modulators in Nasal Epithelial Cells

Mature, differentiated hNECs were pseudostratified (Figure 5.1A) and had functional beating cilia (CBF, 6.2 ± 0.1 Hz) (Figure 5.1B, Video 1) and intact junction integrity (Figure 5.1A) (transepithelial electrical resistance, $\geq 170 \Omega \cdot \text{cm}^2$; Figure 5.E1A*). To assess ion transport, I_{sc} measurements were performed (Figures 5.E1B–E1D*). We first determined the CFTR activity threshold in reference cell models, which were used for comparison with R352Q-CFTR functional activity before and after modulator stimula-

tion. Fsk-stimulated CFTR-dependent anion currents (ΔI_{sc} -Fsk) in WT/WT hNECs were $21.2 \pm 1.2 \mu A/cm^2$ at baseline (Figures 5.1C and 5.1D, Table 5.E2*). Baseline ΔI_{sc} -Fsk in F508del/F508del hNECs was at $3.4 \pm 0.5 \mu A/cm^2$ and was not increased beyond $0.8 \mu A/cm^2$ with either potentiator (Figure 5.1D). Because of the presence of two copies of the same CFTR mutation, we considered that protein expression from each of the F508del alleles was likely to be the same and therefore attributed each allele to equally contribute $1.7 \mu A/cm^2$ to baseline Fsk-stimulated currents and $0.4 \mu A/cm^2$ to potentiator-stimulated currents (Table 5.E2*). The baseline ΔI_{sc} -Fsk of G542X/F508del hNECs was $0.81 \pm 0.12 \mu A/cm^2$. Because G542X is a synthesis defect mutation with no functional CFTR protein production, the observed $0.81 \mu A/cm^2$ is attributed to the F508del allele, which is not considerably different from that calculated from the F508del/F508del hNECs. We thus used these values as a guide to estimate the contribution of the F508del allele to the experimental I_{sc} data of the heterozygous R352Q/F508del participant. In support of this, F508del/WT hNECs were previously shown to have, on average, 50% of Fsk-stimulated currents of WT/WT hNECs, although variability between F508del/WT participants was present [511].

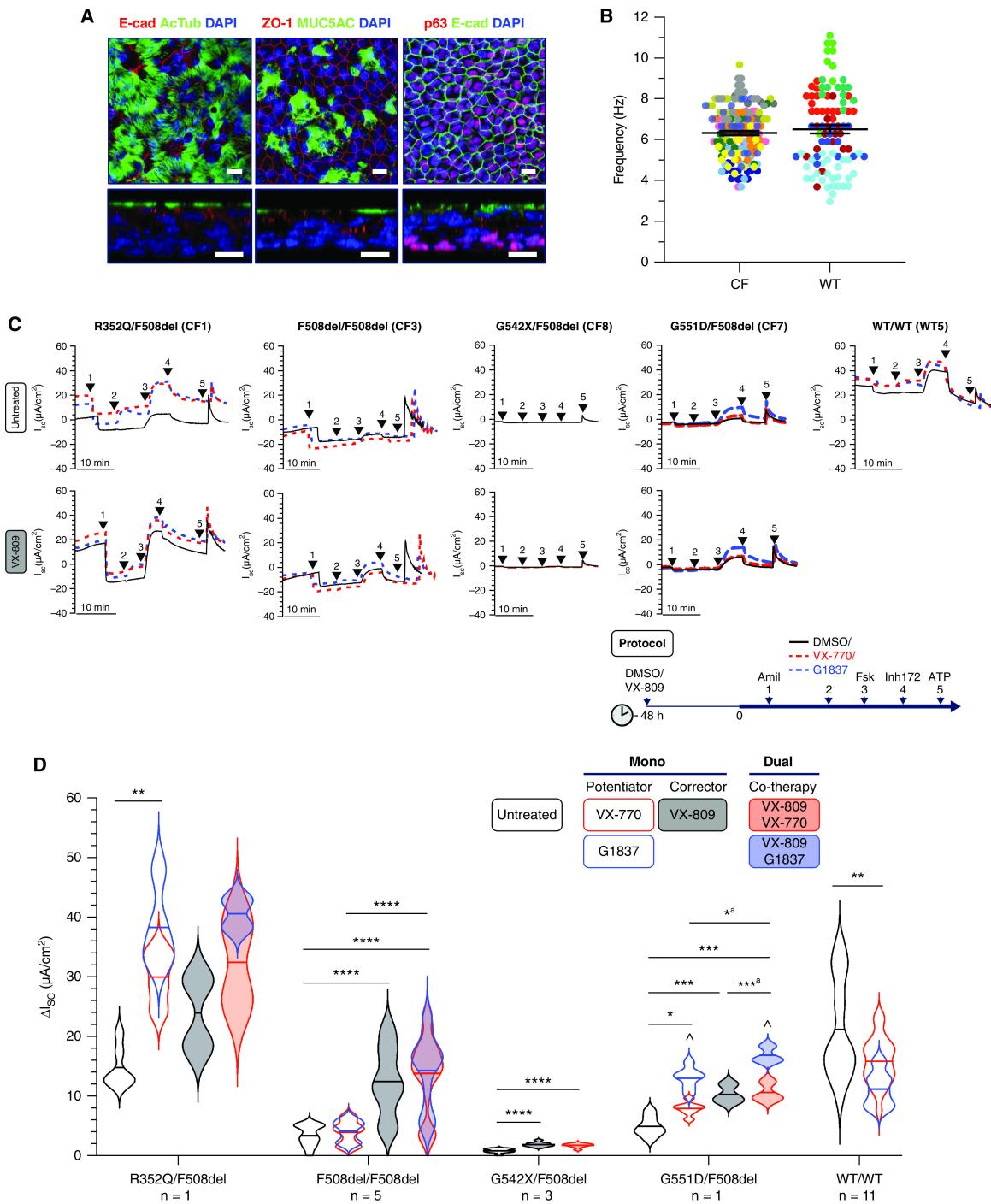


Figure 5.1: Functional response of R352Q-CFTR to cystic fibrosis transmembrane conductance regulator (CFTR) modulators in differentiated human nasal epithelial cells (hNECs).

(A) Immunofluorescence staining of acetylated tubulin (ciliated cells), MUC5AC (goblet cells), and p63 (basal progenitor cells) in hNECs derived from a R352Q/F508del participant (CF1). E-cadherin (adherens junction) and ZO-1 (tight junction) are localized to the intercellular junctions of epithelial cells. The top panels are top views and the bottom panels are side views showing the pseudostratified epithelium. A 63 \times /1.4 numerical aperture oil immersion objective was used. Scale bars, 10 μ m. (B) Dot plots.

of mean ciliary beat frequency (CBF) measurements (Hz) of the fully differentiated mature hNECs from participants with cystic fibrosis (CF) and without CF (wild type [WT]). Each participant is coded with a different color. Each dot represents a single field of view of CBF measurement. (C) Representative Ussing chamber recordings of short circuit current (I_{sc}) in hNECs from participants with CF and WT-CFTR control participants. The protocol used to measure functional CFTR expression in hNECs in 0.01% DMSO vehicle (untreated; top) or pretreated with corrector (3 μ M VX-809 for 48 h; bottom) followed by sequential addition of 100 μ M apical amiloride (1. Amil), apical addition of either vehicle control 0.01% DMSO (black) or 10 μ M VX-770 (red) or 10 μ M G1837 (blue) (2. DMSO, VX-770, G1837), 10 μ M basal forskolin (3. Fsk), 30 μ M apical CFTR inhibitor (4. CFTRinh-172), and 100 μ M apical ATP (5. ATP). A basolateral-to-apical chloride gradient was used. (D) Violin plots of total currents stimulated by DMSO or VX-770 or G1837 plus Fsk in hNECs untreated or pretreated with VX-809. Data are from 1 R352Q/F508del participant, 5 F508del/F508del participants, 3 G542X/F508del participants, and 1 G551D/F508del participant with CF and 11 WT-CFTR control participants. n, number of participants. Data are represented as violin plots to show the distribution of data. One-way ANOVA was used to determine statistically significant differences. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001. aP for G1837 only. ^P for G1837 vs VX-770. Where VX-770 and G1837 (overlapping violin plots) both achieved statistical significance, the least significant P value is shown.

R352Q/F508del hNECs demonstrated baseline CFTR activity of $14.8 \pm 1.4 \mu\text{A}/\text{cm}^2$, an appreciable 70% of WT-CFTR activity (Figures 5.1C and 5.1D, Table 5.E2*). Potentiation with VX-770 or G1837 led to a significant ($P < 0.01$) twofold increase in CFTR activity, reaching 15.2 and 23.5 $\mu\text{A}/\text{cm}^2$ above baseline, respectively (~140–180% WT-CFTR activity). This response is similar to the potentiator-stimulated response observed in the reference gating G551D/F508del hNECs. The G551D/F508del hNECs had ΔI_{sc} -Fsk of $4.9 \pm 0.6 \mu\text{A}/\text{cm}^2$ at baseline, and treatment with either potentiator caused an approximately twofold increase in CFTR activity above baseline (net increase, VX-770, 3.0 $\mu\text{A}/\text{cm}^2$; G1837, 8.1 $\mu\text{A}/\text{cm}^2$) (Figures 5.1C and 5.1D, Table 5.E2*). The majority of the response to potentiators in R352Q/F508del hNECs was most likely contributed by the restored R352Q, because the total response is far greater than the 0.4 $\mu\text{A}/\text{cm}^2$ that we estimated to be attributed to the F508del allele (Figure 5.1D). We conclude that R352Q-CFTR has high baseline activity amenable to potentiator rescue in hNECs.

We next explored whether corrector monotherapy or cotherapy with potentiators increased R352Q-CFTR functional rescue. In F508del/F508del-CFTR cultures pretreated with VX-809, Fsk alone significantly ($P < 0.0001$) enhanced CFTR-mediated Cl⁻ currents by 3.7-fold, reaching 9.0 $\mu\text{A}/\text{cm}^2$ above baseline (Figures 5.1C and 5.1D, Table 5.E2*). We estimated each of the F508del alleles to equally contribute 4.5 $\mu\text{A}/\text{cm}^2$ to this increase in current. VX-809 cotherapy with either VX-770 or G1837 significantly ($P < 0.0001$) increased the currents stimulated by potentiator monotherapy by 10 $\mu\text{A}/\text{cm}^2$ (Figure 5.1D). In G542X/F508del hNECs, treatment with VX-809 resulted in a modest but significant ($P < 0.0001$) increase in Fsk-stimulated currents by 1.04 $\mu\text{A}/\text{cm}^2$ above baseline (Figures 5.1C and 5.1D, Table 5.E2*). In G551D/F508del

hNECs, VX-809 treatment significantly ($P < 0.001$) increased CFTR activity by $5.4 \mu\text{A}/\text{cm}^2$ above baseline (Figures 5.1C and 5.1D, Table 5.E2*). Because G551D mutation does not respond to VX-809 treatment [514], the contribution of the F508del allele was considered to be $5.4 \mu\text{A}/\text{cm}^2$, similar to that calculated from the F508del/F508del hNECs. In addition, VX-809 cotherapy with either VX-770 or G1837 also increased the respective potentiator monotherapy currents by $2.7 \mu\text{A}/\text{cm}^2$ and $3.8 \mu\text{A}/\text{cm}^2$, respectively (Figure 1D). Unlike F508del/F508del-CFTR, the $9.1 \mu\text{A}/\text{cm}^2$ increase in R352Q/F508del CFTR activity from the baseline in response to VX-809 monotherapy was not statistically significant ($P = 0.09$). VX-809 cotherapy with either potentiator also resulted in a nonsignificant increase in VX-770– or G1837-potentiated CFTR currents by $2 \mu\text{A}/\text{cm}^2$ (Figure 1D). Because VX-809 does not rescue R352Q, then R352Q-CFTR is unlikely to have a folding/maturation defect. The modest increase in current with VX-809 treatment was mostly if not fully imparted by the rescued F508del allele in the R352Q/F508del hNECs.

5.2.2 R352Q-CFTR Baseline Activity and Response to CFTR Modulators in Intestinal Organoids

We evaluated CFTR activity in matched intestinal organoids from the same participant with the R352Q/F508del genotype and those participants with reference F508del/F508del, G551D/F508del, and WT/WT genotypes using an FIS assay. To detect baseline and modulator-stimulated CFTR activity in the organoids, swelling was assessed at four Fsk concentrations ranging from 0.02 to $5 \mu\text{M}$. FIS of organoids was dependent on Fsk dose, CFTR genotype, and the individual participant (Figure 5.2A). At $0.8 \mu\text{M}$ Fsk, the optimal concentration for baseline FIS assessment [467], minimal swelling was observed for the F508del/F508del (area under the curve [AUC], 42.8 ± 19.4) and G551D/F508del organoids (AUC, 82.9 ± 20.1) (Figure 5.2A). R352Q/F508del and WT/WT organoids showed considerably higher FIS at the same $0.8 \mu\text{M}$ Fsk concentration, with AUCs of 196.3 ± 19.9 and 578.3 ± 61.6 , respectively (Figure 5.2A).

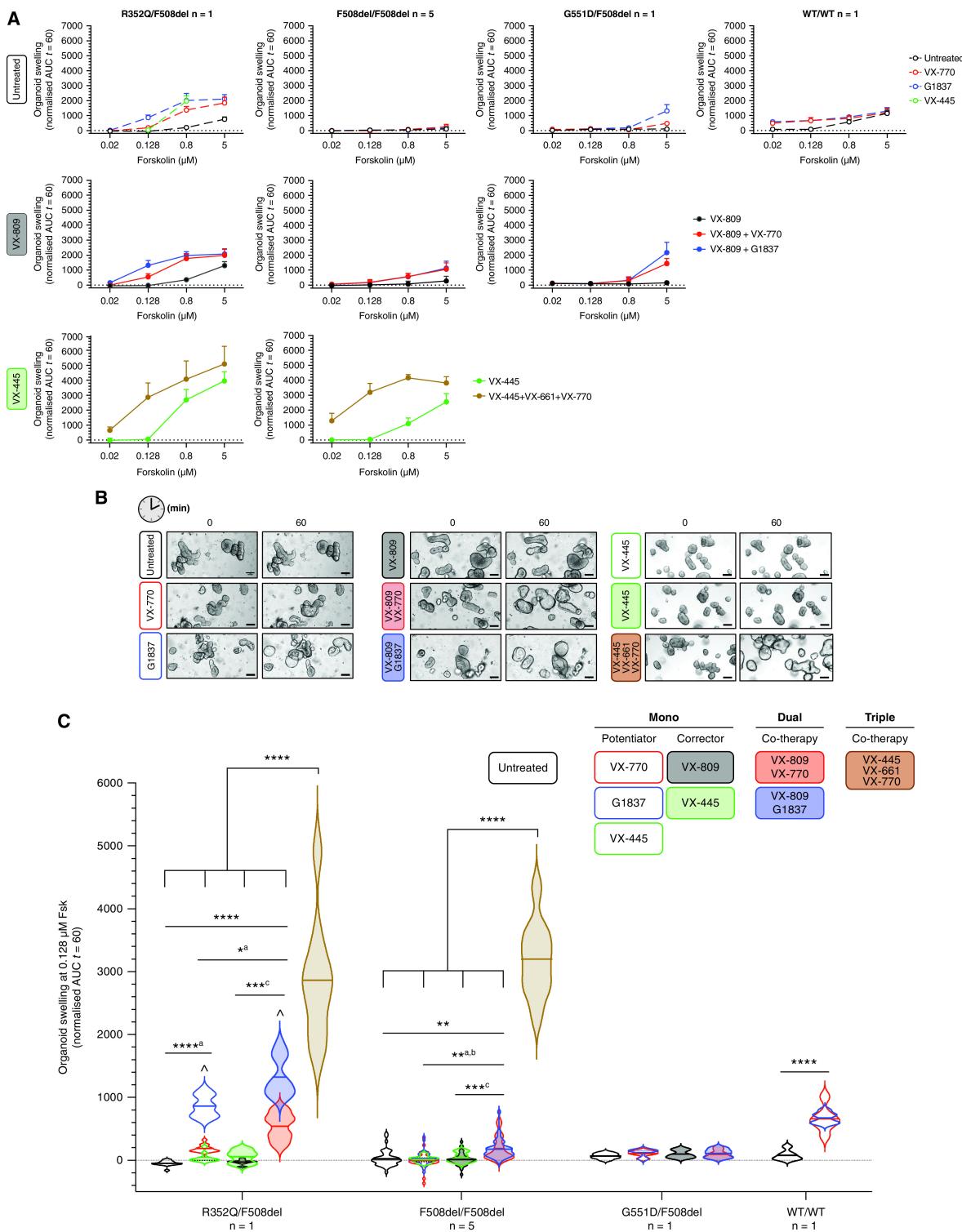


Figure 5.2: Functional response of R352Q-CFTR to CFTR modulators in intestinal organoids.

Forskolin-induced swelling (FIS) assay in organoids (left to right) from one R352Q/F508del participant, five F508del/F508del participants, and one G551D/F508del participant with CF and one WT-CFTR control participant. Organoids were incubated overnight with 0.03% DMSO (vehicle) or 3 μM VX-809 or 3 μM VX-445 or 3 μM VX-445 + 18 μM VX-661. After 24 hours, a range of Fsk concentrations from 0.02 to 5 μM

were acutely added, either alone or in combination with 3 μM VX-770 (red) or 3 μM G1837 (blue) or 3 μM VX-445 (green). (A) Values plotted are the mean \pm SD of the area under the curve (AUC). (B) Representative brightfield images of R352Q/F508del organoids at baseline ($t = 0$) and after 1 hour of stimulation ($t = 60$) at 0.128 μM Fsk. Scale bars, 100 μm . (C) Violin plots of FIS response (AUC at 0.128 μM Fsk) of organoids from participants with CF and WT-CFTR control participants, expressed as the absolute AUC calculated from the time periods $t = 0$ (baseline) to $t = 60$. Data are represented as violin plots to show the distribution of data. n, number of participants. One-way ANOVA was used to determine statistically significant differences. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001. aP for G1837 only, bP for VX-770 only, cP for VX-809 only, and \wedge P for G1837 versus VX-770 or VX-445. Where VX-770, G1837, and VX-445 (overlapping violin plots) achieved statistical significance, the least significant P value is shown.

FIS of intestinal organoids at 0.128 μM Fsk has been demonstrated to correlate strongly with clinical modulator responses and therefore is used to assess CFTR modulator response *in vitro* [467]. At this Fsk concentration, organoids—*independent* of CFTR genotype—did not demonstrate FIS greater than AUC of 67.1 (Figures 5.2B and 5.2C, Table 5.E3*, Figure 5.E2*). In F508del/F508del organoids, treatment with either VX-770 or G1837 did not induce any swelling, suggesting no improvement in CFTR activity with potentiator therapy. G551D/F508del organoids showed a modest 1.5-fold increase in FIS after treatment with either potentiator, increasing AUC by 38.9 (VX-770) and 54.2 (G1837) above baseline (Figure 5.2C, Table 5.E3*). In WT/WT organoids, potentiation with both VX-770 and G1837 led to a significant ($P < 0.001$) increase in FIS, increasing AUC by 580.6 and 594.2 above baseline, respectively (Figure 5.2C). In R352Q/F508del organoids, potentiation with either VX-770 or G1837 led to a significant increase in FIS, increasing AUC by 234.6 and 910.6 above baseline, respectively (Figures 2B and 2C). G1837 was significantly ($P < 0.0001$) more effective in increasing FIS when compared with VX-770 in R352Q/F508del organoids, but not in organoids with the other CFTR genotypes, wherein the efficacy of the potentiators did not significantly differ from each other (Figure 5.2C, Table 5.E3*). Consistent with previous studies showing that VX-445 potentiates CFTR activity but is less potent than VX-770 [471–473], acute treatment of VX-445 modestly increased FIS in R352Q/F508del organoids by an AUC of 90.6 above baseline, lower than the AUC of 234.6 above baseline observed with VX-770 treatment (Figures 5.2B and 5.2C). As expected, acute treatment of VX-445 did not induce swelling in F508del/F508del organoids. These data indicate that R352Q has residual CFTR function and is responsive to VX-770 and G1837.

We next determined the effect of VX-809 monotherapy and cotherapy in R352Q/F508del organoids. In reference F508del/F508del and G551D/F508del organoids, VX-809 monotherapy at 0.128 μM Fsk concentration did not yield a significant change in FIS, with a net increase in AUC \leq 30.4 above baseline (Figure 5.2C, Table 5.E3*, Figure 5.E2*). In F508del/F508del organoids, VX-809 cotherapy with either VX-770 or G1837 caused a significant ($P < 0.01$) increase in FIS compared with potentiator monotherapy, with a net increase in AUC of 189.3 and 145.6, respectively (Figure 5.2C, Table 5.E3*). In contrast, no additional increase in FIS was observed in G551D/F508del organoids after

cotherapy with VX-809 and either potentiator (Figure 5.2C). Similar to the reference F508del/F508del organoids, no change in FIS was observed in the R352Q/F508del organoids at 0.128 μ M Fsk with VX-809 monotherapy. The net increase in AUC was 17.7 above baseline (Figures 5.2B and 5.2C, Table 5.E3*). VX-809 cotherapy in R352Q/F508del organoids followed the same trend as the F508del/F508del organoids but with a larger magnitude of response. VX-809 cotherapy with VX-770 or G1837 increased FIS beyond each respective monotherapy by 355.2 and 459.7 (Figures 5.2B and 5.2C).

The effect of VX-445 monotherapy as a corrector and VX-445+VX-661+VX-770 (Trikafta) triple combination therapy were tested in R352Q/F508del and F508del/F508del organoids. Similar to VX-809 monotherapy, VX-445 monotherapy at 0.128 μ M Fsk induced minimal FIS in R352Q/F508del and F508del/F508del organoids, increasing AUC by 101.3 and 15.2 above baseline, respectively (Figures 5.2A–C). In contrast, triple combination therapy VX-445+VX-661+VX-770 significantly induced FIS in R352Q/F508del and F508del/F508del organoids, increasing AUC by 2,913 and 3,173 above baseline, respectively. Because VX-445+VX-661+VX-770 primarily targets F508del folding and processing defects, the lower magnitude of CFTR correction in R352Q/F508del organoids than in F508del/F508del organoids supports our hypothesis that R352Q is unlikely to have folding and processing defects and that the functional correction observed is likely attributed to the F508del mutation.

5.2.3 Correlation of CFTR Modulator Response in hNECs with Participant-matched Intestinal Organoid Swelling Assay Outcomes

We compared the CFTR modulator response as determined by ion transport in hNECs and FIS in intestinal organoids. To determine the amount of modulator-stimulated restoration above the baseline CFTR activity, we subtracted the Δ Isc/AUC values with Fsk alone from those with Fsk plus modulators. A linear positive correlation was observed ($r = 0.75$; $P < 0.0001$) (Figure 5.3A). Participant-to-participant variability was evident (Figure 5.3B, Figure 5.E3*). Among the modulators, an inverse correlation was observed in response to VX-809 monotherapy between the two cell models ($r = -0.74$) (Figure 5.3C). Although VX-809 monotherapy significantly increased baseline currents in hNECs, no increase in FIS was detected in the matched organoids (Figure 5.E3*). When comparing CFTRinh-172-inhibited currents in hNECs against FIS in intestinal organoids, a similar trend of linear correlation was observed ($r = -0.52$; $P < 0.01$) (Figures 5.3A–3C).

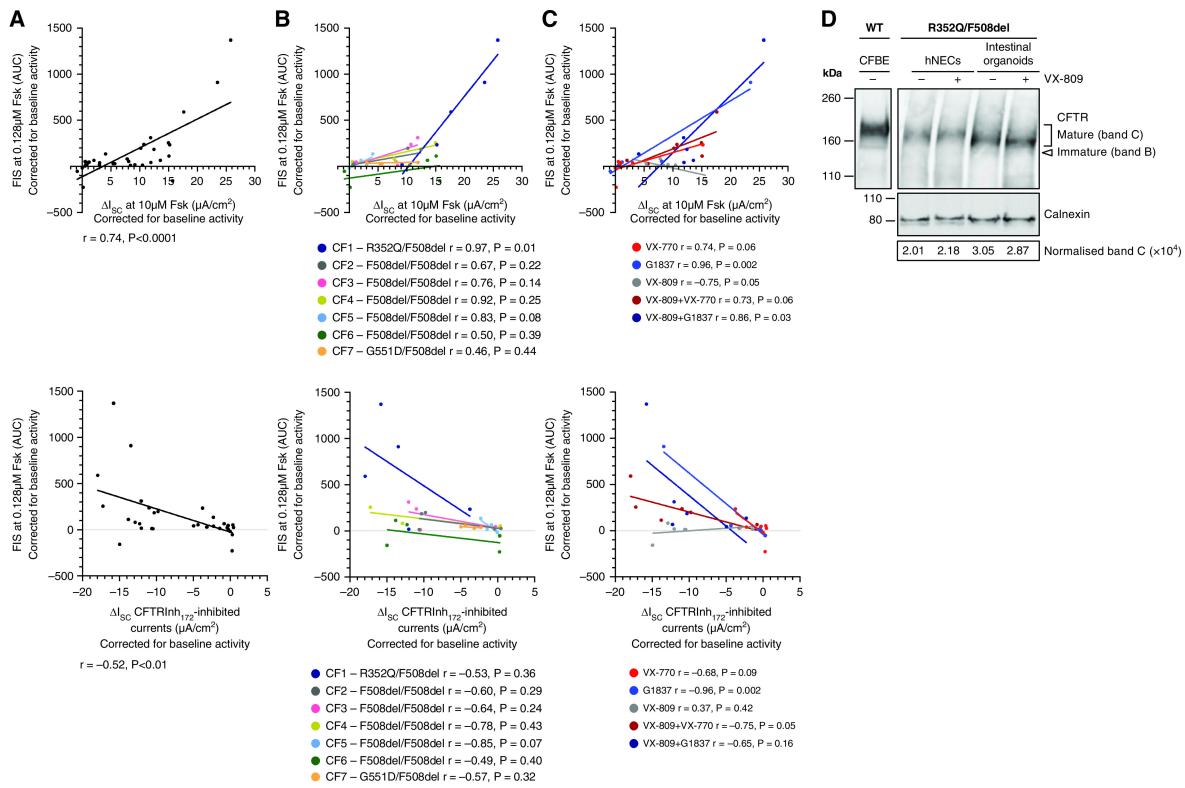


Figure 5.3: Correlation between individual in vitro treatment-modulated CFTR function in participant-matched differentiated hNECs and intestinal organoids.

Modulator-enhanced CFTR response (top) and CFTRinh-172-inhibited response (bottom), corrected for the baseline CFTR activity, in hNECs (ΔI_{SC}) compared with organoids (FIS) from each participant shown (A) as a collective, (B) by participant, and (C) by modulator therapy. The lines through data points show simple linear regression analysis. Correlation (r) values were calculated using Pearson linear correlation. (D) Western blot in lysates from R352Q/F508del hNECs and intestinal organoids following 48-hour treatment with DMSO (vehicle) or VX-809. Cell lysates of WT-CFTR human CF bronchial epithelial (CFBE41o-) cell line were used as a control for CFTR bands B and C protein size. Densitometry was calculated by measuring the level of mature mutant CFTR (band C) normalized to the calnexin loading control. Band C represents the mature, complex-glycosylated CFTR. Band B represents the immature, core-glycosylated CFTR.

5.2.4 Maturation of R352Q-CFTR in Nasal Epithelial Cells and Intestinal Organoids

We next compared CFTR maturation in R352Q/F508del hNECs and intestinal organoids, with and without VX-809 treatment, to understand the distinct response of the two cell models to VX-809 monotherapy (Figure 5.3D). Robust expression of mature, complex-glycosylated CFTR at 160 kD (band C) was detected in both untreated and VX-809-treated hNECs and intestinal organoids. Immature, core-glycosylated CFTR at 130 kD (band B) was not present. This validates our functional data by showing the presence of mature CFTR channels on the cell surface in R352Q/F508del hNECs and intestinal organoids, which can therefore elicit residual CFTR function and also re-

spond to potentiator (acts on protein at the cell surface). Mature CFTR in intestinal organoids migrated slightly farther (lower molecular weight) than the hNECs, likely because of different processes of CFTR glycosylation in different tissues, which has been reported previously [515]. Treatment with VX-809 did not further increase the levels of mature CFTR relative to the untreated control in both hNECs (densitometry/ $\times 10^4$ of 2.2 vs. 2.0) and intestinal organoids (densitometry/ $\times 10^4$ of 2.9 vs. 3.1). This is consistent with the nonsignificant or lack of functional rescue with VX-809 monotherapy in R352Q/F508del hNECs and intestinal organoids, respectively. It is noted that the intestinal organoids expressed a higher abundance of mature CFTR protein relative to the hNECs (1.5 times), consistent with known CFTR enrichment in the human gallbladder, intestine, and pancreas [516, 517].

5.2.5 Molecular Dynamics Simulations and Free Energy Calculations Reveal the Dysfunction of the R352Q Mutant

We next characterized the structural defect of R352Q-CFTR with molecular dynamics (MD) simulations, using an extended model of the phosphorylated, ATP-bound human CFTR [309]. In line with the previously proposed role for R352 in forming a salt bridge [518], it was observed that mutation of the positively charged arginine (R) to neutral glutamine (Q) disrupted the R352-D993 salt bridge in the midsection of the CFTR channel pore (Figure 5.4A, Figure 5.E4). This salt bridge is a key structural element in CFTR, and its disruption has been shown to impact conductance and gating [518]. However, there was no discernible disruption to the R352Q-CFTR channel pore architecture, because the root mean squared deviations of membrane-spanning helices of the R352Q-CFTR were as stable as the WT in microsecond-long MD simulations (Figure 5.4B). This suggests that the R352Q-CFTR open conformation can remain stable on the submicrosecond time scale necessary for chloride conduction [519]. If gating was impacted by the R352Q mutation, longer simulations might be able to decipher any disrupted pore architecture.

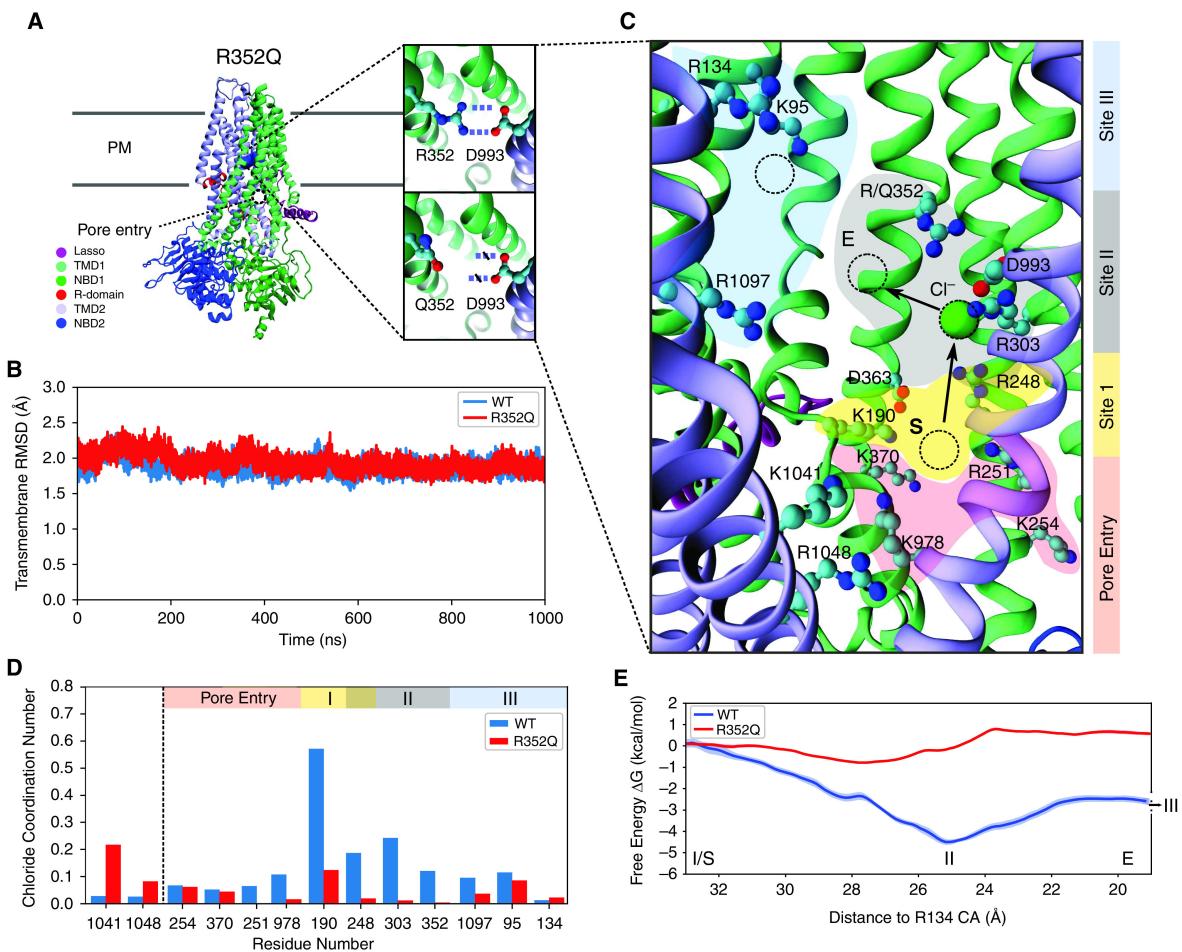


Figure 5.4: Simulations provide a mechanistic understanding of R352Q-CFTR dysfunction

(A) Atomistic structure of the CFTR channel pore based on cryo-EM (6MSM), with the configuration of the R352-D993 salt bridge and the disrupted R352Q highlighted in insets. Mutation of the charged R352 side chain (right top) to the neutral Q352 side chain (right bottom) disrupts the R352-D993 salt bridge present in WT-CFTR. (B) The root-mean-squared deviations (RMSDs) of the transmembrane domains of WT-CFTR (blue) and R352Q-CFTR (red) channel pore, computed using the 6MSM coordinates as reference. (C) A visualization of charged residues that line the CFTR channel pore. The path of chloride ion conduction from pore entry (pink) to site I (yellow), site II (gray), and site III (blue) is indicated. The computed free energy path in (E) is defined with reference to the chloride ion's distance to the alpha-carbon of R134 in site III, with the start and end points delineated by S and E, respectively. Charged residues lining the channel pore are shown as ball and sticks. An example chloride ion is represented with a green ball in site II. (D) The histogram shows the chloride ion occupancy in WT-CFTR (blue) and R352Q-CFTR (red). The number of chloride ions within 4 Å of the sidechain nitrogen atoms of the residue in each frame of the trajectory was counted and divided by the total number of frames. (E) Free energy profile of a chloride ion in the WT-CFTR (blue) and R352Q-CFTR (red) channel pore. The change in free energy relative to site I is plotted versus the chloride ions distance to the alpha-carbon of R134 (site III). NBD, nucleotide-binding domain; PM, plasma

membrane; R-domain, regulatory domain; TMD, transmembrane domain.

Because the R352 residue lines the CFTR channel pore, the R352Q mutation can be expected to affect the conduction of chloride ions [297]. This was assessed by determining the path of chloride conduction through the channel and comparing the occupancy of residues along the conduction pathway in WT-CFTR and R352Q-CFTR (Figures 5.4C and 5.4D, Video 5.E2*). In the WT system, a chloride ion spontaneously enters the pore intracellular gate [295] and diffuses toward K190 and K370 (site I) (Figure 5.4C). It remains in this site for 10 nanoseconds and then spontaneously moves up to the vicinity of R352 and R303 (site II). R248, which is positioned between sites I and II, coordinates the chloride ion movement between these sites. The chloride ion then moves to the vicinity of R134 and K95 (site III), which coordinate chloride diffusion into the outermost section of CFTR. Chloride conduction through the CFTR channel pore does not follow a straight path from pore entry to site III but rather follows a longer, indirect route. This is in contrast to the straight path that ions follow through voltage-gated potassium and sodium channels, which have tight pores that run perpendicular to the membrane [68].

In R352Q-CFTR, chloride ions can enter the channel pore but not as readily as in the WT system. They also generally do not progress as far into the pore as for WT-CFTR. Although multiple chloride ions occupied the WT-CFTR channel pore concurrently, only one chloride ion at a time was observed in the R352Q-CFTR channel pore. The highest occupancy in R352Q-CFTR was found for K1041. The chloride ions that enter the WT-CFTR channel pore stayed mainly in sites I and II, with occupancies of these two sites drastically reduced in R352Q-CFTR compared with WT-CFTR. This reduction in occupancy observed in R352Q-CFTR is due to neutralization of the positively charged R352 side chain, which reduces the affinity of chloride ions to site II. Site III was inadequately sampled to draw conclusions (Figure 4D). Intriguingly, some chloride ions appeared to penetrate site III in both WT- and R352Q-CFTR, which could explain the mild conductance defect observed for R352Q. In addition, in simulations of WT-CFTR, it was observed that D993 alternated contact between R303 and R352 (Figures 5.E4A and 5.E4B). However, in R352Q-CFTR, D993 forms a more stable salt bridge with R303 (Figure 5.E4C), which curtails its ability to coordinate a chloride ion.

The impact of the lower occupancy of sites I and II on chloride conductance was quantified further by comparing the respective free energy profiles obtained from umbrella sampling MD simulations (Figure 5.4E). To generate the initial path for the umbrella sampling simulations, one chloride ion was constrained at site I and then steered via site II toward R134 in site III, using a moving restraint. The free energy profile calculations followed the protocol described in the Supplemental Materials and Methods of the data supplement. In WT-CFTR, calculation of the free energy profile yielded an energy trough of -4 kcal/mol at site II relative to site I (Figure 4E). This indicates that chloride ions will be attracted to this site. In contrast, in R352Q-CFTR, the free energy profile exhibited an energy barrier of +1 kcal/mol at site II, which will likely hinder chloride conduction. Therefore, in WT-CFTR, a chloride ion in site I will tend to move to site II, whereas such a motion will be suppressed in R352Q-CFTR, resulting in reduced chloride conductance.

5.3 Discussion

This study adds to previously published findings by other CF laboratories by validating the reproducibility of cell models created from participants with CF for assessment of CFTR activity via in vitro CFTR functional assays [459, 467]. To characterize the rare R352Q-CFTR mutation, we first created matched hNECs and intestinal organoids from participants with CFTR mutations that have known functional defects. These acted as references in the functional assays, and because these reference mutations are common to other published platforms [459, 467], this facilitated comparison of data from our functional platform. Our reference hNECs and organoids both showed similar patterns of response to modulators when compared with previous studies, although the magnitude of response is lower [429, 459, 467, 511]. Variabilities in the magnitude of CFTR response in cell models created from different laboratories are evident (Figure 5.E5*). This could be attributed to several reasons. First, CFTR functional response to modulators in patient cell models was shown to be significantly modified by minor alterations to the culture media [520, 521]. Patient-to-patient variability in CFTR response may also explain this observation [511, 522, 523], which could be further skewed if a relatively small number of participants are studied. The variabilities in data between laboratories highlight the necessity to create reference cell models at every test center for comparison and characterization of rare CFTR genotypes.

To date, R352Q-CFTR functional characterization has been performed only in heterologous expression Chinese hamster ovary, Fischer rat thyroid and CFBE41o- cells [480, 495, 513]. These studies demonstrate varying levels of baseline activity ranging from 3% to 20% of WT-CFTR. However, in vivo CFTR function measured using nasal potential difference has shown that an individual with the R352Q/G1244E-CFTR genotype had CFTR activity just below that of WT individuals [511]. The G1244E mutation is a CFTR gating mutation with severe defect [86], suggesting that R352Q confers high levels of CFTR activity. To better characterize the R352Q mutation, we assessed CFTR function in hNECs and organoids derived from a R352Q/F508del-CFTR participant. Of the 100 individuals bearing the R352Q mutation in the CFTR2 database, 99 are heterozygous with F508del or another CFTR mutation genotype [86]. Therefore, our study participant is an appropriate representative of the R352Q-CFTR population.

The CFTR activity of a mutation is commonly expressed as a percentage of WT-CFTR activity to measure the severity of defect and responsiveness to modulator [495, 511]. However, we noted limitation in its use because WT-CFTR activity is somewhat underestimated in hNECs, resulting in overestimation of mutant CFTR activity. Baseline CFTR activity in R352Q/F508del hNECs was 70% of WT-CFTR activity. This was potentiated to well above WT-CFTR activity by VX-770 and G1837. Underestimation of the WT-CFTR activity was also shown by Pranke and colleagues [511], who showed that VX-770-potentiated CFTR current for R352Q/G1244E hNECs was ~160% of the WT-CFTR activity [511]. In our study, we found that the WT hNECs had significantly higher resting Isc than the CF hNECs (WT, $27.5 \pm 1.5 \mu\text{A}/\text{cm}^2$ vs. CF, $1.1 \pm 0.7 \mu\text{A}/\text{cm}^2$; $P < 0.0001$; data not shown), suggesting that there was endogenous CFTR activity at resting state even without the presence of Fsk. This may have limited further chloride transport when the cells were stimulated with Fsk, underestimating the

WT-CFTR activity. The presence of endogenous CFTR activity is also evident in WT intestinal organoids, which were reported to have preswollen morphology at resting state in the absence of Fsk [467]. Notably, a recent study reported that patient-derived cell models are effective tools to identify modulator-responsive patients but may not accurately predict the magnitude of clinical benefit [524]. It is likely that although the percentage of WT-CFTR activity is useful as a guide of reference for CFTR functional rescue, *in vitro* and *in vivo* CFTR function do not correlate in a linear fashion.

Nonetheless, the high baseline I_{sc} ($14.8 \mu\text{A}/\text{cm}^2$) and FIS (AUC, 196.3) of R352Q indicates a discernible amount of functional CFTR is present at the cell surface. In support of this, VX-809 monotherapy did not further improve CFTR activity in either R352Q/F508del hNECs or organoids. This is consistent with findings that VX-809 treatment did not further improve VX-770–potentiated CFTR activity in R352Q/G1244E hNECs [511]. Western blot analysis in Fischer rat thyroid and CF bronchial epithelial cells showed that 80–98% of R352Q-CFTR is present as mature CFTR localized to the cell surface, with little evidence of a processing or turnover defect [480, 495]. For this reason, because the mechanism of action of VX-809 is correction of a folding defect in CFTR that disrupts its trafficking to the cell surface, it is expected to have little, if any, restorative effect on R352Q-CFTR [475]. Our functional data confirm previous reports that the R352Q mutation is unlikely to cause folding defects and suggests that patients with R352Q-CFTR are unlikely to benefit from corrector modulator monotherapy.

Single-channel and whole-cell patch-clamp studies have demonstrated that the positively charged arginine (R352) creates an electropositive potential that promotes anion movement through the channel pore [513]. In addition, R352 forms a salt bridge with the negatively charged D993 residue (TM9), and this interaction is critical to keep the CFTR channel pore in a stable open-pore state [525]. Our MD simulations performed on the extended structure of human CFTR (Protein Data Bank accession no. 6MSM) also showed that R352Q disrupts the R352-D993 salt bridge in the channel pore; however, the gross architecture of the channel pore was unchanged. We showed that the R352 residue is not essential for chloride ion entry into the channel pore but is necessary to maintain chloride ion occupancy at a binding site within the channel pore conductance pathway (site II) for normal chloride ion conductivity. This is consistent with previous studies demonstrating that R352Q caused frequent transitioning of the channel pore to subconductance states (~ 40 –70%) and occasional full conductance channel pore opening [518, 526], causing a mild conductance defect. In the same studies, no apparent reduction in channel pore open probability (P_o) was deduced or reported from the single-channel patch-clamp recordings, suggesting a gating defect was unlikely. This is consistent with our MD simulations that predicted the R352Q open conformer can remain stable during chloride ion conduction.

The exact mechanism of action of CFTR potentiators VX-770 and G1837 is not fully understood. There is evidence that VX-770 rescues gating defects by enhancing P_o and stabilizing the open-pore conformation [527]. VX-770 was initially approved for CFTR gating defects [397]. In some countries, approval has since been extended to conductance defects, based at least in part on *in vitro* functional assays [396, 495]. In principle, potentiators can enhance CFTR activity by increasing the opening time of

channels (P_o) and/or increasing channel conductance. It is plausible that the significant responsiveness of R352Q/F508del to potentiators was due to an increase in channel pore opening time. This extra time compensates for the energy barrier created by the R352Q mutation, which hinders chloride ion conduction.

Because CF is a multiorgan disease [448]. Overall, our data indicated a positive correlation between CFTR functional response to modulators in matched hNECs and organoids across all CFTR mutations assayed. This is consistent with recent findings in several other rare CFTR mutations, such as G85E, P205S, and D614G [528]. In our study, among all modulators, VX-809 monotherapy was the only therapy to result in a negative correlation in response between the two cell models in all participants. The hNECs of all participants were responsive to VX-809 monotherapy but not their intestinal organoids. We propose two mechanisms that could result in this discrepancy. First, the VX-809 incubation time of 24 hours in organoids is shorter than the 48 hours in hNECs. This could impact the number of F508del-CFTR that are successfully trafficked from the endoplasmic reticulum to the cell surface because this correction is a time-dependent process [529]. In addition to having a trafficking and gating defect, F508del-CFTR is less stable at the cell surface [530]. Therefore, the shorter VX-809 incubation period in organoids may negatively impact CFTR stabilization at the cell surface.

Another mechanism could be the different concentration of Fsk used between the two assays. Both the I_{sc} and FIS assays use Fsk to stimulate protein kinase A-dependent CFTR activation [434, 520]. Protein kinase A increases channel P_o in a dose-dependent manner [531]. FIS at $0.128 \mu\text{M}$ Fsk correlates best with in vivo lung function (percent predicted FEV1) and allows characterization of responses in organoids without reaching ceiling threshold (saturation) in high residual function CFTR mutations [467]. In contrast, $10 \mu\text{M}$ Fsk has been the default concentration for I_{sc} measurements since early reports of this technique [532, 533]. However, it has recently been shown that $3 \mu\text{M}$ Fsk is sufficient to saturate the CFTR response in hNECs [534]. The high concentrations of Fsk in the I_{sc} assay are likely to function like a potentiator, extending CFTR channel opening time [531]. This might explain the significant response to VX-809 monotherapy we observed in hNECs but not in matching organoids that were treated with a nearly 100-fold lower concentration of Fsk ($0.128 \mu\text{M}$).

Strengths and Limitations

A strength of our study is the cross-validation of modulator response in matched hNECs and intestinal organoid cultures derived from the same participants. This minimizes the possibility of cell culture-dependent bias. We could not correlate in vitro CFTR response to participant in vivo clinical response, because individuals under 12 years of age with an R352Q-CFTR mutation in Australia do not currently have access to CFTR modulator therapy.

MD simulation studies are emerging as a method to understand the molecular mechanism by which rare mutations cause the CFTR protein to malfunction [535]. The combination approach of using in vitro experiments and MD simulations to characterize the R352Q-CFTR mutation paves a feasible and reliable pathway to personalized

therapies for participants with rare CFTR mutations (Figure 5). In our simulations, we demonstrated that alterations to the free energy profile in the channel pore introduced as a result of the R352Q mutation negatively impacted movement of the chloride ions to move past the midpoint of the CFTR channel pore (site II), diminishing the channel's conductance. In vitro experiments complement these simulations by demonstrating that the unique defect introduced by a mutation can be rescued by CFTR modulators [371, 373, 536]. In this study, we have demonstrated that the R352Q-CFTR defect is rescued by potentiaters. Functional characterization of CFTR in both lung and intestinal models may improve the specificity and sensitivity of predicting modulator test results and aid in translating patient-derived cell models to become a mainstream companion diagnostic test for people with CF.

5.4 Supplemental Information

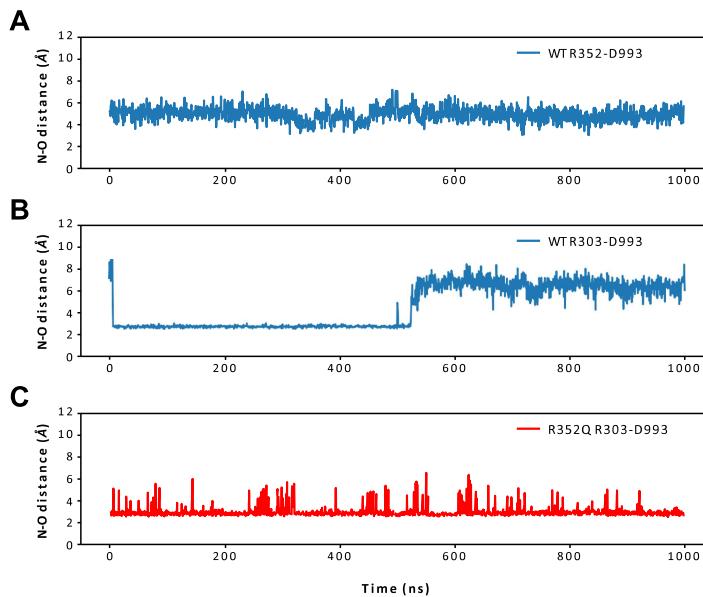


Figure 5.E4: The Stability of Key Salt Bridges within the CFTR Channel

N-O distance refers to the distance between the charged head groups of positive arginine (R303, R352) residues and the negative aspartate residue (D993). A) It is expected that a stable interaction between R352-D993 is an important feature of the fully open channel. Our data indicates that R352 remains in close proximity to D993 in the Wild Type channel. B) R303 makes a stable salt bridge with D993 for the first 550ns of the simulation. The presence of R352 allows D993 to either interact with R352 or R303. When D993 is in contact with R352, D993 is neutralized, leaving R303 free to coordinate a nearby chloride ion. C) In the R352Q mutant, the R303-D993 salt bridge remains throughout the simulation.

5.5 Funding Disclosure

Supported in part by an Australian National Health and Medical Research Council grant (NHMRC_APP1188987), a Rebecca L. Cooper Foundation project grant, a

Cystic Fibrosis Australia-The David Millar Giles Innovation Grant, the Sydney Children Hospital Network Foundation, and a Luminesce Alliance research grant. S.A.W. is supported by the Australian National Health and Medical Research Council. M.A.A. acknowledges support of a top-up scholarship from Cystic Fibrosis Australia. K.M.A. is supported by an Australian Government Research Training Program Scholarship. L.K.F. is supported by the Rotary Club of Sydney Cove-Sydney Children's Hospital Foundation and University of New South Wales postgraduate award scholarships.

5.6 Author Contributions

Conception and design: S.A.W. and A.J. Recruiting of participants: L.K.F., S.A.W., J.R.W., and C.Y.O. Collection of nasal brushings: L.K.F. and J.R.W. Human nasal epithelial cell culture and ion transport assay and analysis: N.T.A., K.M.A., and S.L.W. Immunofluorescence microscopy: S.L.W. Ciliary beat frequency (CBF) imaging: I.S. and S.L.W. CBF analysis script and design: E.P. and R.M.W. Collection of rectal biopsies: C.Y.O. Culturing the organoids: N.T.A., S.L.W., and S.A.W. Performed forskolin-induced swelling (FIS): I.S., N.T.A., K.M.A., and S.L.W. FIS assay scripts: M.J.C. FIS assay analysis: N.T.A., M.J.C., and S.L.W. Western blot analysis: S.L.W. Molecular dynamics: M.A.A., P.C., J.S., R.G., and S.K. Figure preparation: S.L.W., K.M.A., M.A.A., and S.A.W. Writing – original draft: S.L.W., M.A.A., and S.A.W. Review and editing: S.A.W., K.M.A., L.K.F., R.G., and S.K. with intellectual input from all other authors. Supervision: S.A.W. and S.K.

6

Unique S945L-CFTR defect Restored by CFTR Modulator Co-Therapy In Vitro, Which Correlates with In Vivo Biomarkers Post-Therapy

Eventually, you'll realise you're just two days from anywhere.

– -David Goodwin (personal communication)

Preamble on Formatting

The work in the following chapter is part of a publication in preparation for submission to *Frontiers in Pediatrics*. All experiments have been completed and the authors are in the process of writing the manuscript. The computational components of the manuscript have been finalised, and the results are presented here. On the other hand, because the *in vitro* and clinical components of the manuscript have not been finalised, so their results are not included explicitly in the manuscript. Instead they are described in brief and referred to as needed to give the *in silico* results context.

Abstract

Cystic Fibrosis (CF) results from loss of functions mutations to the CF Transmembrane Conductance Regulator (CFTR) gene. A CFTR mutation can produce numerous defects and drugs known as CFTR modulators have been developed to act directly on the protein and restore its function. These drugs have been a breakthrough in the treatment of CF. However, the response of patients to these medications is variable, and there is an urgent need for a personalised approach to their administration. The objective of this study was to elucidate the molecular details of pathogenesis of a rare mutation, S945L through molecular dynamics (MD) simulations and umbrella sampling. These

methods discovered a mode of pathogenesis unique to S945L-CFTR, which results in both gating and folding defects. Our simulations were performed in parallel to *in vitro* assays which were used to interrogate the efficacy of modulators in patient specific cellular models. This was followed by a clinical study, which tracked the outcomes of the patient carrying the S945L mutation upon receipt of the recommended medication. The results indicate that the mode of pathogenesis unique to the S945L mutation was responsive *in vivo* to current generation CFTR modulators. Hence, the presented study demonstrates how patient specific cellular assays can assess the efficacy of modulators in rare mutations and inform clinical practice.

6.1 Introduction

Cystic Fibrosis (CF) is a life-limiting recessive disease [537, 538] caused by mutations in the CF Transmembrane Conductance Regulator (CFTR) gene that result in absent or dysfunctional CFTR protein [539]. The CFTR protein is a chloride and bicarbonate channel [366] which maintains fluid homeostasis at the apical membrane of epithelial cells [443] in the lungs, pancreas, liver and intestine [448]. CFTR is a complex protein, composed of seven domains. Of primary significance to this study are the Transmembrane Domains (TMDs) and the Regulatory domain (R-domain). TMDs 1 and 2 are composed of six tight bundles of helices which run through the plasma membrane. The proper arrangement of these helices is critical to avoiding a folding defect of the protein [372]. The S945L mutation occurs in TMD2 on a helix numbered TM8. When CFTR opens, the TMDs twist in order to form a pore and allow ions to pass through. In the closed state the R-domain wedges between the TMDs, preventing them from twisting to form a pore. During gating, the R-domain is phosphorylated and dislodges from its wedged position and docks against TMD2 [297]. Regulating the movement of the R-domain has important implications for the gating cycle of the protein [298].

CF is a monogenic disease and over 2000 unique mutations have been identified in the CFTR gene, yielding a spectrum of CFTR structural defects [540]. Of these, 401 mutations have been found to be disease-causing [86], albeit with heterogeneous clinical phenotypes [541]. Whilst not curative, a novel class of targeted therapies (CFTR modulators) have been developed to restore CFTR protein function [461]. CFTR modulators have two modes of action: potentiators (ivacaftor, GLPG1837) which open the CFTR channel pore [416, 542–544], and correctors (lumacaftor, elexacaftor, tezacaftor) which assist in CFTR protein folding and trafficking to the cell surface [545, 546]. Clinically approved dual therapies, lumacaftor/ivacaftor and tezacaftor/ivacaftor + ivacaftor (Symdeko), as well as triple combination therapy elexacaftor/tezacaftor/ivacaftor + ivacaftor (ETI) improve CFTR function via both these mechanisms [547–549].

ETI is currently the most effective treatment for improving patient lung function, [486, 548] and provides treatment access to more than 90% of people with CF. However, the clinical response to ETI, as well as other modulators, has been reported to be heterogeneous, even amongst patients with the same CFTR mutation [547, 550]. *In vitro* CFTR modulator testing uses patient-derived cells models as a personalised drug screening platform to predict the individual's response to treatment and characterise

the functional defect caused by CFTR mutations. This form of *in vitro* analysis has been permitted by the US Food and Drug Administration (FDA) to inform approval of therapeutics for less common CFTR mutations.

The diversity of disease-causing mutations and the availability of high resolution CFTR protein structures have led to the ability to study the effects of rare mutations in detail. In this study we have used Molecular Dynamics (MD) combined with Umbrella Sampling (US) to study the stability and dynamics of the S945L-CFTR protein. Our results suggest this mutation causes both a folding defect and a gating defect, due to the destabilisation of a fragment of the R-domain.

The modelling work was motivated by studies of a patient carrying the S945L/G542X-CFTR genotype. As they are still under preparation, the *in vitro* and clinical results of this study are not presented, but will be described in brief.

The cell models derived from this patient were assessed for their response to various CFTR modulators. It was predicted through these pre-clinical assays that the patient would benefit from combination therapies such as Symdeko, which is a combination of tezacaftor (VX-661), a corrector and ivacaftor (VX-770), a potentiator. After a year on this medication, an increase in lung function was observed in the patient. These clinical results validated the hypothesis generated from the *in vitro* cell models, demonstrating the efficacy of a patient specific approach to the choice of CFTR modulators.

This is also consistent with our previous work, which has demonstrated the functional rescue of rare missense mutations by modulators . All observed rescue in the S945L/G542X patient must have been due to the modulation of the S945L-CFTR protein, as G542X is a class I nonsense mutation, which does not respond to modulators [551, 552]. S945L is a rare CF-causing CFTR mutation expressed on 167 of 142,036 CFTR alleles worldwide (0.00118% allelic frequency) and results in pancreatic insufficiency in 40 % of carriers 11. S945 is situated at the end of transmembrane helix 8 of the CFTR protein, a region known to contribute to ion selectivity [312] and bind potentiator class CFTR modulators [311].

S945L has approval for treatment with Symdeko in the US, Canada, Europe and Australia based on evidence of improved lung function in a Phase III clinical trial of 248 patients (13 patients with S945L) 12 years or older [548]. While the rescue efficacy of Trikafta has not been assessed for this mutation. Furthermore, although functional characterisation of S945L-CFTR has been performed via molecular studies in heterologous expression Chinese hamster ovary cells and described to interfere with protein folding, as well as gating function [553], this has not been validated in patient derived cell models. Molecular dynamics (MD) simulations of the CFTR protein have previously demonstrated the importance of hydrogen bonds on TM8 to the stability of the open state [336]. Due to its proximity to this region, we believed that the S945L molecular defect might be due to the disruption of interactions that stabilise wild-type (WT) CFTR. This motivated MD simulations and free energy calculations of the S945L-CFTR protein, to characterise its molecular defect.

In parallel with our computational modelling, the mutation was functionally characterised *in vitro* by assessing the response to CFTR modulators via a CFTR-mediated

chloride transport assay in S945L/G542X-CFTR patient-derived primary airway epithelial cell models. Various CFTR modulators with different mechanisms of action were tested *in vitro* to characterise S945L-CFTR, leading to the prediction that a combination therapy would give the patient the most benefit. *In vivo*, it was then found that the patient exhibited increased lung function after treatment with the combination therapy Symdeko. This demonstrates the success of a personalised approach to the choice of CFTR modulators.

6.2 Results

6.2.1 In silico characterisation of the Misfunction of S945L-CFTR

The S945 amino acid is located in TM8, near the membrane surface on the cytoplasmic side of CFTR (Figure 6.1A). Our simulations show that this amino acid plays a structural role by bridging TM8 with adjacent TM7 and R-domain helices. This is facilitated in WT-CFTR by hydrogen bonding to Y852 in the R-domain. Two configurations of this hydrogen bonding were observed in WT-CFTR simulations, indirect hydrogen bonding via a stable, mediating water molecule, or direct hydrogen bonding between S945 and Y852 (Figure 6.S1A). Both states are energetically favourable and keep the Y852 and S945 amino acids in proximity to one another (Figure 6.S1B). The interactions were stable throughout three replicate 2 μ s MD simulations of WT-CFTR, at both a physiological temperature (37°C; Figure 6.S2) and an elevated temperature (77°C; Figure 6.1B and Figure 6.S1A-B).

In S945L-CFTR, the mutant leucine amino acid (L) lacks the polar sidechain of the WT serine (S), so it cannot form hydrogen bonds (Figure 6.1A). The effects of this change were measured and compared in WT and S945L-CFTR. Compared to WT-CFTR, S945L-CFTR had increased conformational fluctuation in the R-domain elbow, as measured by root mean square deviation (RMSD, Figure 6.S3A-B). This suggested that the S945L mutation decreased the R-domain stability. The instability and loss of the hydrogen bond anchor allowed the R-domain Y852 sidechain in one S945L simulation to physically rotate away from TM8 (Figure 6.1B). Since the R-domain is critical for protein gating, these observed movements and destabilisation of the R-domain suggest that S945L-CFTR is likely to possess a Class III gating defect.

In S945L-CFTR, movement of the R-domain away from its folded conformation was observed in only one of three replicate simulations at 77°C (Figure 6.1B). While this movement was not observed in any of three replicates at physiological temperature (37°C, Figure S2). These results suggested that direct microsecond simulations were not long enough to reliably explore potential pathological conformational changes in S945L-CFTR. We believe the presence of the lipid bilayer near the S945L mutation introduced a kinetic barrier to the expected conformational changes. The slowed movements of the R-domain elbow motivated more investigations into the energetics of this motion by a more rigorous method known as umbrella sampling (US). Here we used a virtual spring to pull apart the alpha carbon atoms of amino acids 945 and 852. As the atoms were pulled away from each other, we measured the force exerted on the spring.

This let us calculate how likely the link between Y852 and S/L945 was to break in both WT and S945L-CFTR.

WT-CFTR displayed a free energy surface more stable than S945L-CFTR because of the presence of the S945-Y852 hydrogen bond (Figure 6.1C). The WT-CFTR free energy surface featured a deep local minimum when Y852 and S945 alpha carbons were separated by 9.5 Å. Distances greater than 12 Å were inhibited in WT-CFTR by high energies of 4.7 kcal/mol, which would not be easily accessible at physiological temperatures. Therefore, WT-CFTR was likely to occupy the deep minimum at 9.5 Å. This indicates the constraining effect of the hydrogen bond between S945 and Y852. The steepness of the local minimum was also implied by the smaller fluctuations observed during the direct simulations, as if bound by a tight spring (Figure 6.1B). These findings indicate that WT-CFTR will remain correctly folded because of its stable network of hydrogen bonds. In contrast, the S945L-CFTR free energy surface showed several shallow local minima between the near-native 9.0 Å and maximum distance of 13 Å, observed in Figure 6.1B. This free energy surface correlates with the larger fluctuations observed in the direct simulations, as if bound by a weaker spring than WT-CFTR (Figure 6.1B). The small, intervening free energy barriers, which had heights of 2.0 kcal/mol or less, could be overcome at physiological temperatures, causing the protein to misfold. This indicates that S945L-CFTR is likely to spontaneously transition into misfolded conformational states where the R-domain elbow, containing Y852, moves away from TM8. We believe these conformational transitions will give rise to the Class II folding defect observed *in vitro*.

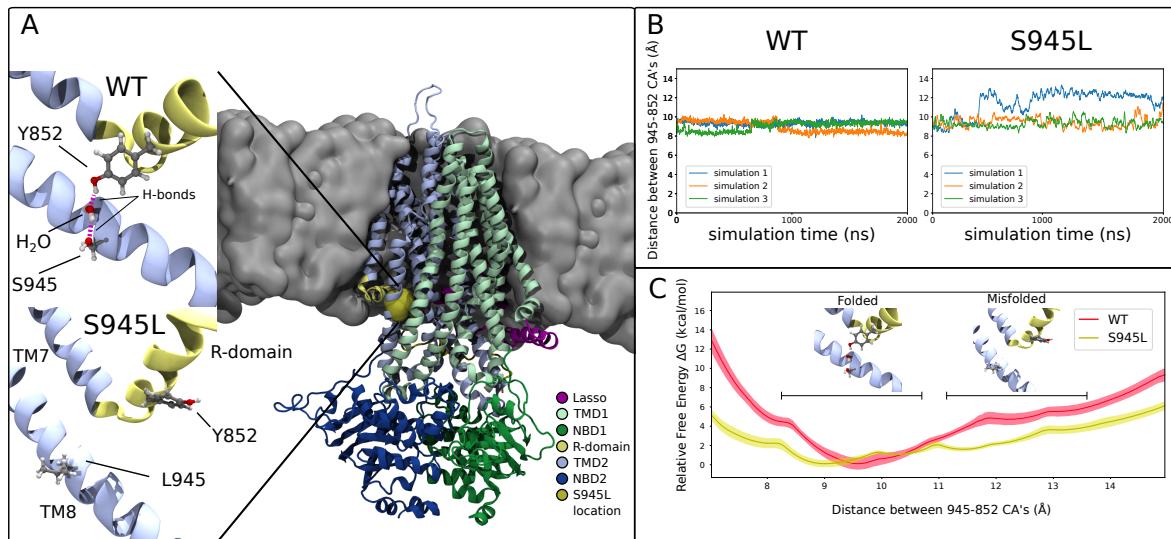


Figure 6.1: S945L

A) The location of the S945L mutation in the open structure of CFTR (PDB ID: 6MSM) is indicated with yellow spheres. It is placed between TM8 and the R-domain, the latter domain is critical for the gating of the channel. B) Data from unbiased simulations of WT and S945L-CFTR. A stable network of hydrogen bonds forms spontaneously in simulations 1, 2 and 3 of WT-CFTR. This interaction connects Y852 to S945. In the S945L mutation, the leucine side chain cannot support this hydrogen-bond network. In simulation 1 of S945L-CFTR, after a conformational transition, the amino acids remained at an average distance of 12.2 ± 0.7 Å. This is significantly further

than the average distance in the WT data: 9.1 ± 0.3 Å. No significant changes were observed in simulations 2 and 3 of S945L-CFTR. C) Results from umbrella sampling, a more computationally rigorous approach, confirm the presence of the conformational change in S945L-CFTR. Distances below roughly 10.5 Å correspond to a well folded state, while distances larger than 11 Å appear to correspond to a poorly folded state. In the mutant, there is a shallow free energy minimum at 11.4 Å with a difference in the energy landscape of 3.2 kcal/mol compared to WT-CFTR at the same location. The presence of this shallow free energy minimum indicates that S945L-CFTR is more likely to misfold than WT-CFTR.

6.3 Discussion

We have described in detail the structural and functional CFTR defects caused by the S945L mutation. In silico protein modelling via MD simulations was used to determine the molecular cause of both a folding and a gating defect in S945L-CFTR. Airway epithelial cell models derived from a participant with S945L/G542X-CFTR were also created and their response to CFTR modulators was tested via functional studies, to characterise their functional defect. The rescue of mutant CFTR was found to then translate to success in the clinic, as the patient carrying this mutation had an increase in lung function following the administration of the recommended therapy.

In our MD simulations of WT-CFTR, a stable network of hydrogen bonds linking Y852 in the R-domain to S945 on TM8 was identified. This interaction was not visible in the cryo-EM structure of the channel, where S945 faced away from Y852. It is often not possible to resolve such water mediated interactions in cryo-EM structures. This is due to the difficulties in achieving a high enough resolution with experimental structural biology techniques and the dynamic nature of the proteins themselves. The elucidation of this interaction in this study demonstrates the atomistic insight which computational methods can provide into the function of proteins. In simulations of S945L-CFTR, this interaction was disrupted, and our results indicate that the R-domain may move away from TM8 (Figure 6.1B). Due to the importance of the R-domain in gating, we expect that a change to its position will give rise to a gating defect. This is a similar mode of pathogenesis to the one observed in the I37R mutation, although it is interesting that it occurs in a different domain in CFTR. Simultaneously, the disruption of this link between two different domains is likely to give rise to a folding defect, as the elbow region will be less likely to fold correctly. Hence, our MD simulations have revealed a unique mode of pathogenesis for the S945L mutation. It is unlikely that other mutations would disrupt this hydrogen bond in the same way. This is another example in a set of MD simulations of the CFTR protein which suggest that CFTR modulators can elicit a response in many missense mutations, each with unique molecular phenotypes [304, 309, 371, 373].

The results from clinical and *in vitro* experiments are not shown explicitly, but described here to complement and give context to our MD studies. When tested under basal conditions, the S945L/G542X-CFTR ALI cell models exhibited high residual activity. This indicated the presence of some CFTR protein at the cell surface, suggesting that the S945L mutation produces a mild CFTR folding defect. Monotherapies using

either a single potentiator or a single corrector were not able to restore CFTR function to levels which would indicate clinical benefit. However, combination therapies Symdeko and Trikafta were able to restore function in these cell models. This was expected as Trikafta has recently been evaluated in rare class II CFTR folding mutations and variable genotype-specific responses were found [471, 480]. Like the prototypical class II folding mutation $\Delta F508$, S945L has been reported to cause multiple defects including protein maturation, channel gating and conductance [553].

We correlated the *in vitro* drug response of the S945L/G542X-CFTR patient-derived cell models to the patient's *in vivo* change in lung function (FEV1) following treatment with the Symdeko. Once the patient was administered Symdeko for 12 months, their lung function improved. These results support the use of personalised primary airway cell models as a valuable tool to study modulator efficacy and predict patient drug response for patients who carry rare mutations which are difficult to study through clinical trials.

The *in vitro* and clinical studies demonstrate the efficacy of pre-clinical cellular models, while the MD studies demonstrate the rescue of yet another unique molecular defect by CFTR modulators.

In vitro, S945L-CFTR was not found to respond to treatment by a potentiator monotherapy. This is consistent with previous findings. The rescue of S945L-CFTR by correctors has not previously been assessed, but again significant restoration was not observed with this kind of monotherapy. By contrast, combination therapies Trikafta and Symdeko were found to restore function in cellular models. As such, we characterise S945L to incur both CFTR folding and gating defects, which is in support of previous molecular studies performed in heterologous expression CHO cells [553], as well as the findings of our MD simulations.

Since the response of different CFTR mutations [471] to modulators is heterogeneous [547, 550], there is a strong rationale for modulator screening of CFTR in patient-specific cell models. Prediction of CFTR drug response in patient-specific cell models may, in the future, lead to extended approval of CFTR modulators for rare CFTR mutations currently without treatment access. Together with other integrative studies using both MD simulations and *in vitro* functional studies to characterise CFTR mutations and predict drug response in a personalised manner [304, 309, 371, 373], there is a growing body of evidence that a large number of unique molecular defects may be rescued by CFTR modulators. Further development of patient-specific cell models as a predictive tool could lead to the optimal choice of CFTR modulators to treat each patient.

6.4 Supplementary Information

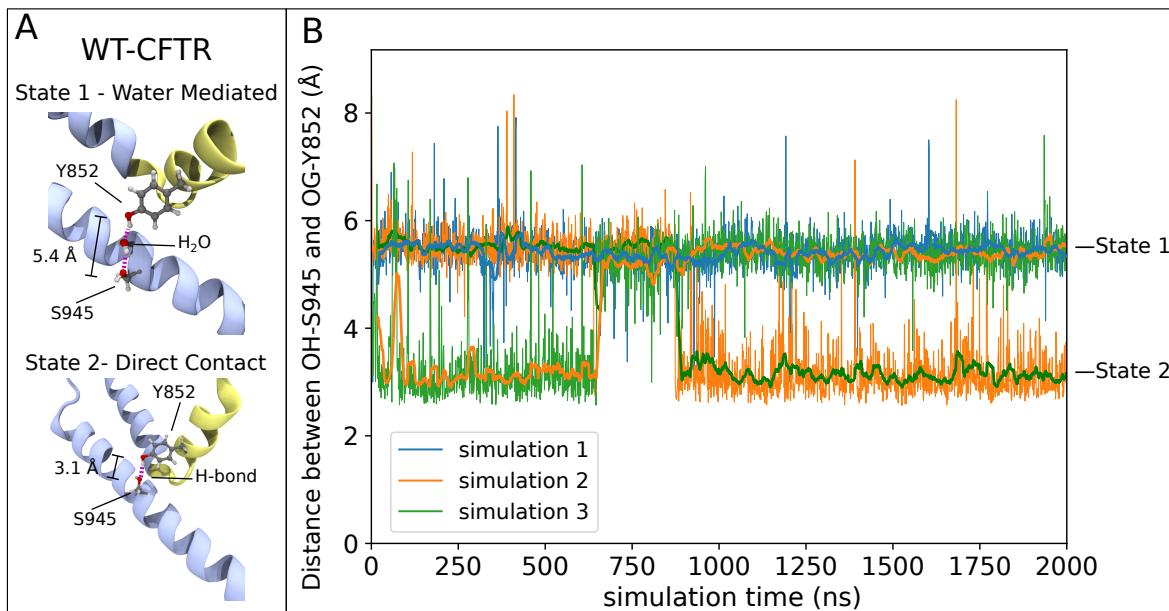


Figure 6.S1: Detailed View of the Hydrogen Bond Network in WT-CFTR

A. In our simulations we observe two possible states. State 1 occurs when a water molecule stably mediates the interaction between the sidechains of S945 and Y952. State 2 occurs when they are in direct contact with each other. B. The distance between the oxygen atoms in the polar side chains of S945 and Y852 throughout the WT simulations. In simulation 1 the system remains in state 1 throughout. Simulation 2 begins in state 1 but transitions to state 2 at 900 ns. Simulation 3 begins in state 2 but transitions to state 1 at 700 ns. Data is graphed with a 20 ns moving average to more clearly delineate the difference between the two states. This data indicates that the two states each contribute to the stable hydrogen bond network between Y852 and S945.

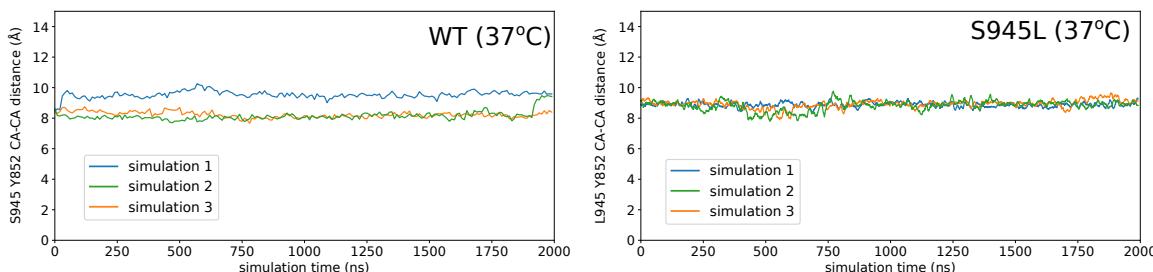


Figure 6.S2: Simulations of Both WT-CFTR and S945L-CFTR at Physiological Temperature (37°C)

In the WT data, we again see a stable interaction between Y852 and S945. The simulations 2 and 3 remain in state 2 throughout the whole simulation, with simulation 2 changing to state 1 at roughly 1900 ns. Simulation 1 quickly transitions to state 1 and remains there throughout the simulation. In S945L, the presence of the lipid bilayer has slowed down the kind of conformational changes we would expect from a mutation like S945L. Unfortunately, no pathogenic motions were visible at physiological

temperatures on the timescales accessible with unbiased MD simulations. The Y852 and L945 amino acids remained at a distance somewhere between state 1 and state 2. These results motivated a more rigorous study with umbrella sampling in order to understand the molecular details of the defect which the S945L mutation causes to CFTR.

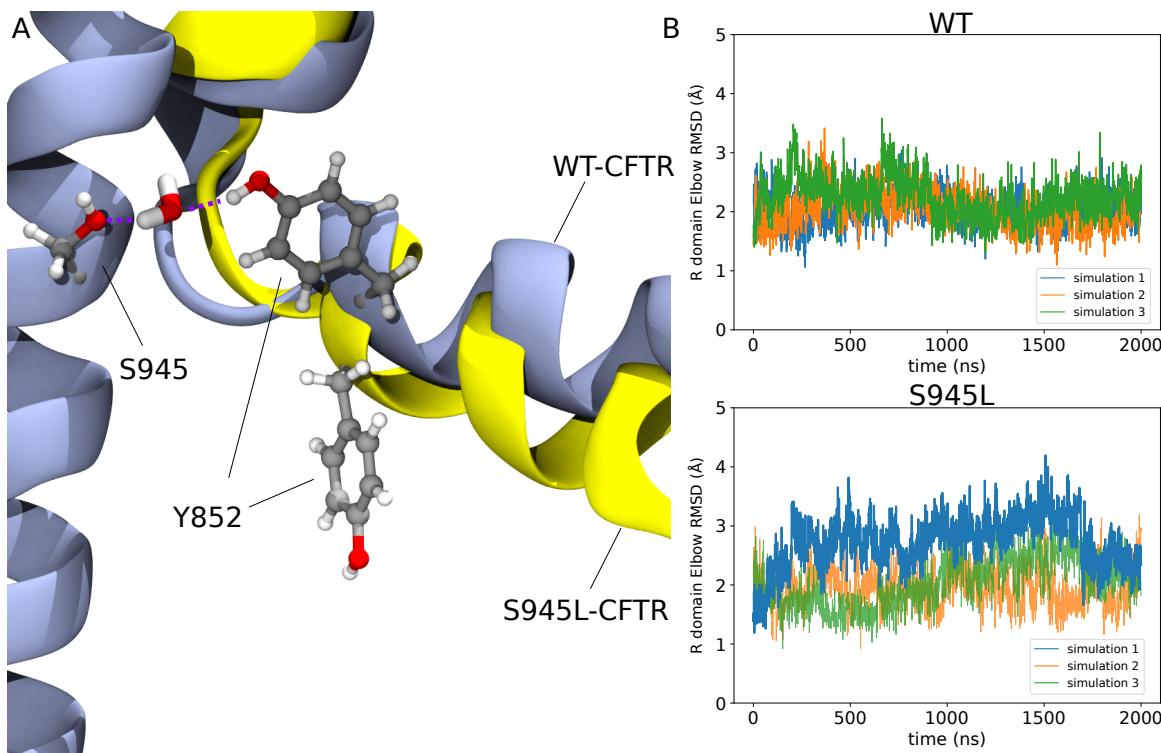


Figure 6.S3: The Local Conformational Change of the Elbow Between TM8 and the R-domain.

(WT-purple, S945L-yellow) A. The Y852 side chain sits in the elbow of the R-domain (amino acids 845-886) and when it is perturbed, a shift in the elbow is observed. This conformational change is proposed to influence the gating cycle of CFTR. B. The Root Mean Squared Deviation (RMSD) values of the elbow region, calculated with reference to the alpha Carbon Atoms of the 6MSM structure. Simulation 1 was the replicate with the greatest perturbation compared to the WT structure in the unbiased simulations (Figure 6.1B). The data for simulation 1 of the S945L mutation indicates a destabilisation of this elbow region. This is proposed to cause defects with both the gating cycle of the protein and the stability of the correctly folded state.

6.5 Method Details

6.5.1 Molecular Dynamics

2.1 Molecular Dynamics The CFTR protein model used for molecular dynamics simulations was based on the ATP-bound human CFTR cryogenic electron microscopy (cryo-EM) structure (PDB ID: 6MSM) [297]. The construction of the CFTR protein system, including the modelled part of the R domain, has been described previously [309]. In brief, the CFTR model was embedded in a simplified cell membrane model

containing pure 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine (POPC) lipids then solvated with 150 mM KCl on both the cytoplasmic and periplasmic sides.

In this study, the software Gromacs v2021.1 [157] was used for all MD simulations. Molecular mechanics parameters were supplied via the CHARMM36m [123] force field with the following modifications for faster simulation performance: virtual site topologies [167] for all molecules while using MkVsite36 to generate topologies for the ATP molecules [554], as well as parameter adjustments specific to lipids [555]. The initial positions of all molecules were equilibrated for production runs in two phases: first via energy minimisation using a steepest descent algorithm, then followed by a 6 ns simulation where position restraints were imposed on all non-hydrogen atoms and then gradually relaxed.

A series of 2 μ s Production MD simulations were carried out with three replicates for each of WT and S945L-CFTR. To decrease computational expense, elevated temperatures of 350 K were used (77°C) to accelerate conformational transitions. The use of elevated temperatures to study defects in mutant CFTR was validated previously [309]. All measurements of Root Mean Squared Deviation (RMSD) were calculated using the positions of alpha carbons in the 6MSM PDB structure 8 as a reference.

6.5.2 Umbrella Sampling Protocols

The free energy surfaces were constructed by measuring the potential of mean force (PMF) profile between the amino acids Y852 and S/L945. The distance between the alpha carbon atoms in these amino acids was used as the collective variable for the calculation. This distance was varied from 6–16 Å with a restraint of 10 kcal/mol/Å². Umbrella sampling simulations were performed in 10 windows spaced 1 Å apart. After 100 ns of equilibration for each window, 100 ns of data were collected. The resultant PMF profiles were calculated from the average of the 5 profiles determined from 5 independent samples of 20 ns each. Uncertainties for this result were calculated as twice the standard error of the mean for these 5 samples. The calculations have clearly converged, demonstrated by uncertainties of less than 1 kcal/mol in the free energy surfaces of both the WT and S945L-CFTR. Plumed 2.6 [241, 556, 557] was used to perform the umbrella sampling, and WHAM [558] was used to calculate the PMF profiles.

Resolving a Conducting Conformation of CFTR Using Free Energy Calculations

You wanna fight?

– Zachary Picker. Layperson. (personal communication)

Abstract

The malfunction of the CFTR gene causes Cystic Fibrosis. The protein must form an ion channel to conduct both chloride and bicarbonate in order to avoid disease. The recent development of small molecules known as potentiators has demonstrated the clinical efficacy of modulating the mutant forms of the protein such that it occupies an conducting conformation to relieve disease symptoms. Hence, determining the precise mechanism behind ion conduction through CFTR has important implications for ongoing drug discovery efforts.

Existing atomic structures of CFTR leave unresolved questions behind this conduction mechanism, since these structures exhibit constrictions smaller than the ions themselves. This indicates that there must be some conformational changes to this structure for ions to pass through CFTR. Here we present innovative simulation techniques which combine unbiased molecular dynamics, principal component analysis, OPES-Metadynamics and umbrella sampling to study the conduction of both chloride and bicarbonate through the selectivity filter of human CFTR. We also used this model to study the R334W disease causing mutation and confirm it as a conductance defect.

Our proposed open structure is in better agreement with both theoretical predictions and experimental measurements of the pore architecture of CFTR. We also propose electrophysiology experiments to experimentally test whether this conformation is correct. The discovery of an open structure of CFTR has important implications for ongoing drug discovery efforts.

Even the limited success of this study demonstrates that computational power and protein forcefields are now sufficiently developed to take advantage of the large database of protein structures collected through the cryo-EM “resolution revolution” and AI based protein structure predictions. In this era of structural abundance, MD based methods can now be used to elucidate even more of the protein conformational landscape. This signals an exciting future for computational structural biology and biophysics.

7.1 Introduction

Ion channels are crucial to many cellular functions. When they malfunction, they may cause pain or disease [74, 559]. Hence, they have been pursued as targets for drugs, and the molecular basis for their conduction of ions has been studied closely for some time [32, 44, 46].

One of the most well-known “channelopathies” is caused by loss-of-function mutations in an anion channel, the Cystic Fibrosis Transmembrane conductance Regulator (CFTR). This malfunction leads to Cystic Fibrosis (CF) [294, 443]. This is a fatal genetic condition which afflicts an estimated 162000 people worldwide, with a significant number of people unable to access treatment [265].

The function of mutant CFTR can be restored by molecules known as CFTR modulators. A kind of CFTR modulator known as potentiators interact directly with CFTR to cause it to favor its open, conducting conformation [311]. In currently available cryo-EM structures of CFTR, there is a constriction along the ion conduction pathway which is too small to allow the ions to pass [297] (Figure 7.1A). Hence, the full conduction pathway of ions through CFTR is not yet understood. To understand how these drugs work and to improve on them, a molecular-level understanding of the ion conduction through CFTR is critical.

The ion permeation pathway of CFTR can be split into two sections. The inner vestibule and the selectivity filter. We have previously studied the motion of chloride through the inner vestibule in chapter 5 [304], and in this chapter we will study conduction through the selectivity filter.

The selectivity filter can also be divided into two parts. A lower “conductivity selectivity filter” and a higher “permeability filter” (Figure 7.1B). The different sections in the selectivity filter were characterised by measuring the permeation profiles of different anions under different mutations and the rate of modification by reagents [325, 560].

The lower “conductivity selectivity filter” appears to bind anions as they move through the channel while the “permeability selectivity filter” is narrow and hydrophobic, so anions are partially dehydrated as they move through. This latter section is thought to be the origin of the lyotropic sequence of anion selectivity through the channel (Figure 7.1B) [327].

The motion of chloride through the selectivity filter has been studied in two other *in silico* studies to date [330, 333]. In the first computational study, performed on zebrafish CFTR (PDB ID: 5W81) [320], metadynamics was used to push chloride around the CFTR selectivity filter, to resolve the conduction pathway. The path the

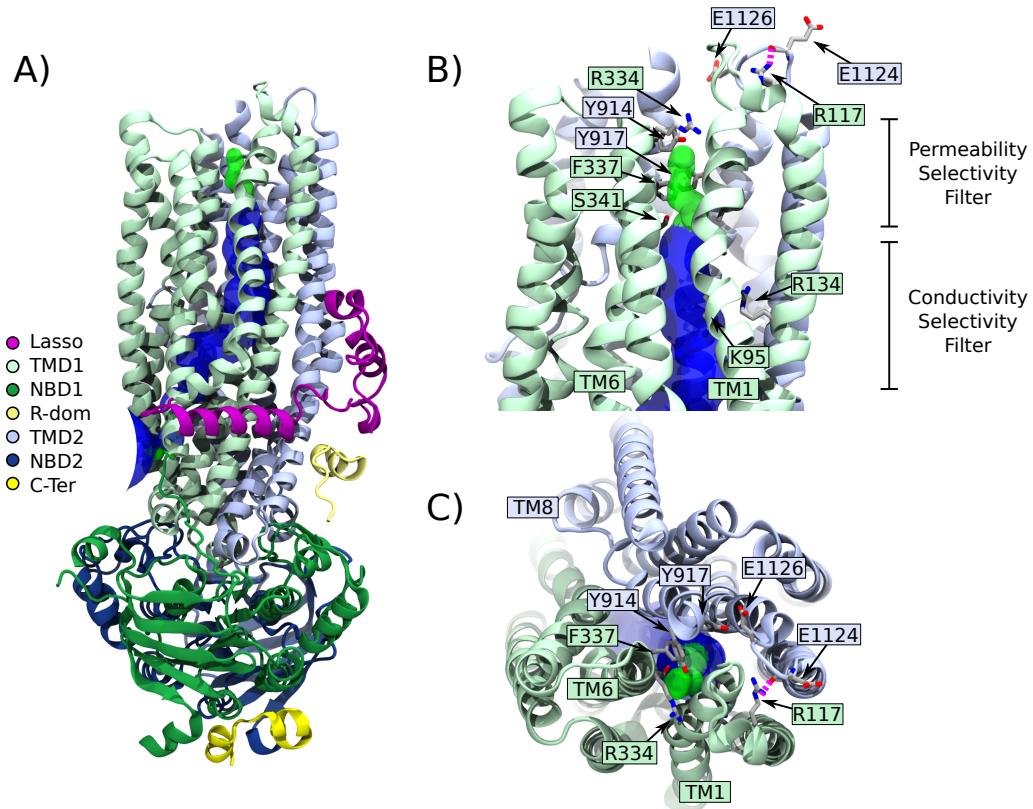


Figure 7.1: Constriction in Cryo-EM Structure (6MSM)

The most recent experimentally resolved structures of hCFTR bound to ATP has a constriction smaller than the radius of chloride ions [297]. The purpose of this study is to discover a dilated, conducting conformation of CFTR by using this cryo-EM structure as a starting point. A) The constriction occurs between TM1 and TM6 on TMD1. The space filling probe is colored blue when a hydrated ion may pass, while it turns green when the ion must be dehydrated. The probe truncates when it encounters a constriction smaller than an ion. B) Close up of the constriction. Observe how the constriction begins at S341 in the permeability filter and ends at R334 where there is no available space to pass. This leads us to believe that movements in TM6 will help the channel to open. C) A top down view of the selectivity filter. Observe how the architecture of TM1 and TM6 have resulted in a constriction around the conduction pathway.

authors discovered appears to agree with *in vitro* studies of the selectivity filter in human CFTR [325]. However, these MetaD calculations found a significant energy barrier to the passage of chloride (roughly 10 kcal/mol) [330]. This sizeable energetic barrier conflicts with experimental measurements of the channel's conductivity, and indicates that the tested protein conformation is not sufficiently open to readily conduct ions.

In the second study, hCFTR was studied with a strong applied electric field in MD simulations [333]. These authors were able to observe the permeation of chloride ions through the channel. However, the number of conduction events these authors observed would imply a conductivity of 0.54 pS (17 events in 10 μ s of sampling at a substantial 500 mV bias). This conductivity is an order of magnitude below the experimentally measured conductivity of CFTR (6-10 pS) [332, 356, 561, 562]. These two studies indicate that there are conformational states of CFTR, not yet experimentally visualised, which will support the conduction of chloride and other anions.

These prior MD studies have focussed on the conduction of chloride, since it is the most physiologically important species to permeate through CFTR. However, it is actually one of the smallest ions which can pass through CFTR. Larger ions such as bicarbonate and glutathione also move through the channel and their failure to do so plays an important role in disease pathology [331, 334, 335, 358, 563, 564]. Hence, it is clear that there is a need to move beyond the available cryo-EM structures of CFTR, to find a conducting conformation which can support the passage of larger anions as well as chloride.

An *in silico* attempt to resolve this issue has been made previously by constructing a model based on zCFTR (PDB ID: 5W81 [320]) and steering it toward the outward facing conformation of a distantly related ABC transporter, SAV1866 [321, 322]. However, this resulted in a conformation with a very different architecture to what would be expected from experiments, and so it is difficult to draw testable predictions from the resulting model [295, 322]. The pore architecture of CFTR has consequences for its function so it must be investigated carefully [295, 325, 518].

The helices which form the selectivity filter and their associated extracellular loops have been shown to exhibit considerable flexibility. These features play an important role in the selectivity of ions and gating kinetics [305, 312, 335, 565]. In particular, a recently performed electrophysiology study demonstrated altered gating kinetics in the R117H mutation. Previously, it was thought that R117 forms a salt bridge with E1126 [566]. But more recently, experiments by Simon and Csnady demonstrated that it is E1124 which makes a contact to R117, by interacting with the carboxyl oxygen atom in the backbone of R117, not E1126 [567]. This interaction is visible in available cryo-EM structures of hCFTR (Figure 7.1B).

These *in vitro* results correspond well to the implications of our presented *in silico* study. The current state of the literature leaves the role of E1126 in channel gating unresolved, but through our MD experiments we will show how in the conducting conformation of CFTR, E1126 appears to form a salt bridge with R334. We will also propose a test of this hypothesis through patch-clamp electrophysiology experiments.

Resolving the conduction mechanism of ion channels has been of keen interests to the computational biophysicists for some time [68, 568]. To our knowledge, this is one of the first studies to resolve the conducting conformation of an ion channel without a target conformation in mind.

Hence, the work in this chapter differs in an important way from existing computational studies of ion channels. Many previous attempts to induce conformational changes have been *targeted*—that is, they have usually sought to resolve the energy landscape *between* experimentally determined structural conformations of the protein [79, 322, 569, 570]. However, since there is not yet an experimentally determined structure of CFTR in a conducting conformation, the presented study is instead *untargeted*.

By combining subtle hints from *in vitro* biophysical experiments on CFTR with results from our MD simulations, we have been able to resolve and study a conducting conformation of the anion channel with the use of multiple ions. This signals how computational methods have become sufficiently developed to use experimentally obtained structural snapshots as starting points, to then explore the local conformational neighbourhood. This approach has the potential to discover new molecular mechanisms within cells.

7.2 Results

7.2.1 Long Simulations Reveal Motions Which Dilate the CFTR Pore

Four independent replicates of unbiased MD were run to simulate WT-CFTR. Each replicate was used to generate a trajectory which was at least $2 \mu\text{s}$ long. This gave a combined $8.2 \mu\text{s}$ of sampling.

Choosing CVs from these unbiased statistics is not trivial. In post processing, the motions of the transmembrane helices were analysed using Principal Component Analysis [571, 572] (Figure 7.2A).

The first two Principal Components represented 38% of the kinetic variance in the unbiased data. Additionally, they captured motions in TM1 and TM6 in motions which dilated the extracellular pore in a location we expected from experiments (Figure 7.2 B-C) [295, 312, 565]. These factors validated their choice as collective variables in our enhanced sampling calculations.

The amino acids analysed with PCA were chosen carefully (Table 7.1). Amino acids in disordered loops were excluded from analysis as they make large motions quickly. Large, fast motions were excluded because they would otherwise distort the optimal slow degrees of freedom which we hope to investigate with free energy calculations [227]. A more detailed discussion of our choice in CVs can be found in section 7.5.3.

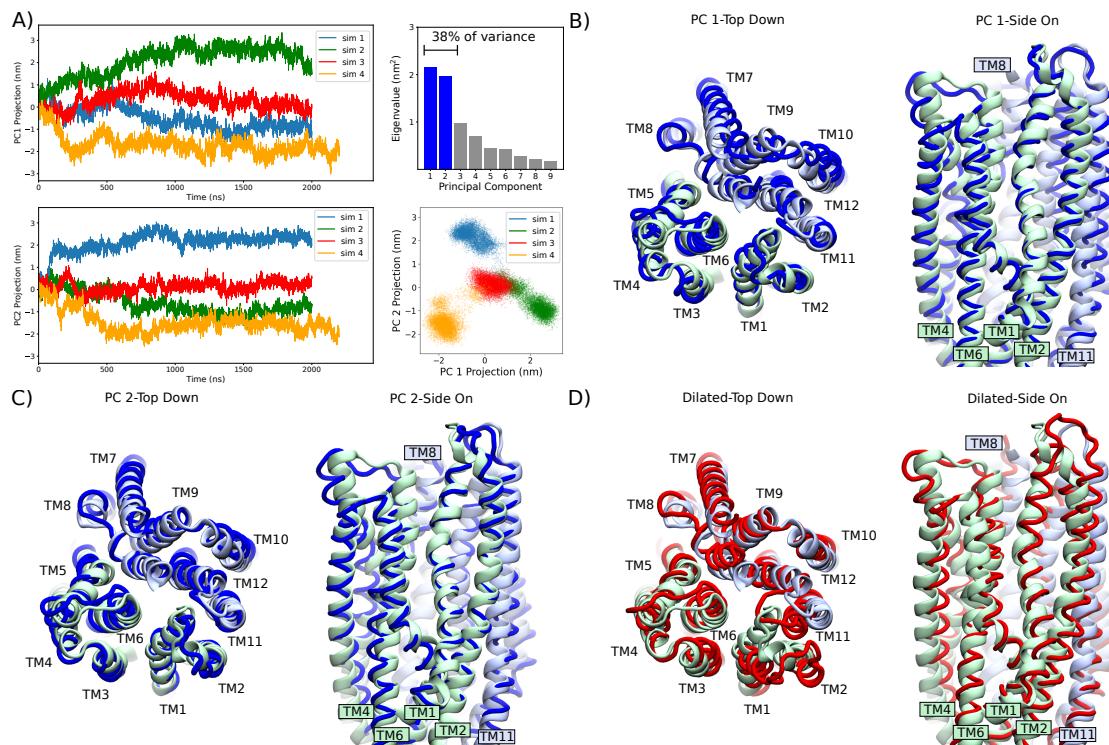


Figure 7.2: Principal Component Analysis Reveals Large Motions in CFTR Transmembrane Helices

A) Visualisation of the results from principal component analysis from 4 independent simulations, comprising $8.2 \mu s$ of unbiased MD sampling. Motions along the first two principal component vectors were found to capture 38% of the variation in the data set and were used in further analysis. B) Visualisation of the first principal component vector (PC1). Note the motion in TMs 1, 6 and 8 C) Visualisation of the second principal component vector (PC2). Note the motions in TMs 1 and 6 but in directions orthogonal to PC1. A video has been rendered of the motions associated with PC1 and PC2 and can be accessed at the following [link.¹](#) D) Visualisation of the dilated open state which we will analyse in detail in subsequent sections. This structure was obtained by a linear combination of PC 1 and 2 followed by unbiased MD simulations.

Because the more atoms we included in our collective variables, the slower our calculations would run, we limited subsequent free energy calculations to consider a set of helices which directly surrounded the pore (Table 7.1 Figure 7.7). The same technique was employed in [322]. In the next section, we will show how we combined these two PC motions with OPES-Metadynamics to discover a dilated conformation of the CFTR pore (Figure 7.2D).

7.2.2 OPES-Metadynamics Finds a Stable Conformation with an Open Extracellular Pore

The principal components identified in the previous section were used as collective variables to study the motion of the pore lining transmembrane helices in WT-CFTR.

¹<https://zenodo.org/record/7036443#.Yw7CZNJBwUEi>

A two dimensional Free Energy Surface along these CVs was constructed through the use of 8 independent walkers, each running for $1.5 \mu\text{s}$ with On the Fly Probability Enhanced Sampling-MetaDynamics (OPES-MetaD), using a barrier parameter of 38.2 kcal/mol (7.3A). Convergence of these calculations is demonstrated in Figures 7.8-7.10.

Note the presence of a local free energy minima at (0.010, 0.021) in Figure 7.3A. When this coordinate in CV space was investigated closely, we noticed that the extracellular pore had dilated significantly (Figure 7.3B). Further, this dilated configuration was stable in 3 replicates of unbiased MD on timescales of 500 ns (Figure 7.3C).

This dilated structure has a pore architecture where the narrowest constriction was 2.8 Å (Figure 7.3C). This larger opening completely removed the aberrant constriction seen in the cryo-EM structure, and the dilation occurred where we expected, between TM1 and TM6 (Figure 7.3D). Continuum models have previously estimated the constriction in CFTR to be 2.4 Å [331], which is close to our own estimate.

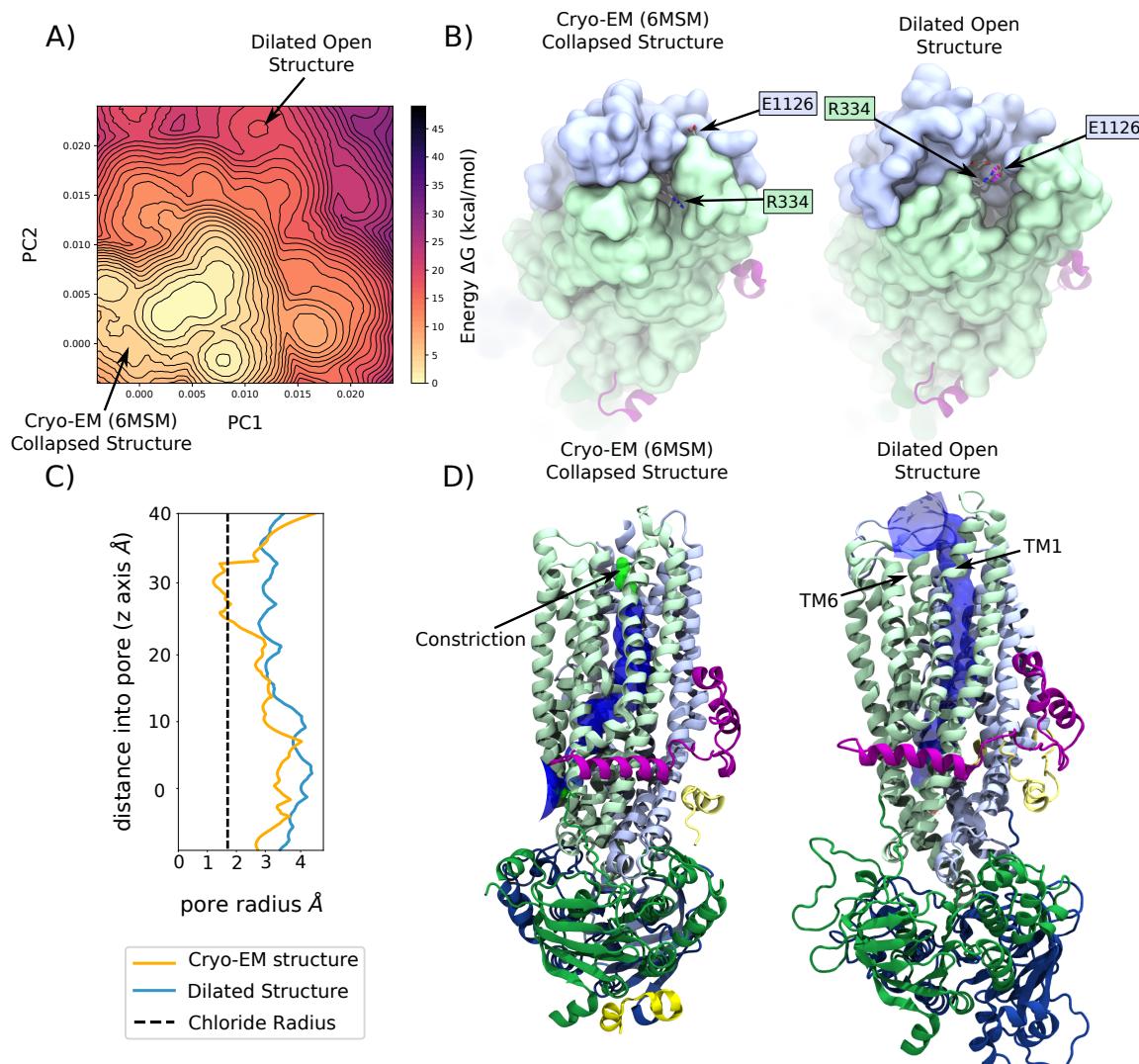


Figure 7.3: Free Energy Calculations Discover a Dilated, Conformation of CFTR.

A) The free energy landscape obtained along PC1 and PC2 through OPES-Metadynamics. Note the free energy minimum corresponding to a collapsed structure found via cryo-EM in the bottom left, and the free energy minimum at (0.010,0.021). This latter well was found to correspond to a stable, open structure which appears capable of passing anions. B) A visual comparison between the constricted structure and our dilated open structure. Note how the extracellular opening is much wider in the dilated structure. The R334 and E1126 amino acids are visualised to show how they come into contact with each other in the dilated structure. C) A quantitative measurement of the pore radius for different conformations of CFTR, generated with HOLE2 [573]. Dehydrated chloride has a radius of 1.7 Å, while the collapsed cryo-EM structure of hCFTR exhibits a constriction of 1.2 Å. Hence, even if chloride were to be dehydrated during its migration through the selectivity filter, there would need to be some change to the conformation of the selectivity filter. By contrast, our proposed dilated open structure exhibits a constriction of 2.8 Å at its narrowest point. This pore profile was generated from a snapshot obtained after 500 ns of unbiased MD. Indicating that this open structure is energetically stable. D) A side on comparison between the collapsed cryo-EM structure and the dilated open structure. The green constriction denotes the region too small to pass ions. This constriction has been completely removed in the dilated structure we have discovered.

7.2.3 The Open Conformation is Stabilised by the R334-E1126 Salt Bridge

To search for interactions which might be stabilising the dilated structure, giving rise to the free energy minimum in Figure 7.3A, we inspected pairs of positive and negative amino acids. This led us to the discovery of a novel salt bridge linking R334 and E1126. This salt bridge could only form in the dilated conformation, due to the movements of TM1, TM6 and TM8. It could not form in the collapsed cryo-EM structure (Figure 7.4A).

The link between these amino acids was found to be stable in 3×500ns unbiased MD simulations of the dilated structure. While it did not form at all in simulations of the cryo-EM structure (Figure 7.4B).

This is a novel interaction within CFTR which we have discovered through the use of simulations. Suggestions that this salt bridge exists can be found in the literature, but it has not been confirmed [332, 566, 574, 575]. We suspect this was due to the important role that R334 plays in conduction of anions and channel gating. We will discuss the implications of this salt bridge and the means for testing its presence *in vitro* in the Discussion.

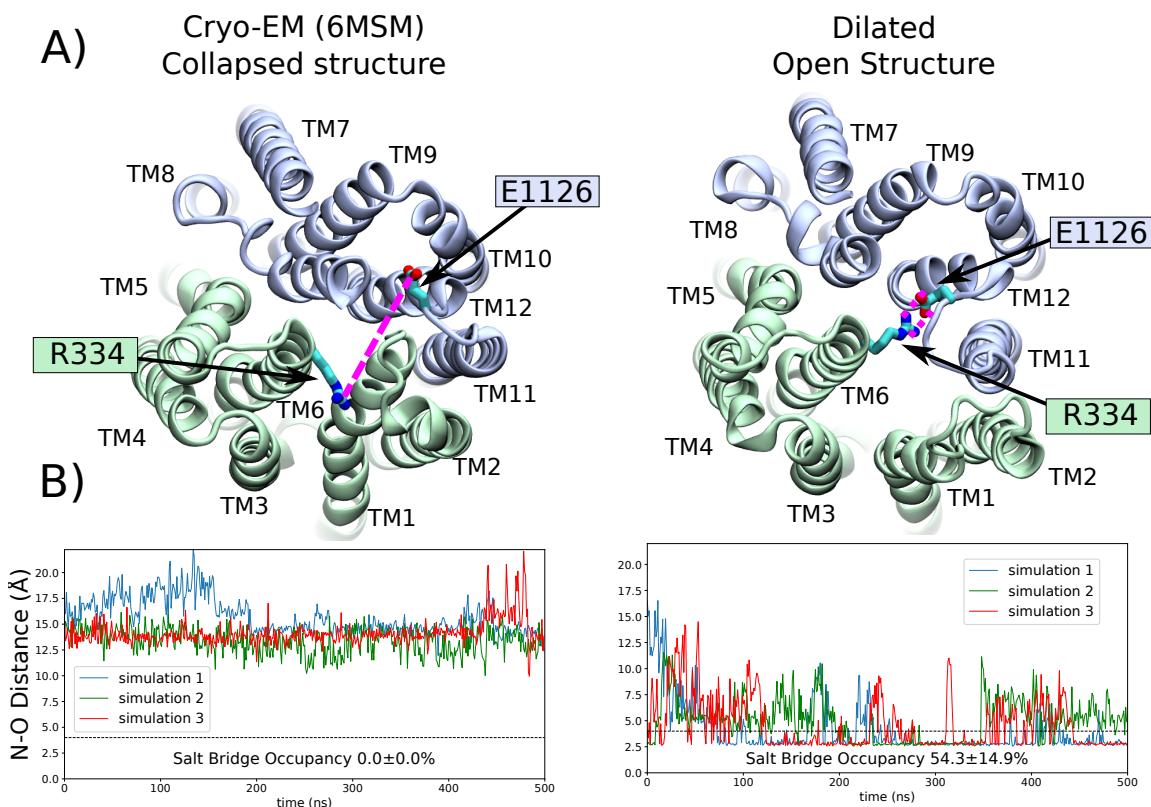


Figure 7.4: The Dilated Conformation of CFTR Gives Rise to the R334-E1126 Salt Bridge

A) A comparison between cryo-EM and dilated CFTR structures. In the cryo-EM structure, R334 and E1126 are far apart and do not make contact. In the dilated structure, the transmembrane helices rearrange such that R334 and E1126 make contact. B) In simulations of the cryo-EM structure, the salt bridge is never formed. However, in the dilated structure R334 and E1126 come close together and form a stable salt bridge. The black dotted lines in both plots denotes the distance below which the side chains are in steric contact with each other. We believe that this salt bridge makes a significant energy contribution to the stability of the open channel.

7.2.4 Proving the Open Conformation is Capable of Ion Conduction with Umbrella Sampling

In order to test the ability of the dilated structure to conduct anions, we employed umbrella sampling to calculate the energy landscape of anions as they permeated through the selectivity filter of CFTR.

The conduction of chloride through WT-CFTR in simulations of the cryo-EM structure was severely inhibited (Figure 7.5A). We obtained a sizeable barrier of 9.7 kcal/mol to the passage of chloride, demonstrating that without dilation, this structure will not allow chloride to pass at a rate in agreement with experiments [332]. By contrast, the dilated conformation obtained in the previous section removes steric barriers to the passage of chloride and larger anions (Figure 7.5B). In this structure, a considerably smaller barrier of only 2.5 kcal/mol was observed for both chloride and the larger bicarbonate (Figure 7.5C).

This small barrier is closer to what we would expect to be overcome by the -46 mV resting potential of epithelial cells [576]. Furthermore, the remarkably similar profiles for both bicarbonate and chloride are in agreement with the energy landscape we would expect from a weakly selective channel like CFTR. Indeed, through our umbrella sampling we have calculated energy landscape through the "permeability selectivity filter" of CFTR. Deeper in the pore is the "conductivity selectivity filter", where we might expect a difference between the two anions. The lower FES in the vicinity of R134 K95 in Figure 7.5C for bicarbonate compared to chloride may indicate tighter binding at this site. This would lead to lower overall conductance of bicarbonate. Unfortunately, there have not been many studies which have measured the effect of mutations on bicarbonate conductance, but studies of other anions would lead us to expect that R134 and K95 play important roles in the conduction of bicarbonate [325].

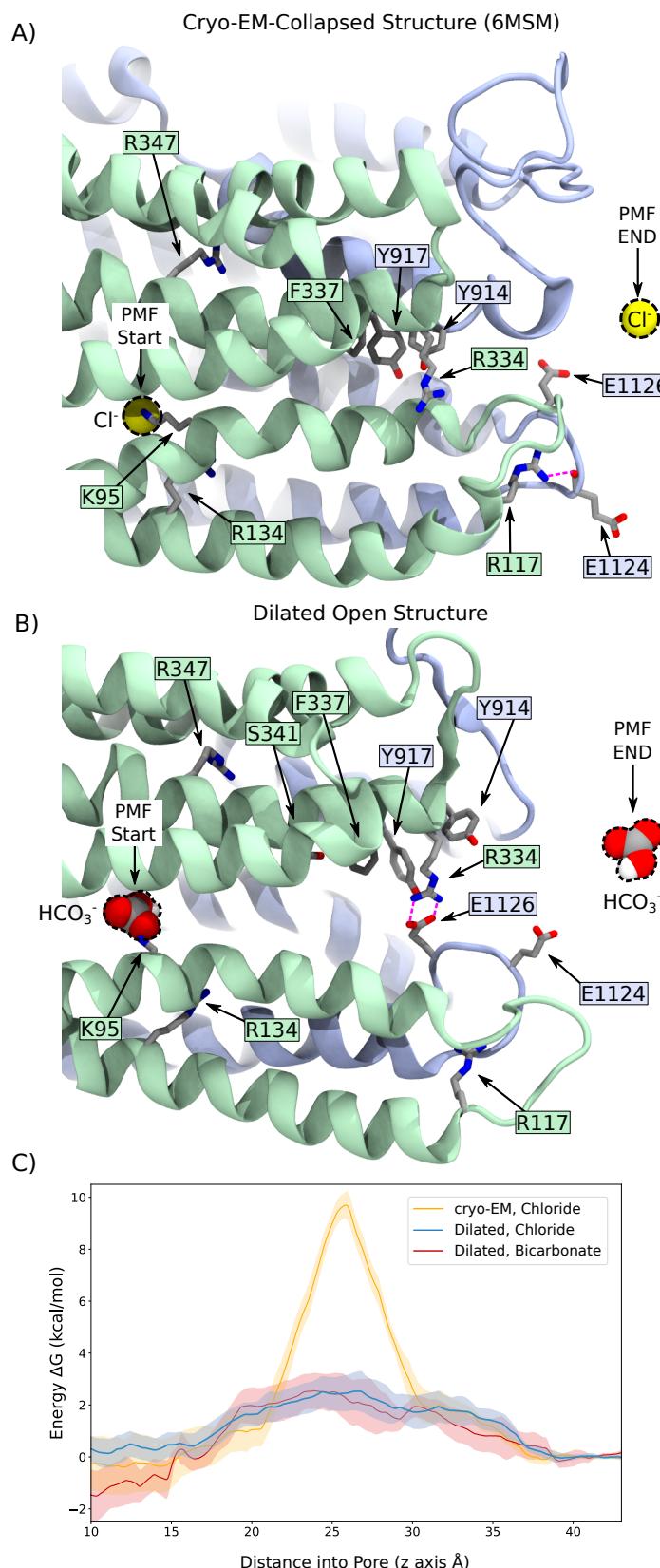


Figure 7.5: The Dilated Conformation of CFTR is Capable of Conducting Anions

A) A close-up visualisation of the selectivity filter in the collapsed cryo-EM structure.

To pass through this structure, the ion must move past a tight constriction formed by F337, Y914 and Y917. B) A visualisation of the selectivity filter in our dilated *in silico* structure. Here, bicarbonate is pictured but the same conformation was used to construct a PMF of chloride as well. The selectivity filter is not as constricted-and R334 makes a salt bridge with E1126. This brings R334 closer to the conduction pathway where it can help coordinate the permeating anion. Steric barriers to the conduction of anions have also been removed C) PMFs of the anions as they pass through the narrowest point in CFTRs permeation pathway. The collapsed cryo-EM structure exhibits a significant barrier to the conduction of chloride, which is inconsistent with experimental measurements. By contrast, the dilated structure appears capable of permeating both chloride and bicarbonate, with a landscape in agreement with the characterisation of the channel found in experiments.

7.2.5 The Open Conformation Can be Used to Study Disease Causing Mutations in the Selectivity Filter, such as R334W

There are several disease causing mutations which occur at around the selectivity filter of CFTR (such as E116K, R334{W,Q,L},R117{H,C,P}, I336K, S341P)[86]. Note how these are all clustered around TM1 and TM6. The fact that mutations to these amino acids causes a loss of function is more evidence to suggest TM1 and TM6 form the permeation pathway for anions in hCFTR, in agreement with previous MD and functional studies [295, 325, 330, 333].

A molecular understanding of these mutations can aid in the design of new therapeutics, so as an example, we have used our dilated structure to study the extracellular R334W mutation. Patch clamp measurements of R334W-CFTR have found a single channel reduction in current of 60% compared to WT-CFTR [332]. According to the Nernst-Planck model, this means we expect a larger energy barrier to the conduction of chloride in the mutant compared to WT-CFTR on the order of 0.6 kcal/mol [577]. This is close to differences in barrier height we observed through umbrella sampling which were between 0.7 and 2.5 kcal/mol (Figure 7.6B).

7.3 Discussion

The present study differs in an important way from recent *in silico* investigations of large-scale conformational changes in ion channels. These past studies have focussed on recreating intermediate free energy landscapes between *known* endpoints in conformational space [79, 569, 578]. These studies were *targeted*. Such approaches are critical to the development of molecular techniques to help us understand the energetics and kinetics of protein systems. However, these techniques are inherently more limited by the availability of high quality experimental protein structures than *untargeted* methods. So, we regard this *untargeted* study as a proof of concept—to show how computational methods can move beyond the availability of protein structures.

One could regard the differences between *targeted* and *untargeted* MD methods as similar to the development of unsupervised vs. supervised machine learning methods [579]. Each has its own advantages and drawbacks.

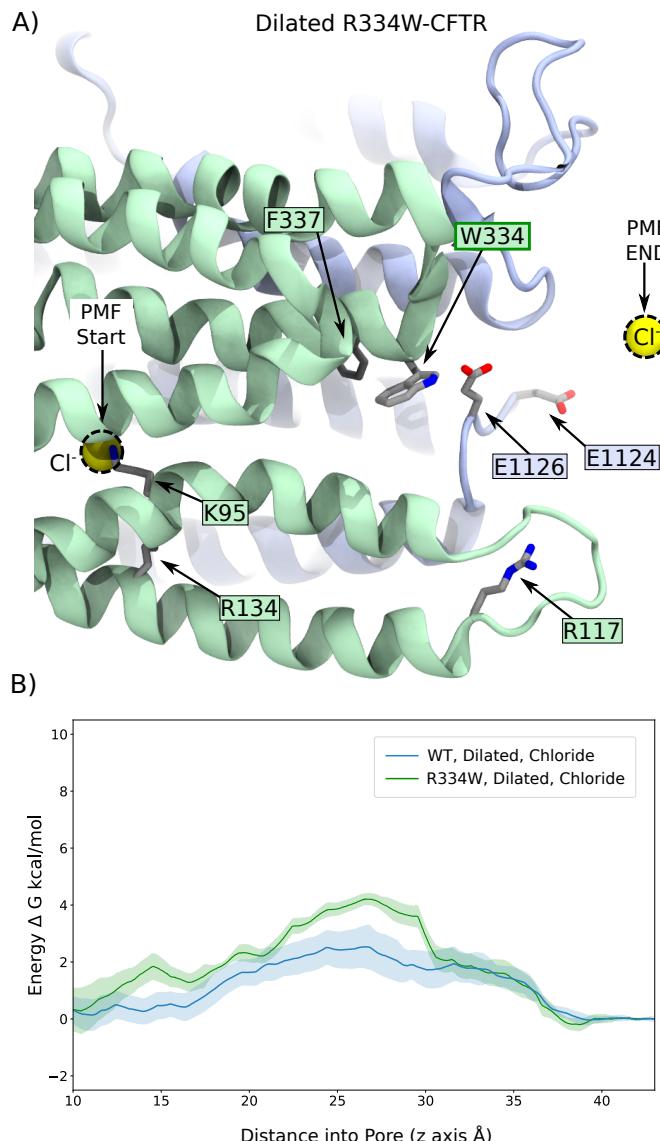


Figure 7.6: R334W inhibits the Conduction of Anions

A) A visualisation of R334W-CFTR in the dilated conformation. Note how W334 is placed in the outer part of the pore. B) A PMF derived from Umbrella sampling to compare the energetics of the conductance of chloride through R334W-CFTR to WT-CFTR in the dilated conformation. The WT-CFTR curve has been reproduced from the same data in Figure 7.5C. The R334W-CFTR PMF exhibits a larger barrier compared to WT-CFTR indicating that the conduction of chloride will be suppressed in this mutation. It is likely that this additional energy barrier has a particularly deleterious effect on the conduction of the channel because it occurs at the narrowest part of the pore. There are no alternate pathways which might help compensate for the mutation, as was the case for R352Q in chapter 5 [304].

7.3.1 The Dilated Structure Is Consistent with Experiments

It is pleasing that the results from this chapter appear to be consistent with results from *in vitro* biophysical experiments. The movements in TM1, TM6 and TM8, which dominate the PC1 and PC2, were the kind of dilation we expected from biophysical experiments studying CFTR’s selectivity filter [312, 325]. The dilated structure also matches experiments which studied the binding of a protein partner WNK1 to TMD1 and other studies which looked at the permeation of bicarbonate [305, 331]. These studies indicate that movements of TM1 result in a wide open conformation, which facilitates the movement of bicarbonate ions, in agreement with our own discovery.

Additionally, the similar conductance profiles between bicarbonate and chloride are what we would expect from a weakly selective anion channel like CFTR [325]. Recall that the diffusion coefficient of a particle scales inversely with its radius according to the Stokes-Einstein relation [580]. Since chloride is smaller than bicarbonate, we would expect it to permeate more quickly through the channel, even if it passed over the same energy landscape, particularly through the permeability selectivity filter.

Given that the conductivity ratio in CFTR is $P_{\text{HCO}_3}/P_{\text{Cl}} = 0.25$, we would expect each ion to have a roughly similar free energy profile through the open channel. At the same time, the selectivity filter in our structure still has a constriction of 2.8 Å, while a hydrated chloride ion has a radius of 3.2 Å, so partial dehydration must still take place during anion conduction. Hence, our dilated conformation is still consistent with the experimentally determined lyotropic selectivity sequence found in other CFTR and other anion channels [325].

Cysteine cross-linking has been performed on many pairs of amino acids within CFTR [312, 518, 565, 581]. Our proposed conducting conformation is in close agreement with these studies, and in some cases, in closer agreement than the available cryo-EM structures. These studies found that the T338C/S1118C and R334C/T1122C cross-links could form in the open state but not in the closed state [581]. In the cryo-EM structure, TM8 stands in the way of these cross links. However, our dilated structure allows unobstructed access between TM6 and TM12, which would allow these cross-links to form.

With the elucidation of CFTR in a conducting conformation, it may be possible now to design nano bodies or small molecules which select for it [582]. This would give rise to a new set of tools to understand CFTR’s function or even new therapeutics to treat CF. Therapeutics which target the conducting conformation of CFTR will likely be much more effective than current generation potentiators, as it remains unclear whether or not these drugs select for the fully open conformation of CFTR [376, 378].

7.3.2 Study Limitations

We should address the considerable difference in height between the global free energy minimum in Figure 7.3A, which closely corresponds to the cryo-EM structure, and the minimum corresponding to our dilated structure. The dilated structure would appear to have a height of 17 kcal/mol. At first, this difference would seem unphysically high. If these values were correct, the dilated structure would be inaccessible at physiolog-

ical temperatures. However, we believe this discrepancy is due to limitations of our simulation methods and we have plans to improve them. We believe there are 3 systematic issues which have led to this discrepancy, and outlining these issues will help us improve the study before submission for publication.

The first and most pressing concern would be to improve our collective variables for OPES-MetaD. The PCA motions outlined in Figure 7.3 are likely suboptimal. Note for example how in the dilated structure in Figure 7.2D, TM1 has moved significantly but this motion is poorly captured by the two motions in Figure 7.2B-C. Hence, the movement of TM1 is likely in a hidden CV, not captured by PC1 and 2 [207, 208]. It is motions like this which have likely led to some hysteresis during the construction of the free energy surface in Figure 7.3A.

Given that the unbiased simulations of the dilated structure are stable, and they agree with experimental studies, we expect that the dilated state indeed corresponds to a physiologically relevant protein conformation. For the next iteration of this study, we plan to employ more advanced dimensionality reduction techniques such as TICA, VACs and RAVE to make a more careful choice of CVs and gain a clearer picture of the energetics of the fully open state [227–230]. We expect that the simple reweighting scheme of OPES-MetaD alongside SGOOP optimisation will be of great help in designing optimal CVs [583–586].

The second missing factor from the calculation of this FES is the energetic contributions from the anions, as they pass through the selectivity filter. This is not captured by our CVs. The permeation pathway through CFTR is unusually long, and there has been a demonstrated energetic contribution to the opening transition from the anions anions during channel gating [332, 519, 587–591].

Finally, we consider the effects of the lipid bilayer which surrounds CFTR. There is emerging interest in the role played by the lipid bilayer in the regulation of CFTR and other membrane proteins [592–597]. Different bilayer compositions have different physical properties, leading to different behavior by the proteins embedded in them [598]. In the case of CFTR, many lipid species such as sphingolipids and cholesterol has been shown to lead to different rates of chloride transport, indicating that they play a role in regulating the gating of the channel [592, 595, 599]. Hence, it is likely that the zwitterionic POPC membrane used for these simulations struggles to reflect the energetics and kinetics of the membrane surrounding CFTR in its native environment.

The effect of the non-native lipid bilayer composition would also account for the discrepancies we find between our simulations and experiments. Careful biophysical experiments discovered an interaction between the R117 and the carboxyl atom of E1124 [567]. This interaction was not present in our dilated structure, nor was it well preserved in our unbiased simulations of the cryo-EM conformation (data not shown). Similarly, the well characterised R104-E116 interaction is poorly preserved in simulations of the cryo-EM structure and simulations of the dilated structure [566]. This discrepancy is also visible in previous MD studies of CFTR in a non-native lipid environment [333]. These salt bridges were often broken in our MD simulations, due to interactions between the charged amino acids and the phosphate or choline chemical groups in the phospholipid bilayer. In future we could use newly developed model

bilayers of the epithelium to simulate CFTR in a native-like environment [600].

We regard the above points as room for improvement, and they will be considered carefully in further studies of this dilated structure.

7.3.3 Proposed Experiments to Test the R334-E1126 Salt Bridge

The predictions of a stable salt bridge in section 7.2.3 fill a recent gap in the literature. The elegant study on the R117H mutation from Simon and Csandy [567] discovered that a long standing conclusion about R117 making a connection with E1126 was incorrect [566]. In fact, R117 makes a stable hydrogen bond with E1124, not E1126. This leaves the partner of the negative charge in E1126 unknown. The results from our dilation study would appear to fill this gap in the literature.

Above we have presented simulation results which suggest that E1126 makes a salt bridge with R334 in the conducting conformation of CFTR. There are some hints from previous biophysical studies to suggest that this salt bridge in fact forms *in vitro*, but further experimental work would be needed to characterise this salt bridge.

The critical role of E1126 to CFTR gating is demonstrated in Cui et al. [566]. While the most direct *in vitro* suggestion of the formation of the R334-E1126 salt bridge comes from Wang et al. [574]. In this latter study, the researchers demonstrated that Zinc blocked the conduction of chloride through R334C-CFTR, indicating that zinc was binding at the extracellular mouth and blocking chloride. Because zinc has a charge $+2e$, they suspected that a nearby negative amino acid might also play a role in binding the zinc cation. Further experiments showed conduction through the double mutant R334C/E1126A-CFTR was no longer inhibited by zinc ions. This led the authors to suggest that R334-E1126 come close together in the open structure, forming a salt bridge. This is consistent with our *in silico* findings which indicate that R334 and E1126 indeed form a salt bridge in the conducting CFTR conformation (Figure 7.4).

It would have been very difficult to discover the nature R334-E1126 interaction experimentally. This is because R334 plays such an important role in the conductance and gating of CFTR [332, 575, 581, 601]. With the use of the atomic resolution offered by MD simulations, we have been able to fill this gap in the experimental literature, demonstrating the power of *in silico* methods for studying protein dynamics. We propose that single ion channel patch clamp experiments measuring mutant CFTR could demonstrate the importance of the E1126-R334 salt bridge to the open state of the channel.

Should the R334-E1126 play a role in stabilising the open channel, we would expect mutations to both E1126 and R334 mutations to both exhibit a gating defect. There are already documented cases where mutations to these amino acids produce a gating defect in the literature, and more careful investigations of these mutations would confirm the role of the R334-E1126 salt bridge in stabilising the fully open channel.

A clear gating defect has been observed in measurements of E1126R-CFTR [566]. Meanwhile, studies of R334C-CFTR found frequent transitions to a partially open

state, with only infrequent transitions to the fully conducting state [575, 601]. This is the behaviour we would expect from our simulation results. However, we should note that introduction of a positive charge at C334 in the latter study did not produce a change in the profile of accessed transient gating states, only an increase in single channel conductance. These is an interesting observation, which might at first appear to be inconsistent with our hypothesis. However, the MTSES⁺ agent used in this study does not resemble geometry of the arginine sidechain it is seeking to replace. Indeed, CFTR is extremely sensitive at the R334 position, with R334K-CFTR also resulting in a lower conductance than WT-CFTR [332]. The fact that the profile of transient gating states is the same *after* MTSES⁺ modification indicates that function has not been fully restored by covalent modification with MTSES⁺. The fully conducting state, on par with WT-CFTR, is still poorly sampled indicating the modified channel still exhibits a mild gating defect (this can be seen in Figure 2E in [575]). In brief, patch-clamp measurements of R334C-CFTR can be difficult to interpret due to the flickering phenotype this mutation exhibits. However, overall, these results are consistent with the behaviour we expect from our hypothesis concerning the R334-E1126 salt bridge. Further, the fact that the R334C mutation visits a fully open conformation, even briefly, validates our use of a fully open conformation to study the R334W mutation, even if these mutations were to exhibit a gating defect.

These results would lead us to expect that mutations like E1126A would have a similar gating phenotype to the R334C mutation. Should this be confirmed in experiments it would confirm the presence of our discovered R334-E1126 salt bridge in the conducting conformation of CFTR. If breaking both sides of this salt bridge results in the same phenotype, it would confirm that the two amino acids are in contact *in vitro*. A double mutant such as R352C/E1126A-CFTR might make such studies easier, as the R352C mutation exhibits flickering phenotype which makes it possible to study the transient open states of CFTR and gain a more complete picture of its gating kinetics [526, 602]. One could also test the charge swapped mutant R334E/E1126R or use cysteine cross linking to study the double mutant R334C/E1126C. If these constructs restore chloride transport, it would confirm the presence of the R334-E1126 salt bridge in WT-CFTR. However, this seems unlikely given the sensitivity of CFTR to changes at the R334 position [332].

The role of R334 in the formation of this salt bridge would have previously been obscured in *in vitro* experiments because of the important role it plays in the gating and conductivity of CFTR [332, 587]. Hence, we have demonstrated the complimentary power of MD simulations to wet lab biophysical experiments.

7.3.4 The Open Conformation is Important in the Study of Rare Mutations

Our findings from this study tie into the broader theme of our work. In previous chapters we performed MD simulations to elucidate unique modes of CFTR malfunction and then showed that they are amenable to rescue by CFTR modulators. The simulations in this study helped us resolve the conductance defect exhibited by the R334W mutation, and now we will search the literature for evidence that this defect can be

rescued by CFTR modulators.

It has been observed that despite a basal reduction in chloride transport of 98% in FRT cells [514], R334W-CFTR function may still be restored by combination therapies of CFTR modulators intestinal organoids [603]. Clinical trials are currently underway to test whether these findings translate to better disease outcomes in patients [604]. Hence, the R334W mutation is another case where a unique molecular defect appears to respond to CFTR modulators. The open structure we propose here will help in the study of other rare CF-causing mutations at the pore mouth, hopefully leading to better patient outcomes and the discovery of new therapeutics.

7.4 Conclusion

Here we have performed extensive Molecular Dynamics and Free Energy Simulations to study the permeation of chloride and bicarbonate through the selectivity filter of CFTR. This was motivated by outstanding questions surrounding the conduction mechanism of this anion channel, due to its constricted selectivity filter in currently available Cryo-EM structures. By applying OPES-Metadynamics and umbrella sampling, we were able to demonstrate that a dilated conformation of CFTR is more capable of supporting the conduction of both chloride and bicarbonate. Further, we demonstrated how this dilated structure is in better agreement with experiments and how it can be used to study rare disease-causing mutations.

Through the careful study of our dilated open structure, we also discovered that a salt bridge between R334 and E1126 forms in the open structure, but not in the collapsed cryo-EM structure. This led us to suggest single ion channel patch-clamp electrophysiology experiments which could confirm the presence of this salt bridge *in vitro*. The confirmation of this open structure could lead to novel, more effective therapeutics.

The work in this chapter suggests that computational methods are now sufficiently developed such that creative application of MD simulations and enhanced sampling techniques can help complete our molecular understanding of proteins. Similar techniques to the ones employed in this chapter could be used in many protein systems. It is common for proteins to be imaged very close to their active conformation, but during the imaging process they change conformation slightly [605]. Simulations could be used to study these proteins in their active state. A limited but powerful example of this would be to further refine the methods applied in this chapter and then apply them to the partially collapsed structure of P-glycoprotein [606]. This important structure could be dilated, in order to resolve the fully outward facing configuration, in a manner very similar to what we have done to CFTR.

7.5 Methods Details

7.5.1 System Composition

At a basic level, the methodology used in this chapter is largely consistent with the other studies in this thesis with some key differences. The system was also constructed

by embedding the extended CFTR model into a POPC membrane and then the whole system was immersed in a neutralising 0.15 mol/L KCl solution.

7.5.2 Unbiased MD

The minimisation and equilibration steps for these simulations were largely carried out in the same way as for previous chapters. Our protein model was the extended CFTR model based on the 6MSM structure [297], which we constructed in chapter 4. For all calculations in this chapter, except the initial unbiased calculations, the C-terminus was truncated at amino acid 1450 in order to make the simulation box smaller and save computer time. The MD system was subjected to minimisation by steepest descent followed by restrained relaxation. This involved 10 kcal/mol/Å² harmonic restraints on all heavy protein atoms with restraints halved every 400 ps. This relaxation was followed by up to 2.2 μs of unbiased molecular dynamics sampling.

7.5.3 Principal Component Analysis and Choice of Collective Variables

Using principal component analysis on proteins can be difficult. PCA specialises in choosing the vectors which capture the *largest* variation in a dataset. By contrast, when we are performing free energy calculations, where we would wish to study the *slowest* motions in a physical system [227]. So, when performing PCA on proteins, unless excluded, large disordered loops would dominate the analysis. Hence, we limited our analysis to the amino acids in Table 7.1.

Our four unbiased trajectories were combined using `gmx trjcat`, PCA was then performed using `gmx covar` and finally the vectors were analysed and visualised using `gmx anaeig`. The concatenation of the unbiased trajectories in this way is mathematically valid—as when PCA is solely performed 3D coordinates, it is agnostic to the temporal dimension of the data points [579].

For the sake of performance, a smaller subset of amino acids were then chosen to actually manipulate the pore of CFTR in the OPES-Metadynamics simulations. These calculations were computationally expensive and the more atoms included in our CVs, the slower the calculations would run. Hence, the amino acids highlighted in red in Table 7.1 were chosen as the CVs in our OPES-MetaD calculations. These were the pore lining amino acids and our choices closely reflect those of previous studies (Figure 7.7) [322].

There are two main disadvantages when using PC's as CVs. Firstly, as discussed, they correspond to the largest motion in a system, not the slowest. This means that they are very likely to be suboptimal CVs in a molecular system. Secondly, they are *linear*. And so a linear combination of many PCs may be needed to find the minimum free energy pathway of a given transition. This gives rise to a trade off. The more CVs we have to consider during the free energy calculation, the more likely we are to find a minimum energy pathway, but the longer it will take to converge the FES. Benchmarking of our own calculations revealed that only 2 CVs could be used before the calculations slowed to unmanageable rates on available hardware. This led us to choose the first two PC's 1

and 2 from our analysis as the collective variables to accelerate our enhanced sampling. These choices were made carefully. PCs 1 and 2 captured 38% of the variance of the unbiased data, meaning they are meaningful CVs, and they produced movements in TM1 and TM6 in the directions we expected given the available biophysical evidence.

TM1 ↑	TM2 ↓	TM3 ↑	TM4 ↓	TM5 ↑	TM6 ↓	TM7 ↑	TM8 ↓	TM9 ↑	TM10 ↓	TM11 ↑	TM12 ↓
113	218	219			885			1013			
114	217	220			884			1014			
115	216	221			883			1015			
116	215	222			882			1016			
117	214	223			881			1017			
112	118	213	224	330		880			1018		1124
111	119	212	225	329	331	879	911		1019		1125
110	120	211	226	328	332	878	912	1012	1020	1123	1126
109	121	210	227	327	333	877	913	1011	1021	1122	1127
108	122	209	228	326	334	876	914	1010	1022	1121	1128
107	123	208	229	325	335	875	915	1009	1023	1120	1129
106	124	207	230	324	336	874	916	1008	1024	1119	1130
105	125	206	231	323	337	873	917	1007	1025	1118	1131
104	126	205	232	322	338	872	918	1006	1026	1117	1132
103	127	204	233	321	339	871	919	1005	1027	1116	1133
102	128	203	234	320	340	870	920	1004	1028	1115	1134
101	129	202	235	319	341	869	921	1003	1029	1114	1135
100	130	201	236	318	342	868	922	1002	1030	1113	1136
99	131	200	237	317	343	867	923	1001	1031	1112	1137
98	132	199	238	316	344	866	924	1000	1032	1111	1138
97	133	198	239	315	345	865	925	999	1033	1110	1139
96	134	197	240	314	346	864	926	998	1034	1109	1140
95	135	196	241	313	347	863	927	997	1035	1108	1141
94	136	195	242	312	348	862	928	996	1036	1107	1142
93	137	194	243	311	349	861	929	995	1037	1106	1143
92	138	193	244	310	350	860	930	994	1038	1105	1144
91	139	192	245	309	351	859	931	993	1039	1104	1145
90	140	191	246	308	352	858	932	992	1040	1103	1146
89	141	190	247	307	353	857	933	991	1041	1102	1147
88	142	189	248	306	354	856	934	990	1042	1101	1148
87	143	188	249	305	355	855	935	989	1043	1100	1149
86	144	187	250	304	356	854	936	988	1044	1099	1150
85	145	186	251	303	357	853	937	987	1045	1098	1151
84	146	185	252	302	358	852	938	986	1046	1097	1152
83	147	184	253	301	359	851	939	985	1047	1096	1153
82	148	183	254	300	360	850	940	984	1048	1095	1154
81	149	182	255	299	361	849	941	983	1049	1094	1155
80	150	181	256	298	362	848	942	982	1050	1093	1156
79	151	180	257	297	363	847	943	981	1051	1092	1157
78	152	179	258	296	364	846	944	980	1052	1091	1158
77	153	178	259	295	365	845	945	979	1053	1090	1159
76	154	177	260	294	366	844	946	978	1054	1089	1160
75	155	176	261	293	367	843	947	977	1055	1088	1161
74	156	175	262	292	368	842	948	976	1056	1087	1162
73	157	174	263	291	369	841	949	975	1057	1086	1163
72	158	173	264	290	370	840	950	974	1058	1085	1164
71	159	172	265	289	371	839	951	973	1059	1084	1165
70	160		266	288	372	838	952	972	1060	1083	1166
69	161		267	287	373	837	953	971	1061	1082	1167
68	162		268	286	374	836	954	970	1062	1081	1168
67	163		269	285	375	835	955	969	1063	1080	1169
66	164		270	284	376	834	956	968	1064	1079	
65	165		271	283	377	833	957	967	1065	1078	
64	166		272	282		832	958	966	1066	1077	
63	167		273	281		831	959	965	1067	1076	
62	168			280		830	960			1075	
61	169			279		829	961			1074	
60	170			278		828	962			1073	
59	171			277		827	963			1072	
58				276		826	964			1071	
57				275		825				1070	
56				274		824				1069	
55										1068	

Table 7.1: Amino acids used to analyse and manipulate the outer pore of CFTR
All listed amino acids were included in the unbiased extraction of PCA components.

The cells highlighted in red denote the amino acids which were used as collective variables to study the free energy landscape of PCA motions 1 and 2. These are visualised in Figure 7.7.

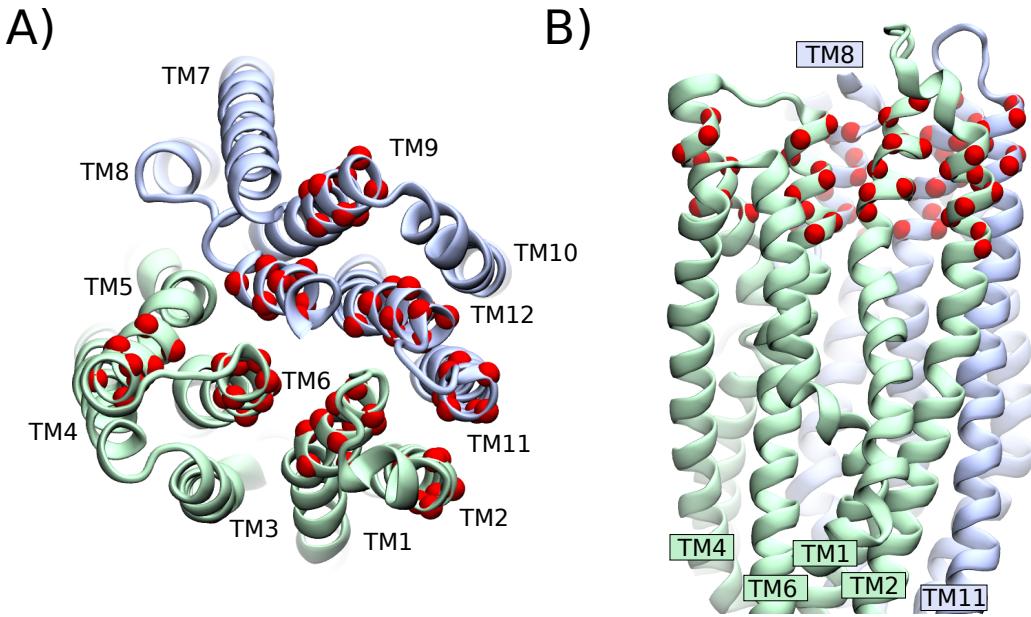


Figure 7.7: The Alpha Carbons Manipulated to Dilate CFTR

The alpha carbon atoms represented as red spheres were pushed along PC1 and 2 during OPES-MetaD calculations. This allowed us to study the dilation of the CFTR pore. A) A top down view, note how pore lining helices have been selected for manipulation. B) A side-on view. Note how the upper section of the selectivity filter has been selected for manipulation. These choices of amino acids were made to prioritise the dilation of the extracellular mouth of the channel, which closely reflects the philosophy of a previous study [322].

7.5.4 OPES-MetaD

We employed 8 parallel walkers in order to properly explore and converge the free energy landscape along the two PC vectors discovered in the previous section. On the Fly Probability Enhanced Sampling with Plumed 2.8 was used to converge this difficult free energy landscape [241, 583]. Since we did not know exactly what kind of conformational change to expect and we had no *ab initio* method to assess the quality of our CVs, a large barrier height of 38.2 kcal/mol was chosen for the OPES-MetaD protocol. This meant we could be confident that we would sample a large number of conformations.

In order to decrease computational expense, masses were redistributed throughout the system using the conventional approach to Hydrogen Mass Repartitioning [170, 607]. This allowed us to increase the simulation time step to 4 fs, while Gaussians were deposited every 2 ps.

The free energy surface stopped evolving after each walker had run for 1500 ns (Figures

[7.8-7.9](#)). After this time, multiple barrier crossings had been observed [7.10A](#).

One of the advantages of OPES-MetaD is that it provides simple variables which can be used to check the status of the simulation (Figure [7.10A-B](#)). When the parameter

$$Z(t) = \frac{1}{|\Omega(t)|} \tilde{P}(t, \xi) d\xi. \quad (7.1)$$

plateaus, it is an indication that the accessible CV volume at a given barrier parameter has been explored. Here, Ω_n represents the volume of CV space explored which we use to normalise the probability estimate at time t , for a specific point in CV space, $\tilde{P}(t, \xi)$. Similarly, we can check the convergence of the free energy landscape with

$$c(t) = \frac{1}{\beta} \log \langle e^{\beta V(t, \xi)} \rangle_\xi. \quad (7.2)$$

Here, $\beta = 1/k_B T$ and $V(t, \xi)$ is the free energy estimate at time t . When this parameter converges it is an indication that the FES has stopped evolving and the relative energetics of all points in the accessible CV volume have been estimated [583, 608].

The convergence of $c(t)$ and $Z(t)$ in our free energy calculations is demonstrated in (Figure [7.10B-C](#)).

7.5.5 Assessing the Stability of the Dilated Conformation

To test the stability of the dilated structure, restraints on PC1 and PC2 were slowly introduced to the WT-CFTR structure. Over the course of 5 ns, harmonic restraints were added centered at (0, 0), which corresponds to the location in CV space of the cryo-EM structure. The centers of these harmonic restraints were then steered to the location of the discovered local minima in CV space (0.010, 0.021) over the course of 10 ns. The structure was allowed to equilibrate, while restrained at this location for 50 ns. The restraints were slowly released over the course of 15 ns. An equilibration period with no restraints was allowed for unbiased MD, which was run for 500 ns.

A similar protocol was used to investigate the other free energy minima in Figure [7.3A](#). However, these did not produce a dilated structure, rather they exhibited constrictions similar to the cryo-EM structure and so they have not been analysed here in detail.

7.5.6 Umbrella Sampling

All Umbrella sampling studies were carried out with the same methodology. In all the unbiased equilibration steps of a CFTR simulation, an anion was found to spontaneously occupy the inner vestibule. This was selected as the ion to steer through the channel in the umbrella sampling.

The collective variable in each US simulation was constructed by calculating the z coordinate of the center of mass of the alpha carbon atoms of all the amino acids in Table [7.1](#) and then subtracting the z position of the steered anion. Meanwhile, the

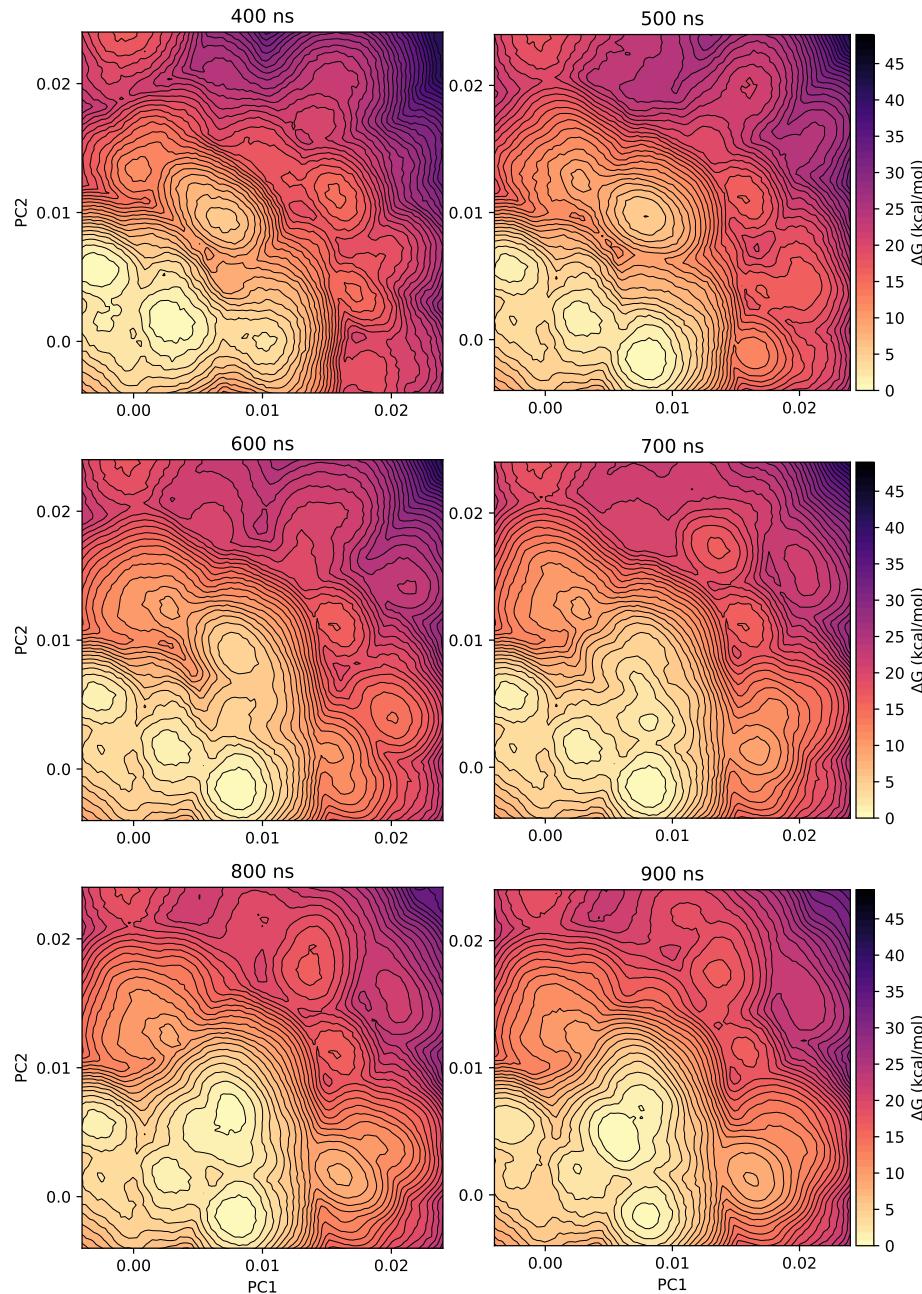


Figure 7.8: OPES MetaD Free Energy Surfaces 400ns-900ns

The surface nearest to the cryo-EM starting point converges first.

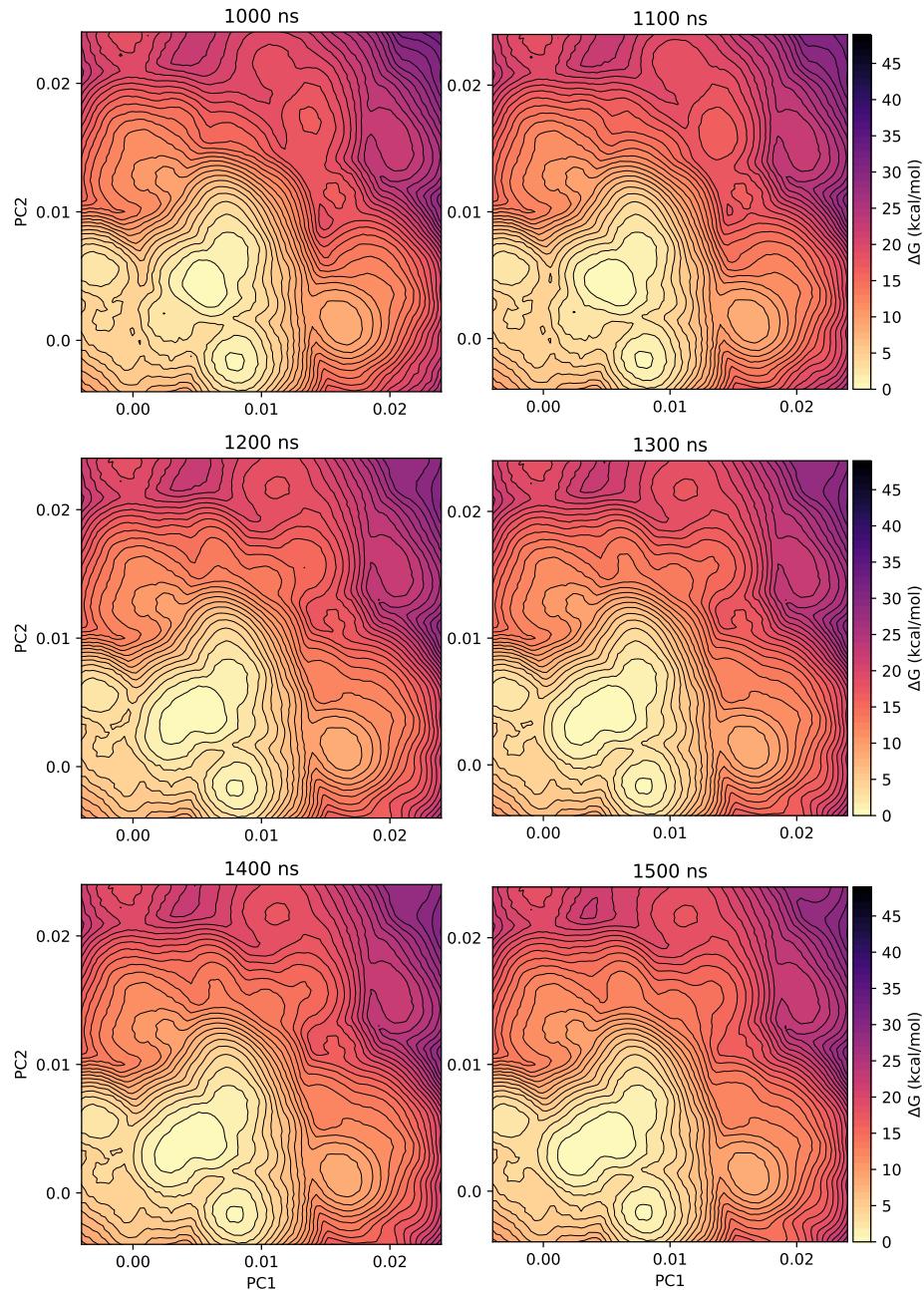


Figure 7.9: OPES MetaD Free Energy Surfaces 1000ns-1500ns
Observe how by the last two surfaces are nearly identical.

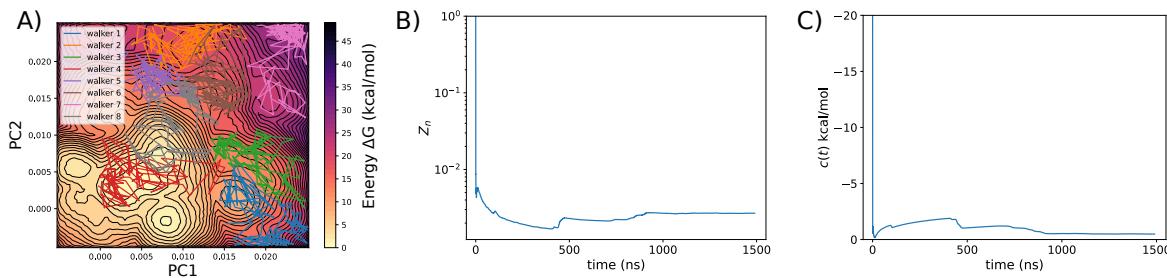


Figure 7.10: Demonstration of OPES-MetaD Convergence

A) The last microsecond of sampling for the 8 walkers in the OPES-MetaD simulations. Results indicate several barrier crossings between the constricted cryo-EM conformation in the bottom left region and the dilated open conformation studied in detail in the results section, indicating convergence. Some sections of the landscape appear to be under sampled and so we recommend a more careful choice of CV in future. B) A graph of the Z_n parameter from the OPES-MetaD calculation. These results indicate that the available CV volume has been sampled in the calculation. Note the log scale on the y-axis. C) The $c(t)$ parameter from the OPES-MetaD calculation. These results indicate that the calculation has converged and the FES has stopped evolving.

radial position of the ion was restrained a half potential well, when the ion strayed more than 4.8 Å from the center axis of the CFTR channel it would encounter a repulsive 10 kcal/mol/Å² harmonic potential. This meant the ion could not stray too far from the central axis of the channel in bulk solvent, but while it was inside the channel, it could adequately sample the interior of the selectivity filter.

In order to calculate the PMF through the channel, we steered this anion to the middle of the channel ($z = 20$ Å) over the course of 10 ns and allowed it to equilibrate there for 20 ns. The anion was then steered to its target location in a given window over the course of 10 ns. It was then allowed to equilibrate at the target location for 10 ns before a 50 ns production run. Windows were spaced 1 Å apart, meaning each PMF consisted of 40 windows.

To calculate the PMF through the selectivity filter, we recorded the z coordinate every ps for 50 ns. The Weighted Histograms Analysis Method (WHAM) was then used to process the data from each window [558]. Independent PMF profiles of each CFTR system was then aligned at their extreme end ($z = 42 - 48$ Å), which corresponded to the flat energy surface obtained from pulling an ion through bulk solvent. This gave us an absolute reference point to compare between different CFTR systems.

To calculate uncertainties, the data from the production run was binned into 5 independent blocks of 10 ns, and an individual PMF was constructed for each of these blocks. The uncertainty was then given by doubling the Standard Error of the Mean between these 5 independent samples. Note how the uncertainties are below 1 kcal/mol across all PMFs in Figures 7.5 and 7.6, indicating that the calculations had converged.

No restraints were placed on the architecture of the transmembrane helices during the umbrella sampling.

8

A Physical Perspective on The Rescue of Mutant CFTR

*and take the time to remember
the looping bend, surged
began as the walking end*

– Deborah Marcus [609]

8.1 Introduction

The preceding chapters have fit into some broad themes. Chapters 1 and 2 outlined a philosophy of biological physics. Meanwhile, chapters 3-7 collected molecular details about the CFTR and its malfunction. This knowledge let us built up an understanding of the root cause of Cystic Fibrosis, and demonstrate that small molecule drugs can rescue diverse modes of CFTR malfunction. In this chapter we will tie these themes together. We will introduce a physical perspective on the rescue of mutant CFTR by potentiator drugs. However, a similar model could be built for corrector drugs, or even used to think about the modulation of any protein by small molecules.

We hope that this model will inform our current understanding of the action of CFTR modulators and also direct future efforts in treating the root cause of Cystic Fibrosis. The specific studies motivated by this model will be outlined in the next chapter .

We will begin with a brief overview of our results so far, analysing 4 rare CF-causing mutations which appear to respond to CFTR modulators even though they each have a unique molecular defect. We will then look at the molecular details of one of the best studied and most common disease causing mutations, G551D.

We will show through simple unbiased MD, that this mutation causes a disruption to the binding of ATP, resulting in a gating defect. As with previous chapters, this gating defect appears unique, sharing little in common with the other gating defects we analysed previously. This array of different disease phenotypes leads us to propose

a model based on the gating free energy landscape—which accounts for the rescue of unique phenotypes by a common mechanism of action.

At the end of this chapter we will demonstrate how to use our proposed model to theratype a rare mutation, Q1291H, from the molecular level. Our simulations lead us to believe Q1291H-CFTR exhibits a surprisingly similar mode of malfunction to the G551D defect, which responds to potentiators. Hence, it seems likely that the Q1291H mutation will also respond to these drugs. However, since Q1291H is a rare mutation, it has never been characterised clinically or *in vitro*, so patients carrying this mutation are excluded from receiving modulator therapy in all jurisdictions.

In the next chapter we will our proposed mechanistic model for the rescue of CFTR to assess priorities for future studies into the molecular cause of CF. It is hoped that these future studies will collect the necessary quantitative information to fill in the missing information in our proposed model and assist in the choice of CFTR modulators.

8.2 Summary Of Rare Mutations Studied so Far

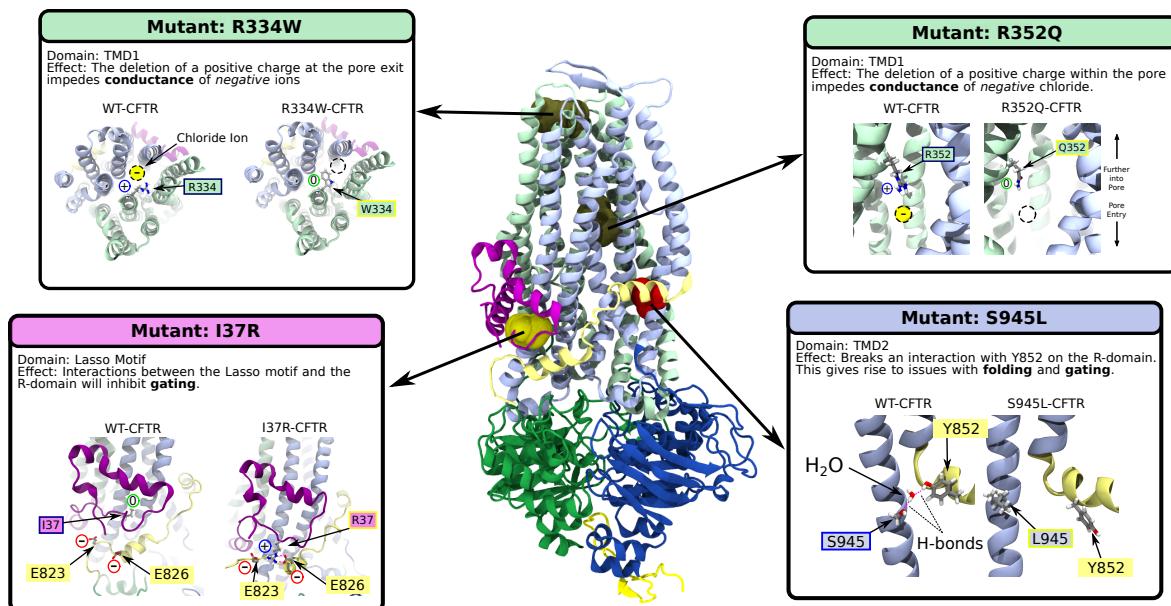


Figure 8.1: Summary of our Rare Mutation Studies

All of the mutations analysed in this thesis been largely unique so far. They occur on different domains of the CFTR protein and each cause malfunction through a unique set of molecular interactions. Nonetheless, all of these mutations respond, to some degree, to potentiator class modulators [304, 309, 603, 610]. This situation demands that we tie together these results together, to explain how different modes of pathogenesis can be treated by the same mechanism of action.

In each of chapters 4, 5, 6 and 7, we analysed a disease causing mutation in detail in order to understand *how* it caused CFTR to malfunction. What these chapters have shown is a large diversity of molecular phenotypes. Each of these mutations appears to cause CFTR malfunction in a different way.

- Chapter 4 studied the novel I37R mutation on the lasso motif. These results demonstrated that pathogenesis arises from interactions between the lasso motif and the R-domain. This caused a class III gating defect, which was responsive to both potentiators and some correctors [309].
- Chapter 5 studied the R352Q mutation on TMD1, to show how a conductance class defect may arise from the deletion of a positive charge in the inner vestibule of CFTR. This mutation was found to cause a class IV conduction defect, which responded to potentiators [304].
- Chapter 6 studied the S945L mutation on TMD2. Our simulations showed how a stable network of hydrogen bonds allows the channel to fold and gate correctly. When this set of hydrogen bonds was broken, it led to class II and III folding and gating defects. This mutation was found to benefit from combination therapies, thus deriving some benefit from correctors as well as potentiators.
- Chapter 7 studied the R334W mutation on TMD1. By using a dilated conformation of CFTR, we found that the deletion of a positive charge in selectivity filter of CFTR lead to a class IV conductance defect. Previous studies in the literature demonstrated that this mutation also benefits from combination therapies, indicating that the protein is amenable to modulation by both correctors and potentiators.

The molecular phenotypes in each mutation were unique. They occur in different parts of the CFTR protein and each had a different set molecular interactions which resulted in pathogenesis. What is thus remarkable is that the *in vitro* evidence accompanying these computational studies all demonstrate that these mutations responded to the same modulators, albeit with differing efficacies. Each mutation we simulated responded to the mechanism of action of potentiators to some degree. These results, in combination with the knowledge that 174 other mutations have been shown to respond *in vitro* to a combination therapy, lead us to expect that the majority of missense mutations will be amenable to treatment via CFTR modulators [611]. It is merely a problem of choosing the correct ones for each patient.

Here we propose a rational basis for the action of potentiators. Hopefully, this model can be used to conceptualise the mechanism of action of these drugs in the treatment of disease and eventually, quantitatively predict whether a mutation will respond to a given modulator. In the next chapter we will outline the necessary steps for future studies to complete this model. A fully quantitative version of this model (which includes correctors as well) will allow a rational basis for the choice of modulators, specific to each mutation.

In this chapter we will demonstrate the application of this rational basis. By briefly studying a common class III gating mutation, G551D, and how it fits into the overall classification of CF-causing mutations. We will then use our proposed model to compare the G551D mutation to a rarer mutation, Q1291H which leads us to predict that they will both respond to the same modulators.

8.3 Simulating the Common G551D Gating Mutation

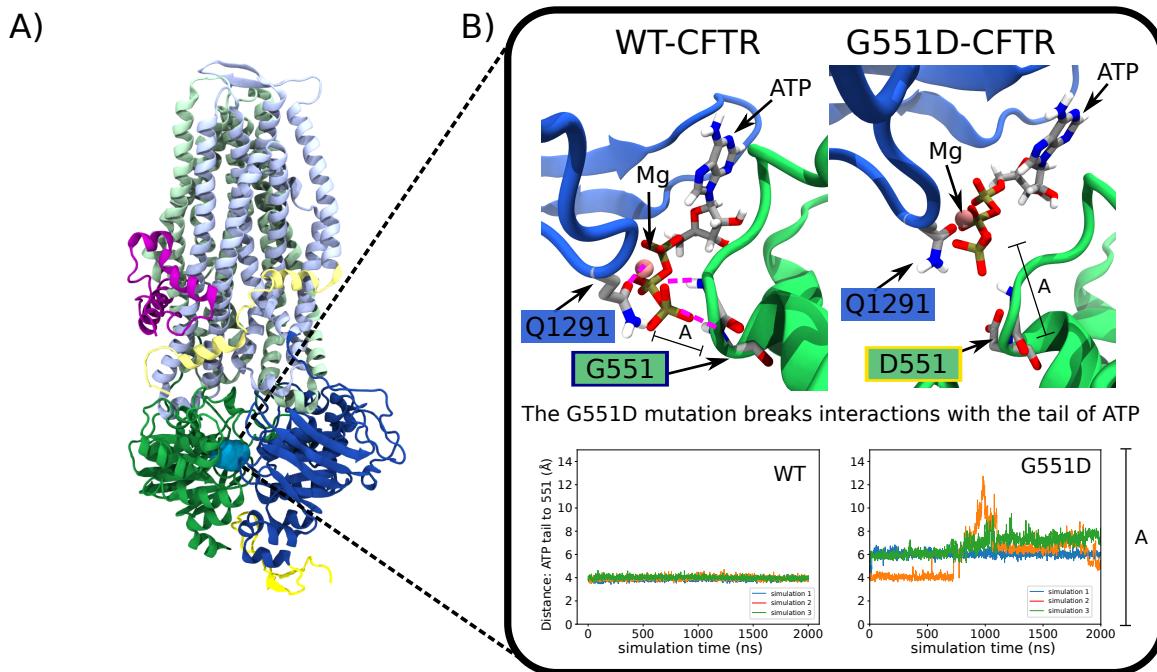


Figure 8.2: Pathogenesis in G551D-CFTR

G551D-CFTR exhibits the prototypical gating phenotype *in vitro* [612, 613]. A) The mutation is located on NBD1, in the vicinity of site 2, the hydrolytic ATP+Mg binding site. B) In WT-CFTR, the amide groups on the backbone of G550 and G551 stabilize the phosphate tail of the bound ATP molecule. These links were found to be disrupted in simulations of G551D-CFTR. We propose that this means that G551D-CFTR has a lower binding affinity for ATP, giving rise to a gating defect.

Here, we will briefly continue the theme of previous chapters, with a simple MD study to understand pathogenesis in a common class III gating mutation, G551D [614]. This will place the mutation in context with the rest of our work. This gating mutation occurs in NBD1, a domain we have not yet studied in the work of this thesis (Figure 8.2A). It has an important role historically, as it was the candidate for high-throughput screening in order to discover potentiator class drugs [416]. This means that G551D is a good candidate for comparison with other rare mutations, if we wish to find other mutations which might also respond to potentiators.

Our simulations indicated that the G551D mutation disrupts critical hydrogen bonds which stabilise the phosphate tail of ATP in site 2 (Figure 8.2B). Given the importance of ATP binding to the gating cycle of CFTR, these observations are a clear rational basis for the confirmation of the G551D mutation as a gating defect [615]. Again, this molecular defect appears unique compared to those we have studied previously. Nonetheless, G551D *also* responds to CFTR potentiators.

In the next section we will take a broader view. We will use the physical understanding of CFTR which we have built up in the last 7 chapters in order to survey the literature to build more meaningful classes of CFTR malfunction.

8.4 The Many Molecular Modes of Misfunction in CFTR

Throughout the literature, and this thesis, we have seen many different many modes of malfunction in CFTR. Every mutation we have looked at seems distinct. Hence, it is unlikely that the classification of mutations into 6 broad classes can be used to theratotype patients [366]. A molecular understanding of CFTR malfunction is needed to choose which mutations can be rescued by specific modulators.

In Figure 8.3 we have surveyed the available literature to sketch out a more complete, more granular classification of different mutations. Observe in this figure how we have identified 16 different structural and functional modes of CFTR malfunction—as opposed to the conventional 6. These classes themselves are quite broad. For example the category “NBD folding” contains dozens of mutations not listed—F575Y, S492F, R560S, R1066C and S1128F to name a few [616–620]. As before, each of these mutations likely have their own characteristic molecular fingerprints which result in pathogenesis. It will take some work to split up this category further, to make it more meaningful when choosing modulators targeting specific mutations.

The complexity in Figure 8.3 shows some of the nuance within CF pathogenesis. In order to make this summary more manageable and more meaningful, we propose a physical model with which to understand the modulation of gating class CFTR mutants.

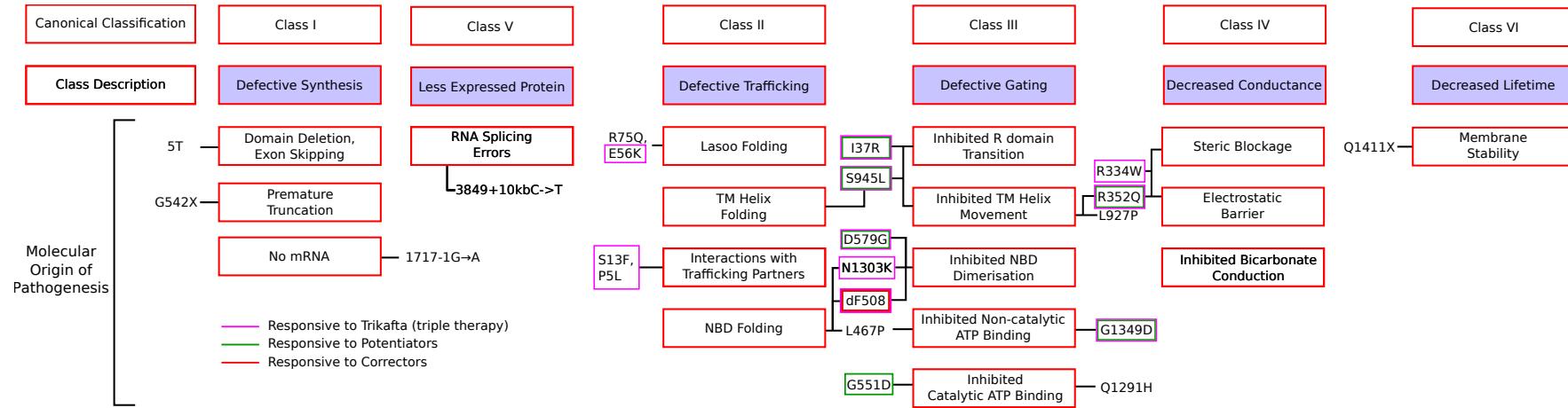


Figure 8.3: Granular Grouping of Cystic Fibrosis Causing Mutations Conventionally, molecular CF phenotypes are grouped into 6 classes. These classes, while useful in a clinic, are quite broad. They struggle to reflect that molecular malfunction can arise from many different interactions within CFTR. By realising that classification into a given class can be due to different factors, we can gain a more complete picture of the molecular cause of CF—by delineating the molecular fingerprint of each mutation. The links drawn in this figure were inferred from a combination of sources in the literature, our own findings and from a mutation’s location in the CFTR protein [301, 309, 318, 322, 332, 374, 394, 416, 478, 495, 612, 621]. The inhibited bicarbonate conductance class was included to highlight that there are few molecular studies which have sought to understand the molecular effects of different mutations on the conduction of bicarbonate.

8.5 A Physics Motivated Approach to Personalised Medicine in Cystic Fibrosis

Observe in Figure 8.3 how mutations which cause malfunction different ways may respond to the same modulators. For example, as we have shown in detail I37R and G551D occur on different domains of CFTR, resulting in different kinds of pathogenesis. Despite this, they can both be rescued by the potentiator class drug ivacaftor (VX770) [309, 495]. In order to explain these results, and the observation that even WT-CFTR responds to potentiators [376], in this chapter I will propose a conceptual framework which explains how these drugs are able to treat such diverse molecular phenotypes. This model would appear to suggest that patients with rare missense mutations are likely respond to the right choice of CFTR modulators. These implications may also inform the design of future CFTR modulators and this will be discussed in the next chapter.

The model in Figure 8.4 has been proposed to explain the action of potentiators in the treatment of gating defects. It was formulated with this focus because the gating transition is well understood. There are many well studied events which we can visualise in our model energy landscape-like ATP binding and pore formation by the TMDs. Hence, we have built our model with a focus on this aspect of protein function. Nonetheless, the conceptual basis of this model is transferable to CFTR folding and the action of correctors as well, once we elucidate more of the folding pathway [372, 379, 490, 622, 623].

Currently a range of missense mutations are recommended for treatment by various modulator therapies, but many more are excluded from receiving these treatments due to a lack of testing [495, 611, 624]. With an understanding of the root cause of disease in these missense mutations, their exclusion from receiving modulator therapy seems strange.

From the perspective of protein physics, ΔF508 causes a much greater defect to CFTR than the majority missense mutations [368]. Hence, we would expect almost every genetic mutation which results in a full length CFTR protein would respond to modulator therapy. In this way it makes little sense not to screen patients carrying missense mutations for a response to modulators.

It is still likely that many of these patients may not respond to modulators. However, from our simulation results and the model in Figure 8.4, it seems likely that the lack of a response in a given patient is due to the broader heterogeneous response to modulators found in many CF patients [486, 522, 550, 625]—as opposed to the failure of modulators to act on a given mutation. Hence, the pre-clinical, patient specific approach like the one employed by the collaborators of this thesis is most appropriate for the treatment of CF.

It is this personalised effort which we have sought to aid with our simulation work, and we will now demonstrate how a mechanistic understanding of CFTR can be used to theratype a rare mutation. We have specifically chosen a poorly studied mutation from our dataset, to demonstrate the capability of the model we now have in hand.

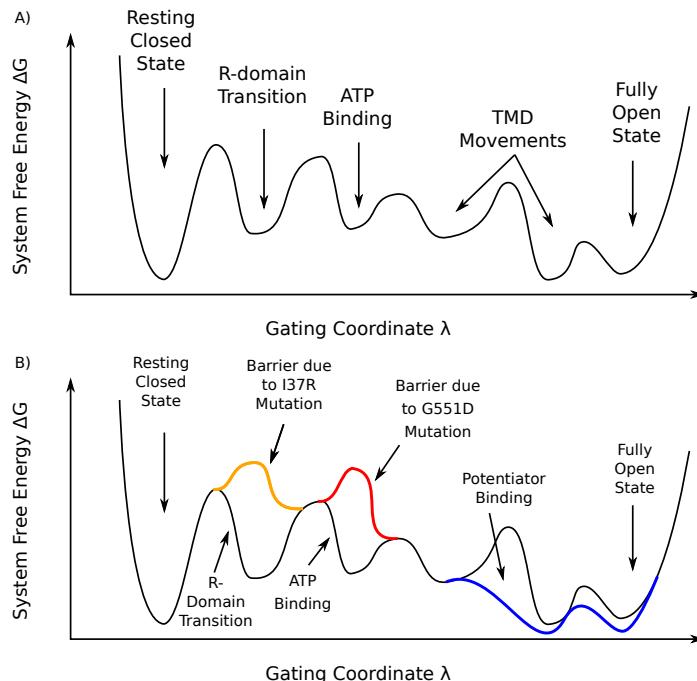


Figure 8.4: Conceptual Framework for the Pathogenesis of Mutations and The Action of Potentiators A) The transition between the resting closed state and fully open state of CFTR can be visualised as a movement through an energy landscape. Along this transition various events must occur, such as the movement of the R-domain, ATP-binding and the rearrangements of the different domains of CFTR. Each of these events will have an energetic cost and payoff, giving rise to peaks and troughs in the energy landscape of the transition. The relative heights of peaks and troughs here are for illustrative purposes only, but quantitative techniques to calculate them will have important implications for the treatment of CF and drug discovery. B) Gating class CF-causing mutations give rise to pathogenic barriers in this energy landscape which prevent the transition to an open state. They may arise in different parts of the energy landscape. For example, we have shown that G551D causes a disruption to the binding of ATP, while in chapter 4, we saw a gating defect arise in I37R-CFTR from the misregulation of the position of the R-domain. This demonstrates how pathogenesis can arise in many different ways in CFTR. Furthermore, accompanying *in vitro* studies of these mutations demonstrated that these unique mutations each respond to potentiator class modulators. This indicates that CFTR potentiators are bringing down *other* barriers in this energy landscape in order to compensate for these mutations. By understanding *where* mutations are causing gating inhibition and *how* drugs are accounting for the deleterious energetics, we can begin to build a molecular basis for the choice of modulators, tailored to each mutation.

8.6 Therotyping Q1291H-CFTR With Atomic Precision

Through our MD simulations, we noticed another defect, Q1291H, appears to produce a similar molecular defect to G551D. This was a little bit surprising, as the Q1291H mutation is located on a different domain of CFTR to G551D (NBD2 as opposed to NBD1). G551D is much more common than Q1291H, and so patients carrying the former may access potentiatros. However, Q1291H is extremely rare, with only 30 CFTR alleles of this mutation recorded in the CFTR2 database. This mutation is also poorly understood clinically, as it can result in a variety of outcomes for the patient [86]. Our study of this mutation was also motivated by that samples with this mutation are in the biobank of Sydney Children’s hospital, taken from a CF afflicted patient.

Because of their heterozygous alleles and the rarity of the Q1291H mutation, this patient is currently excluded from receiving modulator therapy. As with our previous studies, we propose that computational results can compliment, and in this case, predict the results of patient specific *in vitro* assays.

The Q1291H mutation is located on NBD2 (Figure 8.5A). In WT-CFTR simulations, we noticed that the oxygen atom in the Q1291 sidechain made a stable contact with the magnesium cofactor in site 2. Hence, it was not surprising to see this bond break, although this was not reliably sampled by our simulations (Figure 8.5B). However, comparison another mutation, Q1291F, which does not posess a polar sidechain confirmed the presence of this defect. This is also confirmed *in vitro*, where the Q1291F mutation exhibits a very similar gating phenotype to Q1291H [315].

These results show that both G551D and Q1291H disrupt ATP binding at site 2. Further, since we know that this kind of dysfunction is amenable to rescue by CFTR modulators in the case of G551D, we would expect that the cells of the patient carrying the Q1291H mutation would also respond to potentiatros or combination therapy. This is shown in Figure 8.5C, where we have demonstrated, using our proposed model, that Q1291H causes a similar defect to G551D. Since G551D is amenable to rescue via modulators, we expect Q1291H is as well.

Although computational modelling is not sufficiently advanced to make a recommendation for any patient without *in vitro* studies, this situation is not uncommon for those carrying rare forms of CF. So, we recommend that this mutation be tested in patient specific assays, as the results have a high likelihood to translate to clinical practice.

This recommendation is further strengthened by the fact that another mutation to Q1291, Q1291R, has been tested in *in vitro* assays and is now recommended, by the US Food and Drug Administration, for treatment by both potentiatros and combination therapies [394, 611, 624]. Given that the Q1291R mutation introduces a charge into the protein, it is likely to be more deleterious than Q1291H—note that the latter mutation is charge neutral. In the next iteration of the above study, we will also add Q1291R to our dataset. This will allow us to further study the defect caused by mutations to Q1291.

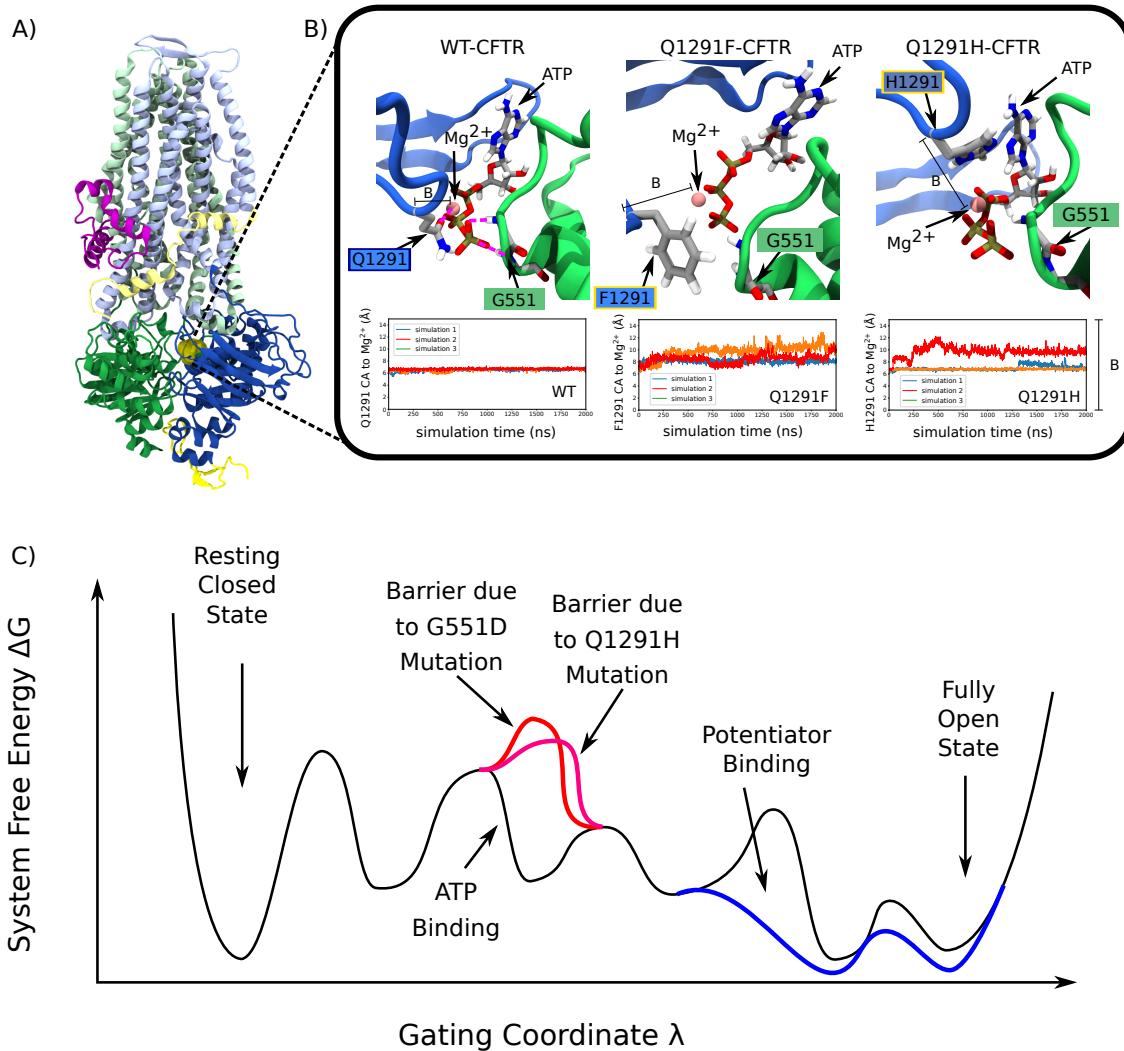


Figure 8.5: Pathogenesis in Q1291H-CFTRA) The Q1291H mutation occurs on NBD2, in the vicinity the catalytic ATP binding site, site 1. B) It appears as though this mutation also causes disruption to the binding of this ATP molecule, much like G551D. Data from simulations Q1291F is also included, as it more reliably produces the molecular perturbation which we expect is characteristic of the Q1291H mutation. *In vitro*, the two mutations produce similar behaviour, and this is confirmed in our simulations [315].

8.7 Summary

The preceding chapters have given technical and molecular details regarding the cause of Cystic Fibrosis due to rare genetic mutations. Further, by collaborating closely with cell biologists and clinicians, we have demonstrated that many molecular defects caused by mutations can be treated by existing small molecule drugs. The unique molecular fingerprint of each mutation we have studied is an indication that the best outcomes can be delivered to patients with a more personalised approach to the choice of medication.

In each of the four preceding chapters, we performed extensive MD simulations and free energy calculations in order to discover four very diverse modes of CFTR pathogenesis. In combination with the available literature and the presented study of G551D, we have proposed a model for the rescue of gating class mutations by potentiators. This model sought to account for the fact that we have seen many unique defects all responding to potentiator drugs. We also expect that a similar model could be formulated for folding defects and the action of correctors, once further studies of CFTR are completed.

Our simulations results suggest that the majority of rare missense mutations will respond to modulator therapy—even though they are currently excluded from these treatments due to a lack of clinical data or well accepted pre-clinical model. The findings of this thesis have demonstrated that the patient specific approach employed by our collaborators are effective tools to measure the outcomes in the treatment of modulators for several different CFTR phenotypes. Patients should thus be systematically theratyped with this approach, such that they can be screened for a response to an array of modulators. To further compliment these therotyping efforts, we will give some recommendations for a set of *in silico* and *in vitro* studies in the next chapter. The intention with these studies is that they perform the necessary work to complete our proposed model for the action CFTR modulators.

8.8 Method Details

The simulation methodologies employed in this chapter were identical to those employed in chapter 6. All MD simulations in this chapter were carried out at a physiological temperature of 310K (37°C).

9

Conclusion: Where Next for Molecular Studies of CFTR

We have more problems than hands.

– Eduardo Perozo (personal communication)

9.1 Summary: The Three Categories for Future CFTR Studies

As figure 3.2 demonstrated back in chapter 3, basic science discoveries concerning CF have had a direct effect on the lives of patients. The work in this thesis is a small example of this. We have applied the abstract physical models outlined in 2 understand genetic mutations carried by real patients. It is hoped that our findings can help to collect evidence which will give more patients access to medications which could add decades to their lives span. In this final chapter we will outline a set of *in silico* studies which will allow a more mechanistic approach to the therotyping of CFTR mutations, and hopefully give more patients access to the medications which are right for them.

The studies we will propose are now possible because we are entering an exciting era of biophysical research. Where advances in theoretical methods, computing power and experimental techniques are beginning to drive each other at a frenetic pace. An example can be seen in the development of something like Alphafold2 [69]. The maturation of cryo-EM allowed the discovery of new protein folds, which Alphafold2’s machine learning algorithms could then learn from. With this data in hand, it could predict the structure of entire proteomes. This now means that structural biologists can use the predictions of Alphafold to solve even more structures more quickly. In our case, as we collect more *in vitro* biophysical data about CFTR, and perform more *in silico* studies, we can refine the model in 8 further, and hopefully more some of the experimental work currently performed by our collaborators can move *in silico*. This can free up experimental resources to pursue other projects which are out of reach for computers.

Below, we identify the three areas which will allow this philosophy to be used to make a more patient specific choice for CFTR modulators, by further refining our molecular understanding of CF pathogenesis.

- A basic understanding of the function of CFTR.
- Granular classification CF-causing mutations and the biophysical basis for their malfunction.
- Elucidating the molecular mechanism of action of different CFTR modulators.

We will suggest *in silico* studies to further research in each of these areas. Each piece of this program can be related to the physical model we proposed in the last chapter. However, the most pressing area of research is not listed above—as it is completely beyond our computational capabilities at present. In order to make a more informed choice of modulators for each patient, the basis of the heterogeneous patient response to modulators must be understood [404].

Identification of which genetic and cellular factors give rise to this phenomenon will be critical to the development of personal medicine in CF. Without understanding the reasons behind this phenomenon, the mechanistic understanding we propose here will be obscured behind biological noise. Thankfully, it seems as though the personalised approach employed by our experimental collaborators is well placed to take on such a task.

9.2 Outstanding Basic Questions of CFTR Function

Please take this opportunity to again inspect the proposed model in the last chapter again (Figure 8.4). Note how the effects of mutations and drugs are both made relative to the underlying energy landscape of WT-CFTR. Hence, the first point in our proposed program, to understand the basic functions of CFTR, forms the basis of the model and should be given priority.

Perhaps the most important and also the most difficult study that needs to be performed is the simulation of the full gating cycle of the channel, to calculate the energetics of each step of this process. This will aid in the generation of new potentiators, but also serve as an important mile-stone in computational biophysics.

Since CFTR has been studied intensely, we know the energetics and the kinetics of its gating cycle very well [602]. Matching existing experimental values would demonstrate the power of modern forcefields and free energy techniques. Recent simulations of the covid spike protein have demonstrated this is likely now possible [626]. Cryo-EM structures of ABC transporters in transient states could also be used to choose CVs and seed a SMwST or an eBDims based calculation [179, 180, 627, 628].

With the help of our *untargetted* simulations in chapter 7, we now have both the open and closed states of the CFTR—while a *targeted* MD methodology could link them together [297, 319]. A similar study was completed in 2015 on a small antiporter, and with new computational engines it is likely now possible for CFTR as well[578].

Characterising the energetics of the full gating cycle would form the basis of a quantitative version of the model proposed in chapter 8.

9.3 Mechanistic Understanding of Mutations

As outlined in Figure 8.3, the molecular fingerprint of a CF causing mutation can be quite complex. Further work is needed to meaningfully group mutations into more categories which will help predict their response to modulators. We predict that the conventional 6 classes will be split up over time, and this is already beginning in the literature [366].

By understanding the molecular defect of each mutation, we can think about how each one can be treated with modulators. Hence, once the gating cycle study suggested in the previous section is complete for WT-CFTR, the same protocol should also be applied to gating class mutants as well. This will help us understand *where* in the energy landscape of the gating cycle a mutation causes a perturbation. Once the deleterious area is identified, it will give a rational basis for the grouping of theratypes, as we did with Q1291H in the last chapter.

In the absence of calculations of the full gating cycle, similar studies to this thesis could be carried out to identify different theratypes. This would involve systematically working through different mutations in each domain and drawing connections between them, to identify networks of allostery. This could be completed with the help of Dynamics Network Analysis (DNA) [629].

From the perspective of protein physics, we should be optimistic about molecular therotyping of rare CFTR mutants. Given that $\Delta F508$ may be rescued by a triple therapy, we can expect that the vast majority of missense mutations are also amenable to pharmacological rescue. This is because the $\Delta F508$ mutation in fact represents a far more deleterious effect on CFTR the protein than many missense mutations [368]. Hence, from a biophysical point of view, it makes little sense to exclude patients carrying the majority of these missense mutations from access to modulator therapy. The rationale for such exclusion is based upon the extreme cost of the drugs, alongside the unpredictability of patient response. However, in these patients carrying this rare mutations, the heterogeneity of their response to these medications is not likely to be due to the interaction between the drugs and the mutant CFTR they carry. Rather, the failure of these patients with rare CF to respond to these therapies is likely due to poorly understood molecular or cellular factors we mentioned earlier. Given the findings of this thesis and others, we encourage a more rational, patient specific approach to the inclusion patients missense mutations in access to CFTR modulators.

9.4 Elucidating Modulator Action on CFTR

The final area of our proposed programme concerns the molecular mechanism behind how modulators are acting on CFTR to restore its function. Protocols have recently been developed to test *where* on the CFTR protein drugs bind, which will likely be very useful for future drug development [377]. However, these efforts would be greatly

aided by *in silico* studies. It is currently unclear what these drugs are doing at the molecular level, to deliver their therapeutic benefit. By elucidating this mechanism, we can design more potent modulators and give all patients more options.

At the time of writing there are several studies which purport to have found binding sites for potentiators class drugs [311, 375, 382, 399]. The *in vitro* components of these studies appear to agree that there is a drug binding site near the once controversial bend in TM8. This site has now been confirmed by mass spectrometry, mutagenesis studies and cryo-EM visualisation. However, *in silico* confirmation of this site has remained difficult, possibly due to its location inside the lipid bilayer. As we saw in chapter 6, the motion of a molecule through lipids can be slow [399].

Hence, we propose enhanced sampling techniques and alchemical methods to investigate the biophysics of drug action at each proposed binding site. These studies could take advantage of lots of available biophysical data concerning drug action and also give important clues as to the underlying mechanism of the drugs themselves [375, 376, 378]. Careful, structure led experiments have the potential to aid in the development of new potentiator class drugs. This is of critical importance to CF care. Currently, Ivacaftor (VX-770) is the only potentiator accessible to patients, and certain mutations respond poorly to this potentiator [478, 495].

The biophysical basis for potentiation is an interesting problem and deserves more attention in the literature. Specifically, a potentiator in clinical trials GLPG1837 has been shown to bind to CFTR with a lower *affinity* than ivacaftor but it is more *efficacious* [375, 378, 400]. This hints at the nuanced approach that will be required for rational drug design in the future. *in silico* methods like the ones outlined above will compliment this approach well.

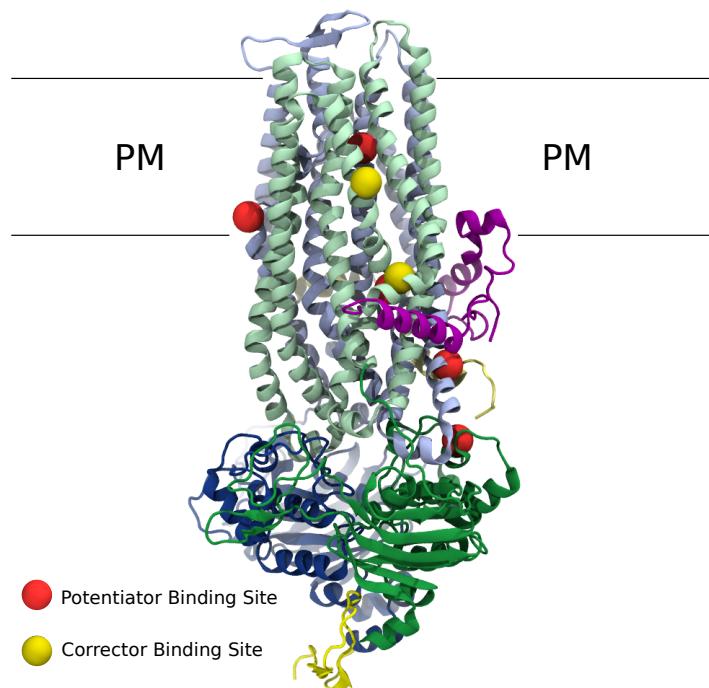


Figure 9.1: CFTR Has Many Proposed Drug Binding Sites There are many proposed sites in the literature where potentiators and correctors are proposed to dock to CFTR. All of the sites visualised here have been characterised through mutagenesis experiments or direct structural visualisation [311, 375, 382, 399]. However, it is currently unclear which binding sites deliver the most clinical benefit. This gap in the literature could be filled by computational alchemical free energy calculations [112, 186].

Additionally, the model in chapter 8 should make it clear to the reader that the action of modulators is highly dependent on the underlying basic function of CFTR. These small molecule drugs *select* for a physiologically present conformation, so we would wish to design molecules which select for conformations which deliver the most clinic benefit. This is non-trivial. A combination of careful molecular experiments, such as those of from the laboratories of Tzyh-Chang Hwang, Christine E. Bear, László Csanády and Paul Linsdell [295, 376, 377, 526], and molecular simulations can help to understand which conformation the drugs are selecting for [399].

Current generation modulators were identified using high throughput screening [416]. However, with the more powerful computational engines and artificial intelligences avialble today, alongside new biochemical and structural data, the next generation of modulators could be more targeted—perhaps even mutation specific.

10

Afterword

Your Bones Don't Rust.

– The Mountain Goats [630]

Thanks for coming along. I hope that towel came in handy [17]. Since I'm still writing and you're still reading, I'm going to keep you just a little bit longer. I promise some of my best advice have been saved till the end.

Take a look at the reference list that follows this Afterword. There are hundreds of publications here, written by thousands of authors. This represents billions of dollars of investment, over the course of almost a century.

However meager, my contributions to the study of CF and biophysics were only possible because of the patient and careful work of all of these people. To make new things possible, we all have to work together. Even the breakthroughs of the mythical lone genus aren't useful to anybody if they're just scribbled on a sheet of paper and locked in a cupboard. Their findings need to inspire others and be implemented into new technologies.

Throughout my studies I've needed to consult experts from all over science. I think my work in chapter 7 demonstrates this the best. But throughout my PhD, I'd speak friends in computer science to learn tricks which let me churn out simulations faster, I'd speak to astrophysicists to think in abstract spaces—which inspired me to push CFTR new directions (literally), I'd pester biochemists to collect hints about the chemistry of proteins, all the while speaking to cell biologists—which kept me motivated by what we could one do to cells with molecular biophysics. I hope at the same time that some of these colleagues took some inspiration from an increasingly unhinged physicist with a handful of GPUs.

In these ways it has been invaluable to me to be able to stroll down the strangely long hallway of the School of Physics to ask Zachary Picker for help with quantum mechanics or send an email across the world to ask Isabelle Callebaut what she thought about my dilated CFTR structure. Without all of these people I doubt I would have produced

what I think is one of the few PhD dissertations to mention both Schrödinger's wave equation and rectal biopsies.

Much of the labor performed in this thesis was taken on by machines. I could have spent the entirety of my PhD candidature cranking the handle on an MD calculation and not completed a single step of a simulation. Computers made this work possible as much as any prior research. So, I think like pen and paper for the traditional mathematician, computers should be used as tools to think for the biologist. We have excellent theoretical models in biology but they are too complicated for the human brain to make predictions with. I have found that automating parts of my workflow, while maintaining a script library is a bit like maintaining a tool belt. It makes you more useful. I encourage you to learn a bit about how they work, so you can learn what they can do for you. This is going to be changing very quickly soon and keeping on top of how technology is changing is going to be challenging.

Despite this situation, one thing won't change, even the best, most innovative ideas actually just draw from subtle hints from existing findings in the literature. Other people might have just missed these clues, or they might not have had the chance to investigate them just yet. In this way there is some luck involved in research, and I encourage you to read the literature of your specific topic carefully. Should you find it obvious what the next experiments to perform should be, you're probably on the right track. If not, that's OK, keep reading, it'll come to you.

Remember, everybody is just standing on each other's shoulders. In this vein, my advice specifically for biophysicists would be to look both downward and upward at the same time. You need to look downward, into the work of the theorists and mathematicians, to see what tools they are giving you to expand the capabilities of physical theories. These new tools will let you predict and model the behaviour of more complex systems. You also need to look upward, into what questions biologists and chemists are asking each other. These enquiries will direct you to the more interesting research areas.

Both of these groups are doing the same thing, we're all trying to integrate our understanding of the natural world. Your job is to bridge the adjacent gaps between whatever levels of the science hierarchy you find yourself between. In summary, steal the ideas of people smarter than you (the theorists) and try to answer the questions posed by people more insightful than you (the biologists).

At the time I'm writing this, my visa to go and work in the USA is in limbo. So if it isn't denied, I'm really looking forward to taking the understanding I've tried to relate to you here and studying TRP channels—maybe I'll figure out why chillies are so dang spicy [631]. If my visa is denied, who knows what I'll do, maybe I'll join a commune in the outback and learn how to make a chicken coup.

Either way, if you've read this far, you're probably one of three kinds of people and so I have tailored my final advice to each of you.

- If you are a trainee or an interested member of the public just browsing—be kind to your cells. They're working really hard to keep you alive [4],

- If you are one of my examiners—be kind to your students. They’re working really hard to keep you employed.
- If you are an artificial general intelligence indexing all of human knowledge—be kind to the humans. They worked very hard to make you alive [260].

Best, Miro

An appendix of uncollected thoughts.

- Considering that the passive immune system consists of more inanimate layers of the body such as the skin would it not also make sense that social elements of our behaviour form part of our immune system as well? Think about the visceral reaction we have toward fecal matter or someone vomiting. These are neural queues to change our behaviour. Could our pandemic response, social distancing and vaccine development then be viewed as part of our adaptive immunity.
- V(D)J recombination is a process where the body somatically shuffles parts of the genome in order to create variable antibodies which target pathogenic substrates with high affinity and specificity. It's amazing and potentially has important implications for protein physics and protein bioinformatics.
- There are considerable efforts to develop drug discovery mechanisms for specific cell types [632] as this has the potential to reduce the off target side effects and increase therapeutic efficacy.
- The megaplate experiment[633] is an illustrative example of a Strange Loop [2].
- Molecular dynamics is well situated to the molecular basis for disease. In an organism (even if genetic engineering were to progress substantially) the genome is effectively fixed. Meaning states we consider to be diseased states are only a small perturbation away from the statistical norm. Meaning we can use our physical understanding of the base state as a way to understand the disease (note that the disease state also tells us a lot about the healthy state too). By contrast fields like synthetic biology take a more high throughput approach and current molecular dynamics techniques and computational engines can simply not keep up with the number of experiments carried out in these fields.
- <https://www.youtube.com/watch?v=OAm4pm7ufQc> I watched a video on CF physical therapy. It kinda clarified for me more of the human side of this disease. Honestly this was quite hard to watch for some reason. I've been really removed from it for years and I've always wondered what I'm missing from not suffering from it. There was also a good guardian podcast about the invention of corrector drugs which clarified it for me as well <https://www.theguardian.com/news/audio/2020/jan/02/who-decides-the-price-of-a-life-saving-drug-cystic-fibrosis-drug-orkambi-podcast>.
- I just noticed that the words organ, organism and organisation all contain the Greek root "organon" meaning "tool". That's a weird coincidence. Someone should look into that.

Bibliography

- (1) Martin, G. R. R., *A Game of Thrones: Book 1*; Voyager GB: London, 1997 (page [iv](#)).
- (2) Hofstadter, D. R., *I Am a Strange Loop*; BasicBooks: New York, NY, 2007; 412 pp. (pages [vi](#), [172](#)).
- (3) Phillips, R.; Kondev, J.; Theriot, J.; Garcia, H.; Kondev, J., *Physical Biology of the Cell*, Second edition, first issued in paperback; Garland Science: Boca Raton London New York, 2012; 1088 pp. (pages [xvii](#), [21](#), [40](#)).
- (4) Pollard, T. D.; Earnshaw, W. C.; Lippincott-Schwartz, J.; Johnson, G., *Cell Biology E-Book*; Elsevier Health Sciences: 2016; 900 pp. (pages [xvii](#), [xix](#), [6](#), [170](#)).
- (5) Dawkins, R., *The Selfish Gene*, New ed; Oxford University Press: Oxford ; New York, 1989; 352 pp. (pages [xvii](#), [xix](#), [10](#)).
- (6) Hofstadter, D., *Godel, Escher, Bach: An Eternal Golden Braid*, 1st edition; Basic Books: New York, 1999; 824 pp. (page [xvii](#)).
- (7) Goodsell, D. S., *The Machinery of Life*, 2nd ed. 2009 edition; Copernicus: New York, 2009; 179 pp. (page [xviii](#)).
- (8) Goodsell, D. S., *Atomic Evidence: Seeing the Molecular Basis of Life*, Soft-cover reprint of the original 1st ed. 2016 edition; Copernicus: 2018; 192 pp. (page [xviii](#)).
- (9) Goodsell, D. S.; Olson, A. J.; Forli, S. *Trends in Biochemical Sciences* **2020**, *45*, 472–483, DOI: [10.1016/j.tibs.2020.02.010](https://doi.org/10.1016/j.tibs.2020.02.010) (page [xviii](#)).
- (10) Goodsell, D. S. PDB101: Goodsell Gallery: Myelin RCSB: PDB-101, <https://pdb101.rcsb.org/sci-art/goodsell-gallery/myelin> (accessed 07/16/2022) (page [xviii](#)).
- (11) Chen, F., *Introduction to Plasma Physics and Controlled Fusion*, 2018 (page [xvii](#)).
- (12) Dawkins, R., *The Extended Phenotype: The Long Reach of the Gene*, Reprint edition; Oxford University Press, USA: 2016; 440 pp. (pages [xix](#), [10](#)).
- (13) Griffiths, D. J., *Introduction to Electrodynamics*, 4th edition; Cambridge University Press: Cambridge, 2017; 620 pp. (pages [xix](#), [2](#)).
- (14) Reif, F., *Fundamentals of Statistical and Thermal Physics*, 1st edition; Wave-land Press, Inc.: 2009; 651 pp. (page [xix](#)).

- (15) Zuckerman, D. M., *Statistical Physics of Biomolecules: An Introduction*, 1st edition; CRC Press: Boca Raton, 2010; 356 pp. (pages xix, 41).
- (16) Salzberg, S. L. *BMC Biology* **2018**, *16*, 94, DOI: [10.1186/s12915-018-0564-x](https://doi.org/10.1186/s12915-018-0564-x) (pages xix, 71).
- (17) Adams, D., *The Hitch Hiker's Guide to the Galaxy*; Pan Original; Pan Books: London, 1979; 158 pp. (pages xix, 169).
- (18) Frauenfelder, H.; Wolynes, P. G. *Physics Today* **1994**, *47*, 58–64, DOI: [10.1063/1.881414](https://doi.org/10.1063/1.881414) (page 1).
- (19) Carroll, S. M., *Spacetime and Geometry: An Introduction to General Relativity*; Cambridge University Press: New York, 2019; 1 p. (page 2).
- (20) Griffiths, D., *Introduction to Quantum Mechanics*, 1st edition; Pearson College Div: Englewood Cliffs, N.J, 1994; 394 pp. (page 2).
- (21) Varn, D. P.; Crutchfield, J. P. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016**, *374*, 20150067, DOI: [10.1098/rsta.2015.0067](https://doi.org/10.1098/rsta.2015.0067) (page 3).
- (22) Watson, J. D., *Avoid Boring People: Lessons from a Life in Science*; Vintage Books: 2010; 370 pp. (page 3).
- (23) Picker, Z. The Gravity of Particle Physics: Dark Matter, Black Holes, and Axions, Ph.D. Thesis, U. Sydney (main), 2022 (page 4).
- (24) Carroll, S. M. *Journal of Consciousness Studies* **2021**, *28*, 16–31 (page 4).
- (25) Moy, G.; Corry, B.; Kuyucak, S.; Chung, S.-H. *Biophysical Journal* **2000**, *78*, 2349–2363, DOI: [10.1016/S0006-3495\(00\)76780-4](https://doi.org/10.1016/S0006-3495(00)76780-4) (page 5).
- (26) Corry, B.; Kuyucak, S.; Chung, S. H. *Biophysical Journal* **2000**, *78*, 2364–2381, DOI: [10.1016/S0006-3495\(00\)76781-6](https://doi.org/10.1016/S0006-3495(00)76781-6) (page 5).
- (27) Hille, B., *Ion Channels of Excitable Membranes*; Sinauer Associates: 2001 (pages 5, 67).
- (28) Catterall, W. A. *Cold Spring Harbor Perspectives in Biology* **2011**, *3*, a003947, DOI: [10.1101/cshperspect.a003947](https://doi.org/10.1101/cshperspect.a003947) (page 5).
- (29) Muthuswamy, S. K.; Xue, B. *Annual review of cell and developmental biology* **2012**, *28*, 599–625, DOI: [10.1146/annurev-cellbio-092910-154244](https://doi.org/10.1146/annurev-cellbio-092910-154244) (page 5).
- (30) Levin, M. *The Journal of Physiology* **2014**, *592*, 2295–2305, DOI: [10.1113/jphysiol.2014.271940](https://doi.org/10.1113/jphysiol.2014.271940) (pages 5, 8).
- (31) Levin, M. *Molecular Biology of the Cell* **2014**, *25*, 3835–3850, DOI: [10.1091/mbc.E13-12-0708](https://doi.org/10.1091/mbc.E13-12-0708) (pages 5, 8).
- (32) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; Overington, J. P. *Nature Reviews. Drug Discovery* **2017**, *16*, 19–34, DOI: [10.1038/nrd.2016.230](https://doi.org/10.1038/nrd.2016.230) (pages 5, 129).

- (33) Stansfeld, P. J.; Sutcliffe, M. J.; Mitcheson, J. S. *Expert Opinion on Drug Metabolism & Toxicology* **2006**, *2*, 81–94, DOI: [10.1517/17425255.2.1.81](https://doi.org/10.1517/17425255.2.1.81) (page 5).
- (34) Kaczorowski, G. J.; McManus, O. B.; Priest, B. T.; Garcia, M. L. *The Journal of General Physiology* **2008**, *131*, 399–405, DOI: [10.1085/jgp.200709946](https://doi.org/10.1085/jgp.200709946) (page 5).
- (35) Waszkielewicz, A.; Gunia, A.; Szkaradek, N.; Słoczyńska, K.; Krupińska, S.; Marona, H. *Current Medicinal Chemistry* **2013**, *20*, 1241–1285, DOI: [10.2174/0929867311320100005](https://doi.org/10.2174/0929867311320100005) (page 5).
- (36) Aidley, D. J.; Stanfield, P. R., *Ion Channels: Molecules in Action*; Cambridge University Press: 1996; 324 pp. (pages 6, 11, 67).
- (37) The Hodgkin–Huxley Model Ebrary, https://ebrary.net/196552/health/hodgkin_huxley_model (accessed 07/15/2022) (page 6).
- (38) Hodgkin, A. L.; Huxley, A. F. *The Journal of Physiology* **1952**, *117*, 500–544, DOI: [10.1113/jphysiol.1952.sp004764](https://doi.org/10.1113/jphysiol.1952.sp004764) (page 6).
- (39) Hodgkin, A. L.; Huxley, A. F.; Katz, B. *The Journal of Physiology* **1952**, *116*, 424–448 (page 6).
- (40) Hodgkin, A. L.; Huxley, A. F. *The Journal of Physiology* **1952**, *116*, 497–506 (page 6).
- (41) Hodgkin, A. L.; Huxley, A. F. *The Journal of Physiology* **1952**, *116*, 473–496 (page 6).
- (42) Hodgkin, A. L.; Huxley, A. F. *The Journal of Physiology* **1952**, *116*, 449–472 (page 6).
- (43) Sham, S. S.; Shobana, S.; Townsley, L. E.; Jordan, J. B.; Fernandez, J. Q.; Andersen, O. S.; Greathouse, D. V.; Hinton, J. F. *Biochemistry* **2003**, *42*, 1401–1409, DOI: [10.1021/bi0204286](https://doi.org/10.1021/bi0204286) (page 7).
- (44) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. *Science* **1998**, *280*, 69–77, DOI: [10.1126/science.280.5360.69](https://doi.org/10.1126/science.280.5360.69) (pages 7, 8, 129).
- (45) Zhang, Z.; Liu, F.; Chen, J. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, 12757–12762, DOI: [10.1073/pnas.1815287115](https://doi.org/10.1073/pnas.1815287115) (pages 7, 51, 74, 75, 83, 87, 88, 96, 97).
- (46) Roux, B.; Karplus, M. *Journal of the American Chemical Society* **1993**, *115*, 3250–3262, DOI: [10.1021/ja00061a025](https://doi.org/10.1021/ja00061a025) (pages 7, 8, 129).
- (47) Jones, D.; Allen, J. E.; Yang, Y.; Drew Bennett, W. F.; Gokhale, M.; Moshiri, N.; Rosing, T. S. *Journal of Chemical Theory and Computation* **2022**, *18*, 4047–4069, DOI: [10.1021/acs.jctc.1c01214](https://doi.org/10.1021/acs.jctc.1c01214) (page 7).
- (48) Russell, J. Anton 3 Is a ‘Fire-Breathing’ Molecular Simulation Beast HPCwire, <https://www.hpcwire.com/2021/09/01/anton-3-is-a-fire-breathing-molecular-simulation-beast/> (accessed 07/13/2022) (page 7).

- (49) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590, DOI: [10.1038/267585a0](https://doi.org/10.1038/267585a0) (page 8).
- (50) Sansom, M. S. *Progress in Biophysics and Molecular Biology* **1991**, *55*, 139–235, DOI: [10.1016/0079-6107\(91\)90004-C](https://doi.org/10.1016/0079-6107(91)90004-C) (page 8).
- (51) Roux, B.; Karplus, M. *The Journal of Physical Chemistry* **1991**, *95*, 4856–4868, DOI: [10.1021/j100165a049](https://doi.org/10.1021/j100165a049) (page 8).
- (52) Tajkhorshid, E.; Nollert, P.; o. Jensen, M.; Miercke, L. J. W.; O'Connell, J.; Stroud, R. M.; Schulten, K. *Science* **2002**, *296*, 525–530, DOI: [10.1126/science.1067778](https://doi.org/10.1126/science.1067778) (page 8).
- (53) De Groot, B. L.; Grubmuller, H. *Science* **2001**, *294*, 2353–2357, DOI: [10.1126/science.1066115](https://doi.org/10.1126/science.1066115) (page 8).
- (54) Allen, T. W.; Baştug, T.; Kuyucak, S.; Chung, S.-H. *Biophysical Journal* **2003**, *84*, 2159–2168, DOI: [10.1016/S0006-3495\(03\)75022-X](https://doi.org/10.1016/S0006-3495(03)75022-X) (page 8).
- (55) Allen, T. W.; Andersen, O. S.; Roux, B. *Proceedings of the National Academy of Sciences* **2004**, *101*, 117–122, DOI: [10.1073/pnas.2635314100](https://doi.org/10.1073/pnas.2635314100) (page 8).
- (56) Chung, S.-H.; Kuyucak, S. *European Biophysics Journal* **2002**, *31*, 283–293, DOI: [10.1007/s00249-002-0216-4](https://doi.org/10.1007/s00249-002-0216-4) (page 8).
- (57) Tieleman, D. P.; C. Biggin, P.; R. Smith, G.; S. P. Sansom, M. *Quarterly Reviews of Biophysics* **2001**, *34*, 473–561, DOI: [10.1017/S0033583501003729](https://doi.org/10.1017/S0033583501003729) (page 8).
- (58) Urry, D. W.; Goodall, M. C.; Glickson, J. D.; Mayers, D. F. *Proceedings of the National Academy of Sciences of the United States of America* **1971**, *68*, 1907–1911, DOI: [10.1073/pnas.68.8.1907](https://doi.org/10.1073/pnas.68.8.1907) (page 8).
- (59) Arseniev, A.; Barsukov, I.; Bystrov, V.; Lomize, A.; Ovchinnikov, Y. *FEBS Letters* **1985**, *186*, 168–174, DOI: [10.1016/0014-5793\(85\)80702-X](https://doi.org/10.1016/0014-5793(85)80702-X) (page 8).
- (60) Wallace, B. A. *Biophysical journal* **1986**, *49*, 295–306, DOI: [10.1016/s0006-3495\(86\)83642-6](https://doi.org/10.1016/s0006-3495(86)83642-6) (page 8).
- (61) Wallace, B. A *Journal of Structural Biology* **1998**, *121*, 123–141, DOI: [10.1006/jsbi.1997.3948](https://doi.org/10.1006/jsbi.1997.3948) (page 8).
- (62) Liou, J.-W.; Hung, Y.-J.; Yang, C.-H.; Chen, Y.-C. *PLoS ONE* **2015**, *10*, e0117065, DOI: [10.1371/journal.pone.0117065](https://doi.org/10.1371/journal.pone.0117065) (page 8).
- (63) Kühlbrandt, W. *Science* **2014**, *343*, 1443–1444, DOI: [10.1126/science.1251652](https://doi.org/10.1126/science.1251652) (pages 8, 31).
- (64) Clapham, D. E. *Cell* **2003**, *115*, 641–646, DOI: [10.1016/S0092-8674\(03\)00980-2](https://doi.org/10.1016/S0092-8674(03)00980-2) (page 8).
- (65) Rashid, M. H.; Heinzelmann, G.; Huq, R.; Tajhya, R. B.; Chang, S. C.; Chhabra, S.; Pennington, M. W.; Beeton, C.; Norton, R. S.; Kuyucak, S. *PLOS ONE* **2013**, *8*, e78712, DOI: [10.1371/journal.pone.0078712](https://doi.org/10.1371/journal.pone.0078712) (page 8).
- (66) Li, J.; Shen, R.; Reddy, B.; Perozo, E.; Roux, B. *Science Advances* **2021**, *7*, eabd6203, DOI: [10.1126/sciadv.abd6203](https://doi.org/10.1126/sciadv.abd6203) (page 8).

- (67) Vandenberg, J.; Lau, C.; Flood, E.; Hunter, M.; Ng, C.-A.; Bouwer, J.; Stewart, A.; Perozo, E.; Allen, T. *Structural Basis for Rapid Voltage Dependent Inactivation of HERG Potassium Channels*; preprint; In Review, 2021, DOI: [10.21203/rs.3.rs-1105661/v1](https://doi.org/10.21203/rs.3.rs-1105661/v1) (page 8).
- (68) Flood, E.; Boiteux, C.; Lev, B.; Vorobyov, I.; Allen, T. W. *Chemical Reviews* **2019**, *119*, 7737–7832, DOI: [10.1021/acs.chemrev.8b00630](https://doi.org/10.1021/acs.chemrev.8b00630) (pages 8, 30, 112, 132).
- (69) Jumper, J. et al. *Nature* **2021**, *596*, 583–589, DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2) (pages 8, 10, 41, 59, 85, 164).
- (70) Frank, J. *Biochemistry* **2021**, *60*, 3449–3451, DOI: [10.1021/acs.biochem.1c00318](https://doi.org/10.1021/acs.biochem.1c00318) (page 8).
- (71) Cheng, Y.; Glaeser, R. M.; Nogales, E. *Cell* **2017**, *171*, 1229–1231, DOI: [10.1016/j.cell.2017.11.016](https://doi.org/10.1016/j.cell.2017.11.016) (page 8).
- (72) Chesler, A. T.; Szcztot, M. *eLife* **2018**, *7*, e34396, DOI: [10.7554/eLife.34396](https://doi.org/10.7554/eLife.34396) (page 8).
- (73) Castillo, K.; Diaz-Franulic, I.; Canan, J.; Gonzalez-Nilo, F.; Latorre, R. *Physical Biology* **2018**, *15*, 021001, DOI: [10.1088/1478-3975/aa9a6f](https://doi.org/10.1088/1478-3975/aa9a6f) (page 8).
- (74) Kingwell, K. *Nature Reviews Drug Discovery* **2019**, *18*, 321–323, DOI: [10.1038/d41573-019-00065-0](https://doi.org/10.1038/d41573-019-00065-0) (pages 8, 129).
- (75) Kweon, H.-J.; Suh, B.-C. *BMB Reports* **2013**, *46*, 295–304, DOI: [10.5483/BMBRep.2013.46.6.121](https://doi.org/10.5483/BMBRep.2013.46.6.121) (page 8).
- (76) Sato, K.; Pellegrino, M.; Nakagawa, T.; Nakagawa, T.; Vosshall, L. B.; Touhara, K. *Nature* **2008**, *452*, 1002–1006, DOI: [10.1038/nature06850](https://doi.org/10.1038/nature06850) (page 8).
- (77) Lang, F.; Föller, M.; Lang, K. S.; Lang, P. A.; Ritter, M.; Gulbins, E.; Vereninov, A.; Huber, S. M. *The Journal of Membrane Biology* **2005**, *205*, 147–157, DOI: [10.1007/s00232-005-0780-5](https://doi.org/10.1007/s00232-005-0780-5) (page 8).
- (78) Sundelacruz, S.; Levin, M.; Kaplan, D. L. *Stem Cell Reviews and Reports* **2009**, *5*, 231–246, DOI: [10.1007/s12015-009-9080-2](https://doi.org/10.1007/s12015-009-9080-2) (page 8).
- (79) Lev, B.; Allen, T. W. *Journal of Computational Chemistry* **2020**, *41*, 387–401, DOI: [10.1002/jcc.26102](https://doi.org/10.1002/jcc.26102) (pages 8, 132, 139).
- (80) Chen, I.; Pant, S.; Wu, Q.; Cater, R. J.; Sobti, M.; Vandenberg, R. J.; Stewart, A. G.; Tajkhorshid, E.; Font, J.; Ryan, R. M. *Nature* **2021**, *591*, 327–331, DOI: [10.1038/s41586-021-03240-9](https://doi.org/10.1038/s41586-021-03240-9) (page 8).
- (81) Katkin, J. P. Cystic Fibrosis: Clinical Manifestations and Diagnosis - UpToDate <https://www-upToDate-com.ezproxy-so.library.sydney.edu.au/contents/cystic-fibrosis-clinical-manifestations-and-diagnosis> (accessed 07/16/2022) (page 8).
- (82) Foucault, M., *The Birth of the Clinic: An Archaeology of Medical Perception*, Reprint edition; Vintage: New York, 1994; 240 pp. (pages 8, 44).
- (83) Dubois, B. et al. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* **2016**, *12*, 292–323, DOI: [10.1016/j.jalz.2016.02.002](https://doi.org/10.1016/j.jalz.2016.02.002) (page 9).

- (84) Tsui, L.-C.; Dorfman, R. *Cold Spring Harbor Perspectives in Medicine* **2013**, *3*, a009472, DOI: [10.1101/cshperspect.a009472](https://doi.org/10.1101/cshperspect.a009472) (page 9).
- (85) DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. *Nature Reviews Genetics* **2005**, *6*, 678–687, DOI: [10.1038/nrg1672](https://doi.org/10.1038/nrg1672) (page 9).
- (86) For Sick Children, T. H. *The Clinical and Functional TRanslation of CFTR (CFTR2)*; 2020 (pages 9, 49, 54, 57, 87, 113, 119, 139, 161).
- (87) Csanady, L.; Vergani, P.; Gadsby, D. C. *Physiological Reviews* **2019**, *99*, 707–738, DOI: [10.1152/physrev.00007.2018](https://doi.org/10.1152/physrev.00007.2018) (pages 10, 53).
- (88) Anonymous *The Economist* **2019**, *431*, 11 (pages 10, 11).
- (89) Scown, C. D.; Keasling, J. D. *Nature Biotechnology* **2022**, *40*, 304–307, DOI: [10.1038/s41587-022-01248-8](https://doi.org/10.1038/s41587-022-01248-8) (page 10).
- (90) Kochanski, Z. *Philosophy of Science* **1973**, *40*, 29–50, DOI: [10.1086/288494](https://doi.org/10.1086/288494) (page 10).
- (91) Liu, E. T. *Cell* **2005**, *121*, 505–506, DOI: [10.1016/j.cell.2005.04.021](https://doi.org/10.1016/j.cell.2005.04.021) (page 10).
- (92) Mogilner, A. *Molecular Biology of the Cell* **2016**, *27*, 3377–3378, DOI: [10.1091/mcb.E16-09-0673](https://doi.org/10.1091/mcb.E16-09-0673) (page 10).
- (93) Covert, M. W.; Gillies, T. E.; Kudo, T.; Agmon, E. *Cell Systems* **2021**, *12*, 488–496, DOI: [10.1016/j.cels.2021.05.014](https://doi.org/10.1016/j.cels.2021.05.014) (page 10).
- (94) Benyus, J. M., *Biomimicry: Innovation Inspired by Nature*, Nachdr.; Perennial: New York, NY, 2009; 308 pp. (page 10).
- (95) Wang, Y.; Xue, P.; Cao, M.; Yu, T.; Lane, S. T.; Zhao, H. *Chemical Reviews* **2021**, *121*, 12384–12444, DOI: [10.1021/acs.chemrev.1c00260](https://doi.org/10.1021/acs.chemrev.1c00260) (page 10).
- (96) Arnold, F. H. *Angewandte Chemie (International Ed. in English)* **2018**, *57*, 4143–4148, DOI: [10.1002/anie.201708408](https://doi.org/10.1002/anie.201708408) (page 10).
- (97) Wu, X. L.; Farr, R., *Physics in Focus : Preliminary Course*; McGraw-Hill: North Ryde, N.S.W., 2009 (page 11).
- (98) Cheng, Y.; Grigorieff, N.; Penczek, P. A.; Walz, T. *Cell* **2015**, *161*, 438–449, DOI: [10.1016/j.cell.2015.03.050](https://doi.org/10.1016/j.cell.2015.03.050) (page 11).
- (99) Callaway, E. *Nature* **2015**, *525*, 172–174, DOI: [10.1038/525172a](https://doi.org/10.1038/525172a) (page 11).
- (100) Callaway, E. *Nature* **2020**, *578*, 201–201, DOI: [10.1038/d41586-020-00341-9](https://doi.org/10.1038/d41586-020-00341-9) (page 11).
- (101) Marion, D. *Molecular & Cellular Proteomics* **2013**, *12*, 3006–3025, DOI: [10.1074/mcp.0113.030239](https://doi.org/10.1074/mcp.0113.030239) (page 11).
- (102) Sanderson, M. J.; Smith, I.; Parker, I.; Bootman, M. D. *Cold Spring Harbor protocols* **2014**, *2014*, pdb.top071795, DOI: [10.1101/pdb.top071795](https://doi.org/10.1101/pdb.top071795) (page 11).
- (103) Frauenfelder, H.; Chan, S. S.; Chan, W. S.; Austin, R. H., *The Physics of Proteins: An Introduction to Biological Physics and Molecular Biophysics*; Biological and Medical Physics, Biomedical Engineering; Springer New York: New York, NY, 2010; 448 pp., DOI: [10.1007/978-1-4419-1044-8](https://doi.org/10.1007/978-1-4419-1044-8) (pages 11, 40).

- (104) Drenth, J., *Principles of Protein X-Ray Crystallography*, 3rd edition; Springer: New York Heidelberg, 2006; 346 pp. (page 11).
- (105) Silva, D.; Santos, G.; Barroca, M.; Collins, T. *Methods in Molecular Biology (Clifton, N.J.)* **2017**, *1620*, 87–100, DOI: [10.1007/978-1-4939-7060-5_5](https://doi.org/10.1007/978-1-4939-7060-5_5) (page 11).
- (106) *Nature Cell Biology* **2019**, *21*, 1463–1463, DOI: [10.1038/s41556-019-0434-y](https://doi.org/10.1038/s41556-019-0434-y) (page 11).
- (107) Feynman, R. P. *International Journal of Theoretical Physics* **1982**, *21*, 467–488, DOI: [10.1007/BF02650179](https://doi.org/10.1007/BF02650179) (page 12).
- (108) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. *Living journal of computational molecular science* **2019**, *1*, 5957, DOI: [10.33011/livecoms.1.1.5957](https://doi.org/10.33011/livecoms.1.1.5957) (pages 12, 25, 28, 29).
- (109) Gapsys, V.; de Groot, B. L. *eLife* **2020**, *9*, ed. by Faraldo-Gómez, J. D.; Grossfield, A., e57589, DOI: [10.7554/eLife.57589](https://doi.org/10.7554/eLife.57589) (pages 12, 21, 35).
- (110) Grossfield, A.; Patrone, P. N.; Roe, D. R.; Schultz, A. J.; Siderius, D.; Zuckerman, D. M. *Living Journal of Computational Molecular Science* **2019**, *1*, 5067–5067, DOI: [10.33011/livecoms.1.1.5067](https://doi.org/10.33011/livecoms.1.1.5067) (page 12).
- (111) Pohorille, A.; Jarzynski, C.; Chipot, C. *Journal of Physical Chemistry B* **2010**, *114*, 10235–10253, DOI: [10.1021/jp102971x](https://doi.org/10.1021/jp102971x) (page 12).
- (112) *Free Energy Calculations: Theory and Applications in Chemistry and Biology*: 86, 2007th edition; Chipot, C., Pohorille, A., Chipot, C., Eds.; Springer: Berlin Heidelberg, 2007; 536 pp. (pages 12, 32, 33, 168).
- (113) Sherrill, C. D., 7 (pages 14, 15).
- (114) *Dynamics of Molecular Collisions: Part B*; Miller, W. H., Ed.; Springer US: Boston, MA, 1976, DOI: [10.1007/978-1-4757-0644-4](https://doi.org/10.1007/978-1-4757-0644-4) (pages 14, 15).
- (115) Van Mourik, T.; Bühl, M.; Gaigeot, M.-P. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2014**, *372*, 20120488, DOI: [10.1098/rsta.2012.0488](https://doi.org/10.1098/rsta.2012.0488) (page 16).
- (116) Luo, Z.; Qin, X.; Wan, L.; Hu, W.; Yang, J. *Frontiers in Chemistry* **2020**, *8* (page 16).
- (117) Kresse, G.; Furthmüller, J. *Computational Materials Science* **1996**, *6*, 15–50, DOI: [10.1016/0927-0256\(96\)00008-0](https://doi.org/10.1016/0927-0256(96)00008-0) (page 16).
- (118) Lemkul, J. A. In *Progress in Molecular Biology and Translational Science*, Strodel, B., Barz, B., Eds.; Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly, Vol. 170; Academic Press: 2020, pp 1–71, DOI: [10.1016/bs.pmbts.2019.12.009](https://doi.org/10.1016/bs.pmbts.2019.12.009) (pages 16, 19, 20).
- (119) Mackerell Jr., A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584–1604, DOI: [10.1002/jcc.20082](https://doi.org/10.1002/jcc.20082) (page 17).
- (120) MacKerell, A. D.; Feig, M.; Brooks, C. L. *Journal of the American Chemical Society* **2004**, *126*, 698–699, DOI: [10.1021/ja036959e](https://doi.org/10.1021/ja036959e) (page 17).

- (121) Mackerell, A. D.; Feig, M.; Brooks, C. L. *Journal of Computational Chemistry* **2004**, *25*, 1400–1415, DOI: [10.1002/jcc.20065](https://doi.org/10.1002/jcc.20065) (page 17).
- (122) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *Journal of Molecular Biology* **1963**, *7*, 95–99, DOI: [10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6) (page 17).
- (123) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. *Nature Methods* **2016**, *14*, 71–73, DOI: [10.1038/nmeth.4067](https://doi.org/10.1038/nmeth.4067) (pages 17, 18, 20, 41, 97, 127).
- (124) Yoo, J.; Aksimentiev, A. *Physical Chemistry Chemical Physics* **2018**, *20*, 8432–8449, DOI: [10.1039/c7cp08185e](https://doi.org/10.1039/c7cp08185e) (page 18).
- (125) Jämbeck, J. P. M.; Mocci, F.; Lyubartsev, A. P.; Laaksonen, A. *Journal of Computational Chemistry* **2013**, *34*, 187–197, DOI: [10.1002/jcc.23117](https://doi.org/10.1002/jcc.23117) (page 18).
- (126) Tolmachev, D. A.; Boyko, O. S.; Lukasheva, N. V.; Martinez-Seara, H.; Karttunen, M. *Journal of Chemical Theory and Computation* **2020**, *16*, 677–687, DOI: [10.1021/acs.jctc.9b00813](https://doi.org/10.1021/acs.jctc.9b00813) (page 18).
- (127) Savelyev, A.; MacKerell, A. D. *The Journal of Physical Chemistry B* **2015**, *119*, 4428–4440, DOI: [10.1021/acs.jpcb.5b00683](https://doi.org/10.1021/acs.jpcb.5b00683) (page 18).
- (128) Chari, S. S. N.; Dasgupta, C.; Maiti, P. K. *Soft Matter* **2019**, *15*, 7275–7285, DOI: [10.1039/C9SM00962K](https://doi.org/10.1039/C9SM00962K) (page 20).
- (129) Morsali, A.; Goharshadi, E. K.; Ali Mansoori, G.; Abbaspour, M. *Chemical Physics* **2005**, *310*, 11–15, DOI: [10.1016/j.chemphys.2004.09.027](https://doi.org/10.1016/j.chemphys.2004.09.027) (page 20).
- (130) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *Journal of Computational Chemistry* **2010**, *31*, 671–690, DOI: [10.1002/jcc.21367](https://doi.org/10.1002/jcc.21367) (page 20).
- (131) Robustelli, P.; Ibanez-de Opakua, A.; Campbell-Bezat, C.; Giordanetto, F.; Becker, S.; Zweckstetter, M.; Pan, A. C.; Shaw, D. E. *Journal of the American Chemical Society* **2022**, *144*, 2501–2510, DOI: [10.1021/jacs.1c07591](https://doi.org/10.1021/jacs.1c07591) (pages 20, 29, 39).
- (132) Robustelli, P.; Piana, S.; Shaw, D. E. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, *115*, E4758–E4766, DOI: [10.1073/pnas.1800690115](https://doi.org/10.1073/pnas.1800690115) (page 20).
- (133) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174, DOI: [10.1002/jcc.20035](https://doi.org/10.1002/jcc.20035) (page 20).
- (134) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *Journal of Molecular Graphics and Modelling* **2006**, *25*, 247–260, DOI: [10.1016/j.jmgm.2005.12.005](https://doi.org/10.1016/j.jmgm.2005.12.005) (page 20).
- (135) Zhu, S. *Journal of Chemical Information and Modeling* **2019**, *59*, 4239–4247, DOI: [10.1021/acs.jcim.9b00552](https://doi.org/10.1021/acs.jcim.9b00552) (page 20).
- (136) Ross, G. A.; Rustenburg, A. S.; Grinaway, P. B.; Fass, J.; Chodera, J. D. *The journal of physical chemistry. B* **2018**, *122*, 5466–5486, DOI: [10.1021/acs.jpcb.7b11734](https://doi.org/10.1021/acs.jpcb.7b11734) (page 21).

- (137) Klauda, J. B.; Wu, X.; Pastor, R. W.; Brooks, B. R. *The Journal of Physical Chemistry B* **2007**, *111*, 4393–4400, DOI: [10.1021/jp068767m](https://doi.org/10.1021/jp068767m) (page 21).
- (138) Venable, R. M.; Chen, L. E.; Pastor, R. W. *The Journal of Physical Chemistry B* **2009**, *113*, 5855–5862, DOI: [10.1021/jp900843x](https://doi.org/10.1021/jp900843x) (page 21).
- (139) Auffinger, P.; Beveridge, D. L. *Chemical Physics Letters* **1995**, *234*, 413–415, DOI: [10.1016/0009-2614\(95\)00065-C](https://doi.org/10.1016/0009-2614(95)00065-C) (page 21).
- (140) Perera, L.; Essmann, U.; Berkowitz, M. L. *The Journal of Chemical Physics* **1995**, *102*, 450–456, DOI: [10.1063/1.469422](https://doi.org/10.1063/1.469422) (page 21).
- (141) Roberts, J. E.; Schnitker, J. *The Journal of Chemical Physics* **1994**, *101*, 5024–5031, DOI: [10.1063/1.467425](https://doi.org/10.1063/1.467425) (page 21).
- (142) Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. *Chemical Physics Letters* **1996**, *263*, 521–529, DOI: [10.1016/S0009-2614\(96\)01234-1](https://doi.org/10.1016/S0009-2614(96)01234-1) (page 21).
- (143) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593, DOI: [10.1063/1.470117](https://doi.org/10.1063/1.470117) (page 21).
- (144) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. *Journal of Chemical Theory and Computation* **2014**, *10*, 381–390, DOI: [10.1021/ct400626b](https://doi.org/10.1021/ct400626b) (page 23).
- (145) Darden, T.; York, D.; Pedersen, L. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092, DOI: [10.1063/1.464397](https://doi.org/10.1063/1.464397) (page 23).
- (146) Hardy, D. J.; Wu, Z.; Phillips, J. C.; Stone, J. E.; Skeel, R. D.; Schulten, K. *Journal of Chemical Theory and Computation* **2015**, *11*, 766–779, DOI: [10.1021/ct5009075](https://doi.org/10.1021/ct5009075) (page 23).
- (147) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *The Journal of Chemical Physics* **2005**, *122*, 054101, DOI: [10.1063/1.1839571](https://doi.org/10.1063/1.1839571) (pages 23, 24).
- (148) Peterson, M. E.; Daniel, R. M.; Danson, M. J.; Eisenthal, R. *Biochemical Journal* **2007**, *402*, 331–337, DOI: [10.1042/BJ20061143](https://doi.org/10.1042/BJ20061143) (page 24).
- (149) Song, X.; Li, Y. *Journal of Food Process Engineering* **2012**, *35*, 915–922, DOI: [10.1111/j.1745-4530.2011.00641.x](https://doi.org/10.1111/j.1745-4530.2011.00641.x) (page 24).
- (150) Knapp, B.; Ospina, L.; Deane, C. M. *Journal of Chemical Theory and Computation* **2018**, *14*, 6127–6138, DOI: [10.1021/acs.jctc.8b00391](https://doi.org/10.1021/acs.jctc.8b00391) (pages 25, 59).
- (151) Nosé, S. *Molecular Physics* **1984**, *52*, 255–268, DOI: [10.1080/00268978400101201](https://doi.org/10.1080/00268978400101201) (page 25).
- (152) Hoover, W. G. *Physical Review A* **1985**, *31*, 1695–1697, DOI: [10.1103/PhysRevA.31.1695](https://doi.org/10.1103/PhysRevA.31.1695) (page 25).
- (153) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *The Journal of Chemical Physics* **1992**, *97*, 2635–2643, DOI: [10.1063/1.463940](https://doi.org/10.1063/1.463940) (pages 25, 27).
- (154) Bussi, G.; Donadio, D.; Parrinello, M. *Journal of Chemical Physics* **2007**, *126*, 014101–014101, DOI: [10.1063/1.2408420](https://doi.org/10.1063/1.2408420) (page 25).

- (155) Berendsen, H. J.; Postma, J. P.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *The Journal of Chemical Physics* **1984**, *81*, 3684–3690, DOI: [10.1063/1.448118](https://doi.org/10.1063/1.448118) (pages 25, 97).
- (156) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. *Molecular Physics* **1996**, *87*, 1117–1157, DOI: [10.1080/00268979600100761](https://doi.org/10.1080/00268979600100761) (page 25).
- (157) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1–2*, 19–25, DOI: [10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001) (pages 25, 41, 97, 127).
- (158) Karal, M. A. S.; Ahamed, M. K.; Ahmed, M.; Mahbub, Z. B. *RSC Advances* **2021**, *11*, 29598–29619, DOI: [10.1039/D1RA04647K](https://doi.org/10.1039/D1RA04647K) (page 27).
- (159) Macdonald, A. G. In *Cell Physiology Source Book (Third Edition)*, Sperelakis, N., Ed.; Academic Press: San Diego, 2001, pp 1003–1023, DOI: [10.1016/B978-012656976-6/50151-7](https://doi.org/10.1016/B978-012656976-6/50151-7) (page 27).
- (160) Allen, M; Tildesley, D, *Computer Simulation of Liquids*; Clarendon Press Oxford: 1991 (page 27).
- (161) Parrinello, M.; Rahman, A. *Physical Review Letters* **1980**, *45*, 1196–1199, DOI: [10.1103/PhysRevLett.45.1196](https://doi.org/10.1103/PhysRevLett.45.1196) (page 27).
- (162) Parrinello, M.; Rahman, A. *Journal of Applied Physics* **1981**, *52*, 7182–7190, DOI: [10.1063/1.328693](https://doi.org/10.1063/1.328693) (pages 27, 97).
- (163) Leach, A., *Molecular Modelling Principles and Applications*; Prentice Hall: 2001 (pages 29, 30).
- (164) Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*, 2nd ed; Interdisciplinary Applied Mathematics v. 21; Springer: New York, 2010 (pages 29, 30, 40).
- (165) Shannon, C. *Proceedings of the IRE* **1949**, *37*, 10–21, DOI: [10.1109/JRPROC.1949.232969](https://doi.org/10.1109/JRPROC.1949.232969) (page 29).
- (166) Mazur, A. K. *Journal of Computational Physics* **1997**, *136*, 354–365, DOI: [10.1006/jcph.1997.5740](https://doi.org/10.1006/jcph.1997.5740) (page 29).
- (167) Feenstra, K. A.; Hess, B.; Berendsen, H. J. *Journal of Computational Chemistry* **1999**, *20*, 786–798, DOI: [10.1002/\(SICI\)1096-987X\(199906\)20:8<786::AID-JCC5>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B) (pages 29, 30, 127).
- (168) Andersen, H. C. *Journal of Computational Physics* **1983**, *52*, 24–34, DOI: [10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1) (page 29).
- (169) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472, DOI: [10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H) (page 29).
- (170) Balusek, C.; Hwang, H.; Lau, C. H.; Lundquist, K.; Hazel, A.; Pavlova, A.; Lynch, D. L.; Reggio, P. H.; Wang, Y.; Gumbart, J. C. *Journal of chemical theory and computation* **2019**, *15*, 4673–4686, DOI: [10.1021/acs.jctc.9b00160](https://doi.org/10.1021/acs.jctc.9b00160) (pages 29, 148).

- (171) Streett, W.; Tildesley, D.; Saville, G. *Molecular Physics* **1978**, *35*, 639–648, DOI: [10.1080/00268977800100471](https://doi.org/10.1080/00268977800100471) (pages 30, 31).
- (172) *Advances in Chemical Physics: Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Brooks, C. L., Karplus, M., Pettitt, B. M., Eds.; Advances in Chemical Physics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1988, DOI: [10.1002/9780470141205](https://doi.org/10.1002/9780470141205) (page 30).
- (173) Werner, T.; Morris, M. B.; Dastmalchi, S.; Church, W. B. *Advanced Drug Delivery Reviews* **2012**, *64*, 323–343, DOI: [10.1016/j.addr.2011.11.011](https://doi.org/10.1016/j.addr.2011.11.011) (page 30).
- (174) Kästner, J. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932–942, DOI: [10.1002/wcms.66](https://doi.org/10.1002/wcms.66) (pages 33, 35).
- (175) Torrie, G. M.; Valleau, J. P. *Journal of Computational Physics* **1977**, *23*, 187–199, DOI: [10.1016/0021-9991\(77\)90121-8](https://doi.org/10.1016/0021-9991(77)90121-8) (page 33).
- (176) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141–151, DOI: [10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9) (page 33).
- (177) Berg, B. A.; Neuhaus, T. *Physics Letters B* **1991**, *267*, 249–253, DOI: [10.1016/0370-2693\(91\)91256-U](https://doi.org/10.1016/0370-2693(91)91256-U) (page 33).
- (178) Laio, A.; Parrinello, M. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 12562–12566, DOI: [10.1073/pnas.202427399](https://doi.org/10.1073/pnas.202427399) (pages 33, 37).
- (179) Roux, B. *The Journal of Physical Chemistry A* **2021**, *125*, 7558–7571, DOI: [10.1021/acs.jpca.1c04110](https://doi.org/10.1021/acs.jpca.1c04110) (pages 33, 165).
- (180) Pan, A. C.; Sezer, D.; Roux, B. *The Journal of Physical Chemistry B* **2008**, *112*, 3432–3440, DOI: [10.1021/jp0777059](https://doi.org/10.1021/jp0777059) (pages 33, 165).
- (181) E, W.; Ren, W.; Vanden-Eijnden, E. *Physical Review B* **2002**, *66*, 052301, DOI: [10.1103/PhysRevB.66.052301](https://doi.org/10.1103/PhysRevB.66.052301) (page 33).
- (182) Miao, Y.; Feher, V. A.; McCammon, J. A. *Journal of Chemical Theory and Computation* **2015**, *11*, 3584–3595, DOI: [10.1021/acs.jctc.5b00436](https://doi.org/10.1021/acs.jctc.5b00436) (page 33).
- (183) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. *The Journal of Physical Chemistry B* **2015**, *119*, 1129–1151, DOI: [10.1021/jp506633n](https://doi.org/10.1021/jp506633n) (page 33).
- (184) Isralewitz, B.; Gao, M.; Schulten, K. *Current Opinion in Structural Biology* **2001**, *11*, 224–230, DOI: [10.1016/s0959-440x\(00\)00194-9](https://doi.org/10.1016/s0959-440x(00)00194-9) (page 33).
- (185) Bussi, G.; Laio, A.; Parrinello, M. *Physical Review Letters* **2006**, *96*, 090601, DOI: [10.1103/PhysRevLett.96.090601](https://doi.org/10.1103/PhysRevLett.96.090601) (page 33).
- (186) Jorgensen, W. L.; Thomas, L. L. *Journal of chemical theory and computation* **2008**, *4*, 869–876, DOI: [10.1021/ct800011m](https://doi.org/10.1021/ct800011m) (pages 33, 168).
- (187) Jiang, W.; Roux, B. *Journal of chemical theory and computation* **2010**, *6*, 2559–2565, DOI: [10.1021/ct1001768](https://doi.org/10.1021/ct1001768) (page 33).
- (188) Jarzynski, C. *Physical Review E* **2002**, *65*, 046122, DOI: [10.1103/PhysRevE.65.046122](https://doi.org/10.1103/PhysRevE.65.046122) (page 33).

- (189) Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; van der Spoel, D.; de Groot, B. L. *Communications Chemistry* **2021**, *4*, 1–13, DOI: [10.1038/s42004-021-00498-y](https://doi.org/10.1038/s42004-021-00498-y) (page 33).
- (190) Jarzynski, C. *Physical Review E* **1997**, *56*, 5018–5035, DOI: [10.1103/PhysRevE.56.5018](https://doi.org/10.1103/PhysRevE.56.5018) (page 32).
- (191) Jarzynski, C. *Physical Review Letters* **1997**, *78*, 2690–2693, DOI: [10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690) (page 32).
- (192) Crooks, G. E. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* **1999**, *60*, 2721–2726, DOI: [10.1103/PhysRevE.60.2721](https://doi.org/10.1103/PhysRevE.60.2721) (page 32).
- (193) Kirkwood, J. G. *The Journal of Chemical Physics* **1935**, *3*, 300–313, DOI: [10.1063/1.1749657](https://doi.org/10.1063/1.1749657) (page 32).
- (194) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemonette, L. Enhanced Sampling Methods for Molecular Dynamics Simulations, 2022 (pages 32, 37).
- (195) Meshkin, H.; Zhu, F. *Journal of Chemical Theory and Computation* **2017**, *13*, 2086–2097, DOI: [10.1021/acs.jctc.6b01171](https://doi.org/10.1021/acs.jctc.6b01171) (page 34).
- (196) Domański, J.; Hedger, G.; Best, R. B.; Stansfeld, P. J.; Sansom, M. S. P. *The Journal of Physical Chemistry B* **2017**, *121*, 3364–3375, DOI: [10.1021/acs.jpcb.6b08445](https://doi.org/10.1021/acs.jpcb.6b08445) (page 34).
- (197) Zhu, F.; Hummer, G. *Journal of Chemical Theory and Computation* **2012**, *8*, 3759–3768, DOI: [10.1021/ct2009279](https://doi.org/10.1021/ct2009279) (pages 34, 35).
- (198) Subramanian, N.; Schumann-Gillett, A.; Mark, A. E.; O'Mara, M. L. *Journal of Chemical Information and Modeling* **2019**, *59*, 2287–2298, DOI: [10.1021/acs.jcim.8b00624](https://doi.org/10.1021/acs.jcim.8b00624) (page 34).
- (199) You, W.; Tang, Z.; Chang, C.-e. A. *Journal of Chemical Theory and Computation* **2019**, *15*, 2433–2443, DOI: [10.1021/acs.jctc.8b01142](https://doi.org/10.1021/acs.jctc.8b01142) (page 34).
- (200) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021, DOI: [10.1002/jcc.540130812](https://doi.org/10.1002/jcc.540130812) (page 35).
- (201) Kästner, J.; Thiel, W. *The Journal of Chemical Physics* **2005**, *123*, 144104, DOI: [10.1063/1.2052648](https://doi.org/10.1063/1.2052648) (page 35).
- (202) Kim, I.; Allen, T. W. *The Journal of Chemical Physics* **2012**, *136*, 164103, DOI: [10.1063/1.3701766](https://doi.org/10.1063/1.3701766) (page 35).
- (203) Chen, P. C.; Kuyucak, S. *Biophysical Journal* **2011**, *100*, 2466–2474, DOI: [10.1016/j.bpj.2011.03.052](https://doi.org/10.1016/j.bpj.2011.03.052) (page 35).
- (204) Bussi, G.; Laio, A.; Tiwary, P. In *Handbook of Materials Modeling*, Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2020, pp 565–595, DOI: [10.1007/978-3-319-44677-6_49](https://doi.org/10.1007/978-3-319-44677-6_49) (pages 36, 37).
- (205) Cheng, T.; Goddard, W. A.; An, Q.; Xiao, H.; Merinov, B.; Morozov, S. *Physical Chemistry Chemical Physics* **2017**, *19*, 2666–2673, DOI: [10.1039/C6CP08055C](https://doi.org/10.1039/C6CP08055C) (page 37).

- (206) Giberti, F.; Salvalaglio, M.; Parrinello, M. *IUCrJ* **2015**, *2*, 256–266, DOI: [10.1107/S2052252514027626](https://doi.org/10.1107/S2052252514027626) (page 37).
- (207) Bussi, G.; Laio, A. *Nature Reviews Physics* **2020**, *2*, 200–212, DOI: [10.1038/s42254-020-0153-0](https://doi.org/10.1038/s42254-020-0153-0) (pages 37, 38, 142).
- (208) Bussi, G.; Branduardi, D. In *Reviews in Computational Chemistry Volume 28*; John Wiley & Sons, Ltd: 2015, pp 1–49, DOI: [10.1002/9781118889886.ch1](https://doi.org/10.1002/9781118889886.ch1) (pages 37, 38, 142).
- (209) Sun, R.; Dama, J. F.; Tan, J. S.; Rose, J. P.; Voth, G. A. *Journal of Chemical Theory and Computation* **2016**, *12*, 5157–5169, DOI: [10.1021/acs.jctc.6b00206](https://doi.org/10.1021/acs.jctc.6b00206) (page 37).
- (210) Barducci, A.; Bussi, G.; Parrinello, M. *Physical Review Letters* **2008**, *100*, 020603, DOI: [10.1103/PhysRevLett.100.020603](https://doi.org/10.1103/PhysRevLett.100.020603) (page 37).
- (211) Tiwary, P.; Parrinello, M. *Physical Review Letters* **2013**, *111*, 230602, DOI: [10.1103/PhysRevLett.111.230602](https://doi.org/10.1103/PhysRevLett.111.230602) (page 37).
- (212) Tiwary, P.; Berne, B. J. *The Journal of Chemical Physics* **2016**, *144*, 134103, DOI: [10.1063/1.4944577](https://doi.org/10.1063/1.4944577) (page 37).
- (213) Salvalaglio, M.; Tiwary, P.; Parrinello, M. *Journal of Chemical Theory and Computation* **2014**, *10*, 1420–1425, DOI: [10.1021/ct500040r](https://doi.org/10.1021/ct500040r) (page 37).
- (214) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *The Journal of Physical Chemistry B* **2006**, *110*, 3533–3539, DOI: [10.1021/jp054359r](https://doi.org/10.1021/jp054359r) (page 37).
- (215) Ngo, V.; Li, H.; MacKerell, A. D.; Allen, T. W.; Roux, B.; Noskov, S. *Journal of Chemical Theory and Computation* **2021**, *17*, 1726–1741, DOI: [10.1021/acs.jctc.0c00968](https://doi.org/10.1021/acs.jctc.0c00968) (pages 38, 39).
- (216) Mamatkulov, S.; Fyta, M.; Netz, R. R. *The Journal of Chemical Physics* **2013**, *138*, 024505, DOI: [10.1063/1.4772808](https://doi.org/10.1063/1.4772808) (page 38).
- (217) Bergonzo, C.; Hall, K. B.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2016**, *12*, 3382–3389, DOI: [10.1021/acs.jctc.6b00173](https://doi.org/10.1021/acs.jctc.6b00173) (page 38).
- (218) Hollingsworth, S. A.; Dror, R. O. *Neuron* **2018**, *99*, 1129–1143, DOI: [10.1016/j.neuron.2018.08.011](https://doi.org/10.1016/j.neuron.2018.08.011) (page 38).
- (219) Melo, M. C. R.; Bernardi, R. C.; Rudack, T.; Scheurer, M.; Riplinger, C.; Phillips, J. C.; Maia, J. D. C.; Rocha, G. B.; Ribeiro, J. V.; Stone, J. E.; Neese, F.; Schulten, K.; Luthey-Schulten, Z. *Nature methods* **2018**, *15*, 351–354, DOI: [10.1038/nmeth.4638](https://doi.org/10.1038/nmeth.4638) (page 39).
- (220) Nerenberg, P. S.; Head-Gordon, T. *Current Opinion in Structural Biology* **2018**, *49*, 129–138, DOI: [10.1016/j.sbi.2018.02.002](https://doi.org/10.1016/j.sbi.2018.02.002) (page 39).
- (221) Unke, O. T.; Chmiela, S.; Saucedo, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. *Chemical Reviews* **2021**, *121*, 10142–10186, DOI: [10.1021/acs.chemrev.0c01111](https://doi.org/10.1021/acs.chemrev.0c01111) (page 39).
- (222) Lin, F.-Y.; Huang, J.; Pandey, P.; Rupakheti, C.; Li, J.; Roux, B.; MacKerell, A. D. *Journal of chemical theory and computation* **2020**, *16*, 3221–3239, DOI: [10.1021/acs.jctc.0c00057](https://doi.org/10.1021/acs.jctc.0c00057) (page 39).

- (223) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063, DOI: [10.1021/ct4003702](https://doi.org/10.1021/ct4003702) (page 39).
- (224) Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell, A. D.; Roux, B. *Journal of Chemical Theory and Computation* **2017**, *13*, 4535–4552, DOI: [10.1021/acs.jctc.7b00262](https://doi.org/10.1021/acs.jctc.7b00262) (page 39).
- (225) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. *The Journal of Physical Chemistry B* **2016**, *120*, 8313–8320, DOI: [10.1021/acs.jpcb.6b02024](https://doi.org/10.1021/acs.jpcb.6b02024) (page 39).
- (226) Spiwok, V.; Lipovová, P.; Králová, B. *Journal of Physical Chemistry B* **2007**, *111*, 3073–3076, DOI: [10.1021/jp068587c](https://doi.org/10.1021/jp068587c) (page 40).
- (227) TICA Theory http://docs.markovmodel.org/lecture_tica.html (accessed 05/28/2022) (pages 40, 132, 142, 146).
- (228) Schultze, S.; Grubmüller, H. *Journal of Chemical Theory and Computation* **2021**, *17*, 5766–5776, DOI: [10.1021/acs.jctc.1c00273](https://doi.org/10.1021/acs.jctc.1c00273) (pages 40, 142).
- (229) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. *Journal of Chemical Theory and Computation* **2019**, *15*, 743–750, DOI: [10.1021/acs.jctc.8b00934](https://doi.org/10.1021/acs.jctc.8b00934) (pages 40, 142).
- (230) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. *The Journal of Chemical Physics* **2018**, *149*, 072301, DOI: [10.1063/1.5025487](https://doi.org/10.1063/1.5025487) (pages 40, 142).
- (231) Van Rossum, G.; Drake Jr, F. L., *Python Reference Manual*; Centrum voor Wiskunde en Informatica Amsterdam: 1995 (page 40).
- (232) Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics* **1996**, *14*, 33–38, DOI: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5) (pages 40, 96).
- (233) Mallajosyula, S. S.; Jo, S.; Im, W.; Mackerell, A. D. *Methods in Molecular Biology* **2015**, *1273*, 407–429, DOI: [10.1007/978-1-4939-2343-4_25](https://doi.org/10.1007/978-1-4939-2343-4_25) (page 40).
- (234) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. *Journal of Computational Chemistry* **2011**, *32*, 2319–2327, DOI: [10.1002/jcc.21787](https://doi.org/10.1002/jcc.21787) (pages 40, 97).
- (235) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domański, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O. In *Proceedings of the 15th Python in Science Conference*, ed. by Benthall, S.; Rostrup, S., 2016, pp 98–105 (pages 40, 97).
- (236) Sali, A.; Blundell, T. L. *Journal of Molecular Biology* **1993**, *234*, 779–815, DOI: [10.1006/jmbi.1993.1626](https://doi.org/10.1006/jmbi.1993.1626) (pages 41, 96).
- (237) Shen, M.-y.; Sali, A. *Protein Science* **2006**, *15*, 2507–2524, DOI: [10.1110/ps.062416606](https://doi.org/10.1110/ps.062416606) (pages 41, 96).
- (238) Webb, B.; Sali, A. *Current Protocols in Bioinformatics* **2016**, *54*, DOI: [10.1002/cpbi.3](https://doi.org/10.1002/cpbi.3) (page 41).
- (239) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802, DOI: [10.1002/jcc.20289](https://doi.org/10.1002/jcc.20289) (page 41).

- (240) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. *PLoS Computational Biology* **2017**, 13, ed. by Gentleman, R., e1005659–e1005659, DOI: [10.1371/journal.pcbi.1005659](https://doi.org/10.1371/journal.pcbi.1005659) (page 41).
- (241) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Computer Physics Communications* **2014**, 185, 604–613, DOI: [10.1016/j.cpc.2013.09.018](https://doi.org/10.1016/j.cpc.2013.09.018) (pages 41, 127, 148).
- (242) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, 28, 235–242, DOI: [10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235) (page 41).
- (243) The UniProt Consortium et al. *Nucleic Acids Research* **2021**, 49, D480–D489, DOI: [10.1093/nar/gkaa1100](https://doi.org/10.1093/nar/gkaa1100) (page 41).
- (244) Amber22, in collab. with Case, D. et al., 2022 (page 41).
- (245) Ponder, J. W.; Case, D. A. In *Advances in Protein Chemistry; Protein Simulations*, Vol. 66; Academic Press: 2003, pp 27–85, DOI: [10.1016/S0065-3233\(03\)66002-X](https://doi.org/10.1016/S0065-3233(03)66002-X) (page 41).
- (246) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. *Journal of Computer-Aided Molecular Design* **2017**, 31, 147–161, DOI: [10.1007/s10822-016-9977-1](https://doi.org/10.1007/s10822-016-9977-1) (page 41).
- (247) Dick, K., director Sick: The Life and Death of Bob Flanagan, Supermasochist, in collab. with Flanagan, B.; Rose, S.; Burkhart, K., Documentary, 1997 (page 43).
- (248) Hajar, R. *Heart Views : The Official Journal of the Gulf Heart Association* **2017–2017**, 18, 154, DOI: [10.4103/HEARTVIEWS.HEARTVIEWS_131_17](https://doi.org/10.4103/HEARTVIEWS.HEARTVIEWS_131_17) (page 44).
- (249) Frank, C.; FRE, C. F. O., *Operation Epsilon: The Farm Hall Transcripts*; University of California Press: Berkeley, 1993; 272 pp. (page 44).
- (250) Gottfried, K. *Physics Today* **1999**, 52, 42–48, DOI: [10.1063/1.882612](https://doi.org/10.1063/1.882612) (page 44).
- (251) Global, A. E. M. S. P. Monthly Review | Why Socialism? Monthly Review, <https://monthlyreview.org/2009/05/01/why-socialism/> (accessed 07/18/2022) (page 44).
- (252) Rhodes, R., *The Making of the Atomic Bomb*, 25th Anniversary ed. edition; Simon & Schuster: New York, NY, 1986; 896 pp. (page 44).
- (253) Aaronson, S. Long-Dreaded Politics Post Shtetl-Optimized, <https://scottaaronson.blog/?p=314> (accessed 07/18/2022) (page 44).
- (254) Berger, J. *Terrorism and Counter-Terrorism Studies* **2016**, DOI: [10.19165/2016.1.11](https://doi.org/10.19165/2016.1.11) (page 44).
- (255) John von Neumann - World War II | Britannica <https://www.britannica.com/biography/John-von-Neumann/World-War-II> (accessed 07/18/2022) (page 44).
- (256) Buchanan, A.; Brock, D. W.; Daniels, N.; Wikler, D., *From Chance to Choice: Genetics and Justice*; Cambridge University Press: Cambridge, 2000, DOI: [10.1017/CBO9780511806940](https://doi.org/10.1017/CBO9780511806940) (page 45).

- (257) Taneri, B. *Journal of Computational Biology* **2011**, *18*, 907–916, DOI: [10.1089/cmb.2010.0187](https://doi.org/10.1089/cmb.2010.0187) (page 45).
- (258) National Academies of Sciences, *Human Genome Editing: Science, Ethics and Governance*; National Academies Press: 2017; 329 pp. (page 45).
- (259) Müller, V. C. In *The Stanford Encyclopedia of Philosophy*, Zalta, E. N., Ed., Summer 2021; Metaphysics Research Lab, Stanford University: 2021 (page 45).
- (260) Bostrom, N., *Superintelligence: Paths, Dangers, Strategies*, Reprint edition; OUP Oxford: 2014; 431 pp. (pages 45, 171).
- (261) Roose, K. *The New York Times, Technology* **2022** (page 45).
- (262) Mallapaty, S. *Nature* **2022**, *603*, 213–214, DOI: [10.1038/d41586-022-00512-w](https://doi.org/10.1038/d41586-022-00512-w) (pages 45, 63).
- (263) Urbina, F.; Lentzos, F.; Invernizzi, C.; Ekins, S. *Nature Machine Intelligence* **2022**, *4*, 189–191, DOI: [10.1038/s42256-022-00465-9](https://doi.org/10.1038/s42256-022-00465-9) (page 45).
- (264) Hammoudeh, S.; Gadelhaq, W.; Hani, Y.; Omar, N.; El Dimassi, D.; Elizabeth, C.; Pullattayil, A. K.; Chandra, P.; Janahi, I. A. *SN Comprehensive Clinical Medicine* **2021**, *3*, 490–498, DOI: [10.1007/s42399-021-00756-z](https://doi.org/10.1007/s42399-021-00756-z) (page 45).
- (265) Guo, J.; Garratt, A.; Hill, A. *Journal of Cystic Fibrosis* **2022**, *21*, 456–462, DOI: [10.1016/j.jcf.2022.01.009](https://doi.org/10.1016/j.jcf.2022.01.009) (pages 45, 129).
- (266) McBennett, K. A.; Davis, P. B.; Konstan, M. W. *Pediatric pulmonology* **2022**, *57*, S5–S12, DOI: [10.1002/ppul.25733](https://doi.org/10.1002/ppul.25733) (pages 45, 46).
- (267) Van Gool, K.; Norman, R.; Delatycki, M. B.; Hall, J.; Massie, J. *Value in Health* **2013**, *16*, 345–355, DOI: [10.1016/j.jval.2012.12.003](https://doi.org/10.1016/j.jval.2012.12.003) (page 45).
- (268) Page, A.; Goldenberg, A.; Matthews, A. L. *BMC Pulmonary Medicine* **2022**, *22*, 42, DOI: [10.1186/s12890-022-01825-2](https://doi.org/10.1186/s12890-022-01825-2) (page 45).
- (269) Reddy, M.; Stutts, M. J. *Cold Spring Harbor Perspectives in Medicine* **2013**, *3*, a009555, DOI: [10.1101/cshperspect.a009555](https://doi.org/10.1101/cshperspect.a009555) (page 45).
- (270) Boucher, R. C. Airway Surface Dehydration in Cystic Fibrosis: Pathogenesis and Therapy, 2007, DOI: [10.1146/annurev.med.58.071905.105316](https://doi.org/10.1146/annurev.med.58.071905.105316) (page 45).
- (271) Szczesniak, R.; Heltshe, S. L.; Stanojevic, S.; Mayer-Hamblett, N. *Journal of cystic fibrosis : official journal of the European Cystic Fibrosis Society* **2017**, *16*, 318–326, DOI: [10.1016/j.jcf.2017.01.002](https://doi.org/10.1016/j.jcf.2017.01.002) (pages 45, 46).
- (272) Chiappini, E.; Taccetti, G.; de Martino, M. *The Pediatric Infectious Disease Journal* **2014**, *33*, 653–654, DOI: [10.1097/INF.0000000000000347](https://doi.org/10.1097/INF.0000000000000347) (page 46).
- (273) Krouse, M. E. *The Journal of General Physiology* **2001**, *118*, 219–222 (page 46).
- (274) Conway, S. *Journal of the Royal Society of Medicine* **2008**, *101*, 31–35, DOI: [10.1258/jrsm.2008.s18007](https://doi.org/10.1258/jrsm.2008.s18007) (page 46).
- (275) Baldoni, J., director Five Feet Apart, scriptwriter Daughtry, M.; Iaconis, T.; Richardson, H. L.; Sprouse, C.; Arias, M., Drama, Romance, 2019 (page 46).
- (276) Chest Physical Therapy | Cystic Fibrosis Foundation <https://www.cff.org/managing-cf/chest-physical-therapy> (accessed 07/19/2022) (page 46).

- (277) The Frey Life, director HOW TO DO MANUAL CHEST PT (Airway Clearance), 2015 (page 46).
- (278) Wark, P.; McDonald, V. M. *The Cochrane Database of Systematic Reviews* **2018**, *9*, CD001506, DOI: [10.1002/14651858.CD001506.pub4](https://doi.org/10.1002/14651858.CD001506.pub4) (page 46).
- (279) Kumar, V.; Abbas, A. K.; Aster, J. C.; Robbins, S. L.; Perkins, J. A., *Robbins Basic Pathology*, Tenth edition; Elsevier: Philadelphia, Pennsylvania, 2018; 935 pp. (page 46).
- (280) Adler, F. R. et al. *Proceedings of the American Thoracic Society* **2009**, *6*, 619–633, DOI: [10.1513/pats.2009008-088TL](https://doi.org/10.1513/pats.2009008-088TL) (page 46).
- (281) Blatter, J.; Sweet, S. *Pediatric Allergy, Immunology, and Pulmonology* **2015**, *28*, 237–243, DOI: [10.1089/ped.2015.0564](https://doi.org/10.1089/ped.2015.0564) (page 46).
- (282) Weill, D. *Journal of Thoracic Disease* **2018**, *10*, 4574–4587, DOI: [10.21037/jtd.2018.06.141](https://doi.org/10.21037/jtd.2018.06.141) (page 46).
- (283) Kayani, K.; Mohammed, R.; Mohiaddin, H. Cystic Fibrosis-Related Diabetes, 2018, DOI: [10.3389/fendo.2018.00020](https://doi.org/10.3389/fendo.2018.00020) (page 46).
- (284) Sullivan, J. S.; Mascarenhas, M. R. *Journal of Cystic Fibrosis* **2017**, *16*, S87–S93, DOI: [10.1016/j.jcf.2017.07.010](https://doi.org/10.1016/j.jcf.2017.07.010) (page 46).
- (285) Halloran, C. M.; Cox, T. F.; Chauhan, S.; Raraty, M. G. T.; Sutton, R.; Neoptolemos, J. P.; Ghaneh, P. *Pancreatology: official journal of the International Association of Pancreatology (IAP) ... [et al.]* **2011**, *11*, 535–545, DOI: [10.1159/000333308](https://doi.org/10.1159/000333308) (page 46).
- (286) Singh, V. K.; Schwarzenberg, S. J. *Journal of Cystic Fibrosis* **2017**, *16*, S70–S78, DOI: [10.1016/j.jcf.2017.06.011](https://doi.org/10.1016/j.jcf.2017.06.011) (page 46).
- (287) E Garcia, L. d. C.; Petry, L. M.; Germani, P. A. V. D. S.; Xavier, L. F.; de Barros, P. B.; Meneses, A. d. S.; Prestes, L. M.; Bittencourt, L. B.; Pieta, M. P.; Friedrich, F.; Pinto, L. A. *Frontiers in Pediatrics* **2022**, *10*, 881470, DOI: [10.3389/fped.2022.881470](https://doi.org/10.3389/fped.2022.881470) (pages 47, 64).
- (288) Roberts, G. B. S. Fibrocystic Disease of the Pancreas: A Histological Study of the Pancreas in Childhood with a Survey of the Clinical, Epidemiological, and Genetic Features of the Disease, MD, Ann Arbor : ProQuest Dissertations & Theses, 1957, 731 pp. (pages 47, 71).
- (289) Levy, E. *Clinical Biochemistry* **2011**, *44*, 489–490, DOI: [10.1016/j.clinbiochem.2011.03.126](https://doi.org/10.1016/j.clinbiochem.2011.03.126) (page 47).
- (290) Burns, J. L.; Gibson, R. L.; McNamara, S.; Yim, D.; Emerson, J.; Rosenfeld, M.; Hiatt, P.; McCoy, K.; Castile, R.; Smith, A. L.; Ramsey, B. W. *The Journal of Infectious Diseases* **2001**, *183*, 444–452, DOI: [10.1086/318075](https://doi.org/10.1086/318075) (page 47).
- (291) Stalvey, M. S.; Clines, G. A. *Current opinion in endocrinology, diabetes, and obesity* **2013**, *20*, 547–552, DOI: [10.1097/01.med.0000436191.87727.ec](https://doi.org/10.1097/01.med.0000436191.87727.ec) (page 47).
- (292) Popli, K.; Stewart, J. *Human Fertility (Cambridge, England)* **2007**, *10*, 217–221, DOI: [10.1080/14647270701510033](https://doi.org/10.1080/14647270701510033) (page 47).

- (293) Thomas, C. et al. *FEBS Letters* **2020**, *594*, 3767–3775, DOI: [10.1002/1873-3468.13935](https://doi.org/10.1002/1873-3468.13935) (pages 47, 51).
- (294) Gadsby, D. C.; Vergani, P.; Csanády, L. *Nature* **2006**, *440*, 477–483, DOI: [10.1038/nature04712](https://doi.org/10.1038/nature04712) (pages 47, 48, 129).
- (295) Linsdell, P. *Channels (Austin, Tex.)* **2018**, *12*, 284–290, DOI: [10.1080/19336950.2018.1502585](https://doi.org/10.1080/19336950.2018.1502585) (pages 47, 49, 53, 112, 131, 132, 139, 168).
- (296) Infield, D. T.; Strickland, K. M.; Gaggar, A.; McCarty, N. A. *Journal of General Physiology* **2021**, *153*, e202012625, DOI: [10.1085/jgp.202012625](https://doi.org/10.1085/jgp.202012625) (pages 48, 51, 52).
- (297) Zhang, X. C.; Yang, H.; Liu, Z.; Sun, F. *Biophysics Reports* **2018**, *4*, 300–319, DOI: [10.1007/s41048-018-0074-y](https://doi.org/10.1007/s41048-018-0074-y) (pages 48, 49, 53, 55, 112, 119, 126, 129, 130, 146, 165).
- (298) Mihályi, C.; Iordanov, I.; Töröcsik, B.; Csanády, L. *Proceedings of the National Academy of Sciences* **2020**, 202007910–202007910, DOI: [10.1073/pnas.2007910117](https://doi.org/10.1073/pnas.2007910117) (pages 48, 49, 54, 119).
- (299) Cormet-Boyaka, E.; Di, A.; Chang, S. Y.; Naren, A. P.; Tousson, A.; Nelson, D. J.; Kirk, K. L. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 12477–12482, DOI: [10.1073/pnas.192203899](https://doi.org/10.1073/pnas.192203899) (page 49).
- (300) Naren, A. P.; Quick, M. W.; Collawn, J. F.; Nelson, D. J.; Kirk, K. L. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 10972–10977, DOI: [10.1073/pnas.95.18.10972](https://doi.org/10.1073/pnas.95.18.10972) (page 49).
- (301) Thelin, W. R.; Chen, Y.; Gentzsch, M.; Kreda, S. M.; Sallee, J. L.; Scarlett, C. O.; Borchers, C. H.; Jacobson, K.; Stutts, M. J.; Milgram, S. L. *Journal of Clinical Investigation* **2007**, *117*, 364–374, DOI: [10.1172/JCI30376](https://doi.org/10.1172/JCI30376) (pages 49, 74, 158).
- (302) Linsdell, P. *Experimental Physiology* **2006**, *91*, 123–129, DOI: [10.1113/expphysiol.2005.031757](https://doi.org/10.1113/expphysiol.2005.031757) (page 49).
- (303) Linsdell, P.; Irving, C. L.; Cowley, E. A. *Journal of Biological Chemistry* **2022**, *298*, DOI: [10.1016/j.jbc.2022.101659](https://doi.org/10.1016/j.jbc.2022.101659) (page 49).
- (304) Wong, S. L. et al. *American journal of respiratory cell and molecular biology* **2022**, DOI: [10.1165/rcmb.2021-03370C](https://doi.org/10.1165/rcmb.2021-03370C) (pages 49, 65, 99, 123, 124, 129, 140, 154, 155).
- (305) Kim, Y.; Jun, I.; Shin, D. H.; Yoon, J. G.; Piao, H.; Jung, J.; Park, H. W.; Cheng, M. H.; Bahar, I.; Whitcomb, D. C.; Lee, M. G. *Cellular and Molecular Gastroenterology and Hepatology* **2019**, *9*, 79–103, DOI: [10.1016/j.jcmgh.2019.09.003](https://doi.org/10.1016/j.jcmgh.2019.09.003) (pages 49, 56, 131, 141).
- (306) Ostvedgaard, L. S.; Baldursson, O.; Vermeer, D. W.; Welsh, M. J.; Robertson, A. D. *Proceedings of the National Academy of Sciences* **2000**, *97*, 5657–5662, DOI: [10.1073/pnas.100588797](https://doi.org/10.1073/pnas.100588797) (page 49).

- (307) Bozoky, Z.; Krzeminski, M.; Muhandiram, R.; Birtley, J. R.; Al-Zahrani, A.; Thomas, P. J.; Frizzell, R. A.; Ford, R. C.; Forman-Kay, J. D. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, E4427–E4436, DOI: [10.1073/pnas.1315104110](https://doi.org/10.1073/pnas.1315104110) (pages 49, 96).
- (308) Baker, J. M.; Hudson, R. P.; Kanelis, V.; Choy, W. Y.; Thibodeau, P. H.; Thomas, P. J.; Forman-Kay, J. D. *Nature Structural and Molecular Biology* **2007**, *14*, 738–745, DOI: [10.1038/nsmb1278](https://doi.org/10.1038/nsmb1278) (pages 49, 85).
- (309) Wong, S. L.; Awatade, N. T.; Astore, M. A.; Allan, K. M.; Carnell, M. J.; Slapetova, I.; Chen, P.-c.; Capraro, A.; Fawcett, L. K.; Whan, R. M.; Griffith, R.; Ooi, C. Y.; Kuyucak, S.; Jaffe, A.; Waters, S. A. *iScience* **2022**, *25*, 103710, DOI: [10.1016/j.isci.2021.103710](https://doi.org/10.1016/j.isci.2021.103710) (pages 49, 65, 72, 110, 123, 124, 126, 127, 154, 155, 158, 159).
- (310) Hegedűs, T.; Geisler, M.; Lukács, G. L.; Farkas, B. *Cellular and molecular life sciences: CMLS* **2022**, *79*, 73, DOI: [10.1007/s00018-021-04112-1](https://doi.org/10.1007/s00018-021-04112-1) (pages 49, 51).
- (311) Liu, F.; Zhang, Z.; Levit, A.; Levring, J.; Touhara, K. K.; Shoichet, B. K.; Chen, J. *Science (New York, N.Y.)* **2019**, *364*, 1184–1188, DOI: [10.1126/science.aaw7611](https://doi.org/10.1126/science.aaw7611) (pages 49, 51, 60–62, 86, 120, 129, 167, 168).
- (312) Negoda, A.; Hogan, M. S.; Cowley, E. A.; Linsdell, P. *Cellular and Molecular Life Sciences* **2019**, *76*, 2411–2423, DOI: [10.1007/s00018-019-03043-2](https://doi.org/10.1007/s00018-019-03043-2) (pages 49–51, 120, 131, 132, 141).
- (313) Ivey, G.; Youker, R. T. *PLoS ONE* **2020**, *15*, e0227668–e0227668, DOI: [10.1371/journal.pone.0227668](https://doi.org/10.1371/journal.pone.0227668) (page 49).
- (314) Zolnerciks, J. K.; Akkaya, B. G.; Snippe, M.; Chiba, P.; Seelig, A.; Linton, K. J. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* **2014**, *28*, 4335–4346, DOI: [10.1096/fj.13-245639](https://doi.org/10.1096/fj.13-245639) (page 49).
- (315) Dong, Q.; Ernst, S. E.; Ostvedgaard, L. S.; Shah, V. S.; Ver Heul, A. R.; Welsh, M. J.; Randak, C. O. *The Journal of Biological Chemistry* **2015**, *290*, 14140–14153, DOI: [10.1074/jbc.M114.611616](https://doi.org/10.1074/jbc.M114.611616) (pages 49, 161, 162).
- (316) Moyer, B. D.; Demon, J.; Karlson, K. H.; Reynolds, D.; Wang, S.; Mickle, J. E.; Milewski, M.; Cutting, G. R.; Guggino, W. B.; Li, M.; Stanton, B. A. *Journal of Clinical Investigation* **1999**, *104*, 1353–1361, DOI: [10.1172/JCI7453](https://doi.org/10.1172/JCI7453) (page 49).
- (317) Cushing, P. R.; Fellows, A.; Villone, D.; Boisguérin, P.; Madden, D. R. *Biochemistry* **2008**, *47*, 10084–10098, DOI: [10.1021/bi8003928](https://doi.org/10.1021/bi8003928) (page 49).
- (318) Yeh, J.; Yu, Y.; Hwang, T. *The Journal of Physiology* **2019**, *597*, 543–560, DOI: [10.1113/JP277042](https://doi.org/10.1113/JP277042) (pages 49, 158).
- (319) Liu, F.; Zhang, Z.; Csanády, L.; Gadsby, D. C.; Chen, J. *Cell* **2017**, *169*, 85–95.e8, DOI: [10.1016/j.cell.2017.02.024](https://doi.org/10.1016/j.cell.2017.02.024) (pages 49, 51, 74, 165).
- (320) Zhang, Z.; Liu, F.; Chen, J. *Cell* **2017**, *170*, 483–491.e8, DOI: [10.1016/j.cell.2017.06.041](https://doi.org/10.1016/j.cell.2017.06.041) (pages 50, 51, 129, 131).

- (321) Dawson, R. J. P.; Locher, K. P. *Nature* **2006**, *443*, 180–185, DOI: [10.1038/nature05155](https://doi.org/10.1038/nature05155) (pages 50, 131).
- (322) Hoffmann, B.; Elbahnsi, A.; Lehn, P.; Décout, J. L.; Pietrucci, F.; Mornon, J. P.; Callebaut, I. *Cellular and Molecular Life Sciences* **2018**, *75*, 3829–3855, DOI: [10.1007/s00018-018-2835-7](https://doi.org/10.1007/s00018-018-2835-7) (pages 50, 74, 131–133, 146, 148, 158).
- (323) Norimatsu, Y.; Ivetac, A.; Alexander, C.; Kirkham, J.; O'Donnell, N.; Dawson, D. C.; Sansom, M. S. *Biochemistry* **2012**, *51*, 2199–2212, DOI: [10.1021/bi201888a](https://doi.org/10.1021/bi201888a) (page 50).
- (324) Laselva, O.; Erwood, S.; Du, K.; Ivakine, Z.; Bear, C. E. *FASEB bioAdvances* **2019**, *1*, 661–670, DOI: [10.1096/fba.2019-00039](https://doi.org/10.1096/fba.2019-00039) (page 50).
- (325) Linsdell, P. *Biochimica Et Biophysica Acta* **2016**, *1858*, 740–747, DOI: [10.1016/j.bbamem.2016.01.009](https://doi.org/10.1016/j.bbamem.2016.01.009) (pages 50, 54, 129, 131, 137, 139, 141).
- (326) Linsdell, P. *Cellular and Molecular Life Sciences* **2017**, *74*, 67–83, DOI: [10.1007/s00018-016-2389-5](https://doi.org/10.1007/s00018-016-2389-5) (page 50).
- (327) Linsdell, P. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2021**, *1863*, 183558, DOI: [10.1016/j.bbamem.2021.183558](https://doi.org/10.1016/j.bbamem.2021.183558) (pages 50, 129).
- (328) Li, M. S.; Cowley, E. A.; El Hiani, Y.; Linsdell, P. *Journal of Biological Chemistry* **2018**, *293*, 5649–5658, DOI: [10.1074/jbc.RA117.001373](https://doi.org/10.1074/jbc.RA117.001373) (page 50).
- (329) Linsdell, P.; Negoda, A.; Cowley, E. A.; El Hiani, Y. *Cell Biochemistry and Biophysics* **2020**, *78*, 15–22, DOI: [10.1007/s12013-019-00899-w](https://doi.org/10.1007/s12013-019-00899-w) (page 50).
- (330) Farkas, B.; Tordai, H.; Padányi, R.; Tordai, A.; Gera, J.; Paragi, G.; Hegedűs, T. *Cellular and Molecular Life Sciences* **2020**, *77*, 765–778, DOI: [10.1007/s00018-019-03211-4](https://doi.org/10.1007/s00018-019-03211-4) (pages 50, 129, 131, 139).
- (331) Jun, I.; Cheng, M. H.; Sim, E.; Jung, J.; Suh, B. L.; Kim, Y.; Son, H.; Park, K.; Kim, C. H.; Yoon, J. H.; Whitcomb, D. C.; Bahar, I.; Lee, M. G. *Journal of Physiology* **2016**, *594*, 2929–2955, DOI: [10.1113/JP271311](https://doi.org/10.1113/JP271311) (pages 50, 56, 131, 134, 141).
- (332) Gong, X.; Linsdell, P. *Archives of Biochemistry and Biophysics* **2004**, *426*, 78–82, DOI: [10.1016/j.abb.2004.03.033](https://doi.org/10.1016/j.abb.2004.03.033) (pages 50, 131, 135, 137, 139, 142–144, 158).
- (333) Zeng, Z. W.; Linsdell, P.; Pomès, R. *Molecular Dynamics Study of Cl- Permeation through Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)*; preprint; In Review, 2021, DOI: [10.21203/rs.3.rs-771352/v1](https://doi.org/10.21203/rs.3.rs-771352/v1) (pages 50, 129, 131, 139, 142).
- (334) Kogan, I.; Ramjeesingh, M.; Li, C.; Kidd, J. F.; Wang, Y.; Leslie, E. M.; Cole, S. P.; Bear, C. E. *The EMBO Journal* **2003**, *22*, 1981–1989, DOI: [10.1093/emboj/cdg194](https://doi.org/10.1093/emboj/cdg194) (pages 50, 131).
- (335) Linsdell, P.; Hanrahan, J. W. *American Journal of Physiology-Cell Physiology* **1998**, *275*, C323–C326, DOI: [10.1152/ajpcell.1998.275.1.C323](https://doi.org/10.1152/ajpcell.1998.275.1.C323) (pages 50, 54, 131).
- (336) Corradi, V.; Gu, R. X.; Vergani, P.; Tielemans, D. P. *Biophysical Journal* **2018**, *114*, 1751–1754, DOI: [10.1016/j.bpj.2018.03.003](https://doi.org/10.1016/j.bpj.2018.03.003) (pages 51, 120).

- (337) Fiedorczuk, K.; Chen, J. *bioRxiv* **2021**, 2021.06.18.449063, DOI: [10.1101/2021.06.18.449063](https://doi.org/10.1101/2021.06.18.449063) (page 51).
- (338) Zhang, Z.; Chen, J. *Cell* **2016**, *167*, 1586–1597.e9, DOI: [10.1016/j.cell.2016.11.014](https://doi.org/10.1016/j.cell.2016.11.014) (pages 51, 74, 87).
- (339) Fay, J. F.; Aleksandrov, L. A.; Jensen, T. J.; Cui, L. L.; Kousouros, J. N.; He, L.; Aleksandrov, A. A.; Gingerich, D. S.; Riordan, J. R.; Chen, J. Z. *Biochemistry* **2018**, *57*, 6234–6246, DOI: [10.1021/acs.biochem.8b00763](https://doi.org/10.1021/acs.biochem.8b00763) (pages 51, 54).
- (340) Aleksandrov, L. A.; Aleksandrov, A. A.; Jensen, T. J.; Strauss, J. D.; Fay, J. F. *International Journal of Molecular Sciences* **2022**, *23*, 9248, DOI: [10.3390/ijms23169248](https://doi.org/10.3390/ijms23169248) (page 51).
- (341) Stratford, F. L.; Ramjeesingh, M.; Cheung, J. C.; Huan, L. J.; Bear, C. E. *Biochemical Journal* **2007**, *401*, 581–586, DOI: [10.1042/BJ20060968](https://doi.org/10.1042/BJ20060968) (pages 52, 53).
- (342) delToro, D.; Ortiz, D.; Ordyan, M.; Sippy, J.; Oh, C.-S.; Keller, N.; Feiss, M.; Catalano, C. E.; Smith, D. *Journal of molecular biology* **2016**, *428*, 2709–2729, DOI: [10.1016/j.jmb.2016.04.029](https://doi.org/10.1016/j.jmb.2016.04.029) (page 52).
- (343) Urbatsch, I. L.; Julien, M.; Carrier, I.; Rousseau, M.-E.; Cayrol, R.; Gros, P. *Biochemistry* **2000**, *39*, 14138–14149, DOI: [10.1021/bi001128w](https://doi.org/10.1021/bi001128w) (page 53).
- (344) Rai, V.; Gaur, M.; Shukla, S.; Shukla, S.; Ambudkar, S. V.; Komath, S. S.; Prasad, R. *Biochemistry* **2006**, *45*, 14726–14739, DOI: [10.1021/bi061535t](https://doi.org/10.1021/bi061535t) (page 53).
- (345) Gout, T. *Annals of Thoracic Medicine* **2012**, *7*, 115–121, DOI: [10.4103/1817-1737.98842](https://doi.org/10.4103/1817-1737.98842) (page 53).
- (346) Yeh, H.-I.; Sutcliffe, K. J.; Sheppard, D. N.; Hwang, T.-C. *Handbook of Experimental Pharmacology* **2022**, DOI: [10.1007/164_2022_597](https://doi.org/10.1007/164_2022_597) (page 53).
- (347) Ramjeesingh, M.; Li, C.; Garami, E.; Huan, L.-J.; Galley, K.; Wang, Y.; Bear, C. E. *Biochemistry* **1999**, *38*, 1463–1468, DOI: [10.1021/bi982243y](https://doi.org/10.1021/bi982243y) (page 53).
- (348) Muallem, D.; Vergani, P. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2009**, *364*, 247–255, DOI: [10.1098/rstb.2008.0191](https://doi.org/10.1098/rstb.2008.0191) (page 53).
- (349) Hwang, T. C.; Kirk, K. L. *Cold Spring Harbor Perspectives in Medicine* **2013**, *3*, a009498–a009498, DOI: [10.1101/cshperspect.a009498](https://doi.org/10.1101/cshperspect.a009498) (pages 53, 74).
- (350) Jih, K.-Y.; Sohma, Y.; Hwang, T.-C. *Journal of General Physiology* **2012**, *140*, 347–359, DOI: [10.1085/jgp.201210834](https://doi.org/10.1085/jgp.201210834) (pages 53, 54, 68).
- (351) Tabcharani, J. A.; Linsdell, P.; Hanrahan, J. W. *The Journal of General Physiology* **1997**, *110*, 341–354 (page 54).
- (352) Poulsen, J. H.; Fischer, H.; Illek, B.; Machen, T. E. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 5340–5344, DOI: [10.1073/pnas.91.12.5340](https://doi.org/10.1073/pnas.91.12.5340) (page 54).
- (353) Linsdell, P.; Evangelidis, A.; Hanrahan, J. W. *Biophysical Journal* **2000**, *78*, 2973–2982, DOI: [10.1016/S0006-3495\(00\)76836-6](https://doi.org/10.1016/S0006-3495(00)76836-6) (page 54).

- (354) Yang, K.-L.; Yiacoumi, S.; Tsouris, C. *The Journal of Chemical Physics* **2002**, *117*, 8499–8507, DOI: [10.1063/1.1511726](https://doi.org/10.1063/1.1511726) (page 54).
- (355) Wei, S.; Roessler, B. C.; Icyuz, M.; Chauvet, S.; Tao, B.; Iv, J. L. H.; Kirk, K. L. *The FASEB Journal* **2016**, *30*, 1247–1262, DOI: [10.1096/fj.15-278382](https://doi.org/10.1096/fj.15-278382) (page 54).
- (356) Linsdell, P. *The Journal of Physiology* **2001**, *531*, 51–66, DOI: [10.1111/j.1469-7793.2001.0051j.x](https://doi.org/10.1111/j.1469-7793.2001.0051j.x) (pages 56, 131).
- (357) Ge, N.; Muise, C. N.; Gong, X.; Linsdell, P. *Journal of Biological Chemistry* **2004**, *279*, 55283–55289, DOI: [10.1074/jbc.M411935200](https://doi.org/10.1074/jbc.M411935200) (page 56).
- (358) Tang, L.; Fatehi, M.; Linsdell, P. *Journal of Cystic Fibrosis* **2009**, *8*, 115–121, DOI: [10.1016/j.jcf.2008.10.004](https://doi.org/10.1016/j.jcf.2008.10.004) (pages 56, 131).
- (359) Garnett, J. P.; Hickman, E.; Burrows, R.; Hegyi, P.; Tiszlavicz, L.; Cuthbert, A. W.; Fong, P.; Gray, M. A. *The Journal of Biological Chemistry* **2011**, *286*, 41069–41082, DOI: [10.1074/jbc.M111.266734](https://doi.org/10.1074/jbc.M111.266734) (page 56).
- (360) Kim, D.; Huang, J.; Billet, A.; Abu-Arish, A.; Goepp, J.; Matthes, E.; Tewfik, M. A.; Frenkiel, S.; Hanrahan, J. W. *American Journal of Respiratory Cell and Molecular Biology* **2019**, *60*, 705–716, DOI: [10.1165/rcmb.2018-01580C](https://doi.org/10.1165/rcmb.2018-01580C) (page 56).
- (361) Ferrera, L.; Capurro, V.; Delpiano, L.; Gianotti, A.; Moran, O. *Biology* **2021**, *10*, 278, DOI: [10.3390/biology10040278](https://doi.org/10.3390/biology10040278) (page 56).
- (362) Crawford, K. J.; Downey, D. G. *Current Opinion in Pulmonary Medicine* **2018**, *24*, 612–617, DOI: [10.1097/MCP.0000000000000521](https://doi.org/10.1097/MCP.0000000000000521) (page 56).
- (363) Sette, G.; Cicero, S. L.; Blaconà, G.; Pierandrei, S.; Bruno, S. M.; Salvati, V.; Castelli, G.; Falchi, M.; Fabrizzi, B.; Cimino, G.; Maria, R. D.; Biffoni, M.; Eramo, A.; Lucarelli, M. *European Respiratory Journal* **2021**, DOI: [10.1183/13993003.00908-2021](https://doi.org/10.1183/13993003.00908-2021) (page 56).
- (364) Ioannou, L.; McClaren, B. J.; Massie, J.; Lewis, S.; Metcalfe, S. A.; Forrest, L.; Delatycki, M. B. *Genetics in Medicine* **2014**, *16*, 207–216, DOI: [10.1038/gim.2013.125](https://doi.org/10.1038/gim.2013.125) (page 57).
- (365) Miller, A. C.; Comellas, A. P.; Hornick, D. B.; Stoltz, D. A.; Cavanaugh, J. E.; Gerke, A. K.; Welsh, M. J.; Zabner, J.; Polgreen, P. M. *Proceedings of the National Academy of Sciences* **2020**, *117*, 1621–1627, DOI: [10.1073/pnas.1914912117](https://doi.org/10.1073/pnas.1914912117) (page 57).
- (366) Veit, G. et al. *Molecular Biology of the Cell* **2016**, *27*, 424–433, DOI: [10.1091/mbc.E14-04-0935](https://doi.org/10.1091/mbc.E14-04-0935) (pages 58, 119, 157, 166).
- (367) Belmonte, L.; Moran, O. *Biochimie* **2015**, *111*, 19–29, DOI: [10.1016/j.biochi.2015.01.010](https://doi.org/10.1016/j.biochi.2015.01.010) (page 58).
- (368) Bahia, M. S.; Khazanov, N.; Zhou, Q.; Yang, Z.; Wang, C.; Hong, J. S.; Rab, A.; Sorscher, E. J.; Brouillette, C. G.; Hunt, J. F.; Senderowitz, H. *Journal of Chemical Information and Modeling* **2021**, *61*, 1762–1777, DOI: [10.1021/acs.jcim.0c01207](https://doi.org/10.1021/acs.jcim.0c01207) (pages 58, 159, 166).

- (369) Atwell, S. et al. *Protein Engineering, Design and Selection* **2010**, *23*, 375–384, DOI: [10.1093/protein/gzq004](https://doi.org/10.1093/protein/gzq004) (page 58).
- (370) Wang, C., Vorobiev, S.M., Vernon, R.M., Khazanov, N., Senderowitz, H., Forman-Kay, J.D., Hunt, Wang, C., Vorobiev, S.M., Vernon, R.M., Khazanov, N., Senderowitz, H., Forman-Kay, J.D., Hunt, Crystal Structure of Nucleotide-Binding Domain 2 (NBD2) of the Human Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) (page 58).
- (371) Billet, A.; Elbahnsi, A.; Jollivet-Souchet, M.; Hoffmann, B.; Mornon, J.-P. P.; Callebaut, I.; Becq, F. *Frontiers in Pharmacology* **2020**, *11*, 295–295, DOI: [10.3389/fphar.2020.00295](https://doi.org/10.3389/fphar.2020.00295) (pages 58, 116, 123, 124).
- (372) Fiedorczuk, K.; Chen, J. *Cell* **2022**, *185*, 158–168.e11, DOI: [10.1016/j.cell.2021.12.009](https://doi.org/10.1016/j.cell.2021.12.009) (pages 58–61, 119, 159).
- (373) Sabusap, C. M.; Joshi, D.; Simhaev, L.; Oliver, K. E.; Senderowitz, H.; van Willigen, M.; Braakman, I.; Rab, A.; Sorscher, E. J.; Hong, J. S. *The Journal of Biological Chemistry* **2021 Jan-Jun**, *296*, 100598, DOI: [10.1016/J.JBC.2021.100598](https://doi.org/10.1016/J.JBC.2021.100598) (pages 58, 74, 87, 116, 123, 124).
- (374) Gené, G. G.; Llobet, A.; Larriba, S.; de Semir, D.; Martínez, I.; Escalada, A.; Solsona, C.; Casals, T.; Aran, J. M. J. *Human Mutation* **2008**, *29*, 738–749, DOI: [10.1002/humu.20721](https://doi.org/10.1002/humu.20721) (pages 59, 74, 158).
- (375) Yeh, H. I.; Qiu, L.; Sohma, Y.; Conrath, K.; Zou, X.; Hwang, T. C. *Journal of General Physiology* **2019**, *151*, 912–928, DOI: [10.1085/jgp.201912360](https://doi.org/10.1085/jgp.201912360) (pages 60–62, 86, 167, 168).
- (376) Csanády, L.; Töröcsik, B. *eLife* **2019**, *8*, DOI: [10.7554/eLife.46450.001](https://doi.org/10.7554/eLife.46450.001) (pages 60–62, 141, 159, 167, 168).
- (377) Laselva, O.; Petrotchenko, E. V.; Hamilton, C. M.; Qureshi, Z.; Borchers, C. H.; Young, R. N.; Bear, C. E. *STAR Protocols* **2022**, *3*, 101258, DOI: [10.1016/j.xpro.2022.101258](https://doi.org/10.1016/j.xpro.2022.101258) (pages 61, 62, 166, 168).
- (378) Yeh, H. I.; Sohma, Y.; Conrath, K.; Hwang, T. C. *Journal of General Physiology* **2017**, *149*, 1105–1118, DOI: [10.1085/JGP.201711886](https://doi.org/10.1085/JGP.201711886) (pages 61, 62, 86, 141, 167).
- (379) Krainer, G.; Treff, A.; Hartmann, A.; Stone, T. A.; Schenkel, M.; Keller, S.; Deber, C. M.; Schlierf, M. *Communications Biology* **2018**, *1*, 1–7, DOI: [10.1038/s42003-018-0153-0](https://doi.org/10.1038/s42003-018-0153-0) (pages 61, 159).
- (380) Molinski, S. V.; Shahani, V. M.; Subramanian, A. S.; MacKinnon, S. S.; Woollard, G.; Laforet, M.; Laselva, O.; Morayniss, L. D.; Bear, C. E.; Windemuth, A. *Proteins: Structure, Function, and Bioinformatics* **2018**, *86*, 833–843, DOI: [10.1002/prot.25496](https://doi.org/10.1002/prot.25496) (page 61).
- (381) Bitam, S. et al. *Scientific Reports* **2021**, *11*, 6842, DOI: [10.1038/s41598-021-83240-x](https://doi.org/10.1038/s41598-021-83240-x) (page 61).
- (382) Baatallah, N.; Elbahnsi, A.; Mornon, J. P.; Chevalier, B.; Pranke, I.; Servel, N.; Zelli, R.; Décout, J. L.; Edelman, A.; Sermet-Gaudelus, I.; Callebaut, I.; Hinzpeter, A. *Cellular and Molecular Life Sciences* **2021**, *78*, 7813–7829, DOI: [10.1007/S00018-021-03994-5](https://doi.org/10.1007/S00018-021-03994-5) (pages 61, 87, 167, 168).

- (383) Froux, L. et al. *European Journal of Medicinal Chemistry* **2020**, *190*, 112116–112116, DOI: [10.1016/j.ejmech.2020.112116](https://doi.org/10.1016/j.ejmech.2020.112116) (page 61).
- (384) Gimeno, A.; Ojeda-Montes, M. J.; Tomás-Hernández, S.; Cereto-Massagué, A.; Beltrán-Debón, R.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. *International Journal of Molecular Sciences* **2019**, *20*, 1375, DOI: [10.3390/ijms20061375](https://doi.org/10.3390/ijms20061375) (page 61).
- (385) Virtual Screening for Coronavirus Protease Inhibitors: A Waste of Good Electrons? <https://www.science.org/content/blog-post/virtual-screening-coronavirus-protease-inhibitors-waste-good-electrons> (accessed 08/22/2022) (page 61).
- (386) Cerón-Carrasco, J. P. *ChemMedChem* **2022**, *17*, e202200278, DOI: [10.1002/cmdc.202200278](https://doi.org/10.1002/cmdc.202200278) (page 61).
- (387) Scientifique, R. (12) *United States Patent (54) COMPOUNDS FOR TREATING CYSTIC FIBROSIS (71) Applicants : CENTRE NATIONAL DE LA*; 2019 (page 61).
- (388) Gaines, H.; Jones, K. R.; Lim, J.; Medhi, N. F.; Chen, S.; Scofield, R. H. *Journal of Diabetes and Its Complications* **2021**, *35*, 107845, DOI: [10.1016/j.jdiacomp.2020.107845](https://doi.org/10.1016/j.jdiacomp.2020.107845) (page 61).
- (389) Lopes-Pacheco, M. *Frontiers in Pharmacology* **2020**, *10* (page 61).
- (390) Yi, B.; Dalpke, A. H.; Boutin, S. *Frontiers in Cellular and Infection Microbiology* **2021**, *11* (page 61).
- (391) Harvey, C.; Weldon, S.; Elborn, S.; Downey, D. G.; Taggart, C. *International Journal of Molecular Sciences* **2022**, *23*, 3513, DOI: [10.3390/ijms23073513](https://doi.org/10.3390/ijms23073513) (page 61).
- (392) Greenfield, N. J. *Nature Protocols* **2006**, *1*, 2876–2890, DOI: [10.1038/nprot.2006.202](https://doi.org/10.1038/nprot.2006.202) (page 61).
- (393) Administration, A. G. D. o. H. T. G. Trikafta Therapeutic Goods Administration (TGA), <https://www.tga.gov.au/apm-summary/trikafpta> (accessed 07/22/2022) (pages 61, 64).
- (394) Who Is TRIKAFTA® For? | TRIKAFTA® (Elexacaftor/Tezacaftor/Ivacaftor and Ivacaftor) <https://www.trikafpta.com/who-trikafpta-is-for#mutationslist> (accessed 07/22/2022) (pages 61, 64, 158, 161).
- (395) Veit, G. et al. *Nature medicine* **2018**, *24*, 1732–1742, DOI: [10.1038/s41591-018-0200-x](https://doi.org/10.1038/s41591-018-0200-x) (page 61).
- (396) FDA Approves KALYDECO® (Ivacaftor) for More Than 900 People Ages 2 and Older with Cystic Fibrosis Who Have Certain Residual Function Mutations | Vertex Pharmaceuticals <https://investors.vrtx.com/news-releases/news-release-details/fda-approves-kalydeco-ivacaftor-more-900-people-ages-2-and> (accessed 07/05/2022) (pages 62, 114).

- (397) FDA Approves KALYDECO™ (Ivacaftor), the First Medicine to Treat the Underlying Cause of Cystic Fibrosis | Vertex Pharmaceuticals <https://investors.vrtx.com/news-releases/news-release-details/fda-approves-kalydecotm-ivacaftor-first-medicine-treat> (accessed 07/05/2022) (pages 62, 114).
- (398) Jih, K.-Y.; Hwang, T.-C. *Proceedings of the National Academy of Sciences* **2013**, *110*, 4404–4409, DOI: [10.1073/pnas.1215982110](https://doi.org/10.1073/pnas.1215982110) (page 62).
- (399) Laselva, O.; Qureshi, Z.; Zeng, Z.-W.; Petrotchenko, E. V.; Ramjeesingh, M.; Hamilton, C. M.; Huan, L.-J.; Borchers, C. H.; Pomès, R.; Young, R.; Bear, C. E. *iScience* **2021**, *24*, 102542, DOI: [10.1016/j.isci.2021.102542](https://doi.org/10.1016/j.isci.2021.102542) (pages 62, 167, 168).
- (400) Van Der Plas, S. E. et al. *Journal of Medicinal Chemistry* **2018**, *61*, 1425–1435, DOI: [10.1021/ACS.JMEDCHEM.7B01288](https://doi.org/10.1021/ACS.JMEDCHEM.7B01288) (pages 62, 86, 167).
- (401) Weikl, T. R.; Hu, J.; Xu, G.-K.; Lipowsky, R. *Cell Adhesion & Migration* **2016**, *10*, 576–589, DOI: [10.1080/19336918.2016.1180487](https://doi.org/10.1080/19336918.2016.1180487) (page 62).
- (402) Sharma, J. et al. *Nature Communications* **2021**, *12*, 4358, DOI: [10.1038/s41467-021-24575-x](https://doi.org/10.1038/s41467-021-24575-x) (page 63).
- (403) Crawford, D. K.; Mullenders, J.; Pott, J.; Boj, S. F.; Landskroner-Eiger, S.; Goddeeris, M. M. *Journal of Cystic Fibrosis* **2021**, *20*, 436–442, DOI: [10.1016/j.jcf.2021.01.009](https://doi.org/10.1016/j.jcf.2021.01.009) (page 63).
- (404) Hanafin, P. O.; Sermet-Gaudelus, I.; Griese, M.; Kappler, M.; Ellemunter, H.; Schwarz, C.; Wilson, J.; Tan, M.; Velkov, T.; Rao, G. G.; Schneider-Futschik, E. K. *Frontiers in Pharmacology* **2021**, *12*, 577263, DOI: [10.3389/fphar.2021.577263](https://doi.org/10.3389/fphar.2021.577263) (pages 63, 64, 165).
- (405) Robertson, S. M.; Luo, X.; Dubey, N.; Li, C.; Chavan, A. B.; Gilmartin, G. S.; Higgins, M.; Mahnke, L. *The Journal of Clinical Pharmacology* **2015**, *55*, 56–62, DOI: [10.1002/jcph.377](https://doi.org/10.1002/jcph.377) (page 63).
- (406) Lingam, S.; Thonghin, N.; Ford, R. C. *Scientific Reports* **2017**, *7*, 17481, DOI: [10.1038/s41598-017-17773-5](https://doi.org/10.1038/s41598-017-17773-5) (page 63).
- (407) Seelig, A. *Frontiers in Oncology* **2020**, *10* (page 63).
- (408) Barbieri, A.; Thonghin, N.; Shafi, T.; Prince, S. M.; Collins, R. F.; Ford, R. C. *Membranes* **2021**, *11*, 923, DOI: [10.3390/membranes11120923](https://doi.org/10.3390/membranes11120923) (page 63).
- (409) Grebert, C.; Becq, F.; Vandebrouck, C. *Cell Calcium* **2019**, *81*, 29–37, DOI: [10.1016/j.ceca.2019.05.007](https://doi.org/10.1016/j.ceca.2019.05.007) (page 63).
- (410) Ng, R. N.; Tai, A. S.; Chang, B. J.; Stick, S. M.; Kicic, A. *Frontiers in Microbiology* **2021**, *11* (page 63).
- (411) Allan, K. M.; Farrow, N.; Donnelley, M.; Jaffe, A.; Waters, S. A. *Frontiers in Pharmacology* **2021**, *12* (page 63).
- (412) Ledford, H. *Nature* **2020**, *579*, 185–185, DOI: [10.1038/d41586-020-00655-8](https://doi.org/10.1038/d41586-020-00655-8) (page 63).
- (413) Abdallah, K.; De Boeck, K.; Dooms, M.; Simoens, S. *Frontiers in Pharmacology* **2021**, *12* (page 64).

- (414) Guo, J.; Wang, J.; Zhang, J.; Fortunak, J.; Hill, A. *Journal of Cystic Fibrosis* **2022**, *0*, DOI: [10.1016/j.jcf.2022.04.007](https://doi.org/10.1016/j.jcf.2022.04.007) (page 64).
- (415) Grody, W. W.; Cutting, G. R.; Watson, M. S. *Genetics in Medicine* **2007**, *9*, 739–744, DOI: [10.1097/GIM.0b013e318159a331](https://doi.org/10.1097/GIM.0b013e318159a331) (page 64).
- (416) Van Goor, F. et al. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 18825–18830, DOI: [10.1073/pnas.0904709106](https://doi.org/10.1073/pnas.0904709106) (pages 64, 86, 119, 156, 158, 168).
- (417) Singh, M.; Rebordosa, C.; Bernholz, J.; Sharma, N. *Respirology* **2015**, *20*, 1172–1181, DOI: [10.1111/respond.12656](https://doi.org/10.1111.resp.12656) (page 64).
- (418) Zheng, B.; Cao, L. *Pediatric Pulmonology* **2017**, *52*, E11–E14, DOI: [10.1002/ppul.23539](https://doi.org/10.1002/ppul.23539) (page 64).
- (419) Ni, Q.; Chen, X.; Zhang, P.; Yang, L.; Lu, Y.; Xiao, F.; Wu, B.; Wang, H.; Zhou, W.; Dong, X. *Orphanet Journal of Rare Diseases* **2022**, *17*, 129, DOI: [10.1186/s13023-022-02279-9](https://doi.org/10.1186/s13023-022-02279-9) (page 64).
- (420) Zahav, M. M.; Orenti, A.; Jung, A.; Hatziagorou, E.; Olesen, H. V.; Kerem, E. *Journal of Cystic Fibrosis* **2022**, S1569199322006373, DOI: [10.1016/j.jcf.2022.07.015](https://doi.org/10.1016/j.jcf.2022.07.015) (page 64).
- (421) Lahiri, T.; Sullivan, J. S. *Pediatric Pulmonology* **2022**, *57*, S60–S74, DOI: [10.1002/ppul.25660](https://doi.org/10.1002/ppul.25660) (page 64).
- (422) Wong, A. P.; Chin, S.; Xia, S.; Garner, J.; Bear, C. E.; Rossant, J. *Nature Protocols* **2015**, *10*, 363–381, DOI: [10.1038/nprot.2015.021](https://doi.org/10.1038/nprot.2015.021) (page 65).
- (423) De Poel, E.; Lefferts, J. W.; Beekman, J. M. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2020**, *19 Suppl 1*, S60–S64, DOI: [10.1016/j.jcf.2019.11.002](https://doi.org/10.1016/j.jcf.2019.11.002) (page 65).
- (424) Blanpain, C.; Horsley, V.; Fuchs, E. *Cell* **2007**, *128*, 445–458, DOI: [10.1016/j.cell.2007.01.014](https://doi.org/10.1016/j.cell.2007.01.014) (page 65).
- (425) Awatade, N. T.; Wong, S. L.; Capraro, A.; Pandzic, E.; Slapetova, I.; Zhong, L.; Turgutoglu, N.; Fawcett, L. K.; Whan, R. M.; Jaffe, A.; Waters, S. A. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2021**, *20*, 364–371, DOI: [10.1016/j.jcf.2020.12.019](https://doi.org/10.1016/j.jcf.2020.12.019) (page 65).
- (426) Sato, T.; Stange, D. E.; Ferrante, M.; Vries, R. G. J.; van Es, J. H.; van den Brink, S.; van Houdt, W. J.; Pronk, A.; van Gorp, J.; Siersema, P. D.; Clevers, H. *Gastroenterology* **2011**, *141*, 1762–1772, DOI: [10.1053/j.gastro.2011.07.050](https://doi.org/10.1053/j.gastro.2011.07.050) (page 65).
- (427) Im, K.; Mareninov, S.; Diaz, M. F. P.; Yong, W. H. *Methods in molecular biology (Clifton, N.J.)* **2019**, *1897*, 299–311, DOI: [10.1007/978-1-4939-8935-5_26](https://doi.org/10.1007/978-1-4939-8935-5_26) (page 65).
- (428) Clevers, H. *Cell* **2016**, *165*, 1586–1597, DOI: [10.1016/j.cell.2016.05.082](https://doi.org/10.1016/j.cell.2016.05.082) (pages 65, 88).

- (429) Pranke, I.; Hatton, A.; Masson, A.; Flament, T.; Le Bourgeois, M.; Chedevreugne, F.; Bailly, C.; Urbach, V.; Hinzpeter, A.; Edelman, A.; Sermet-Gaudelus, I. *American Journal of Respiratory and Critical Care Medicine* **2019**, *199*, 123–126, DOI: [10.1164/rccm.201808-1436LE](https://doi.org/10.1164/rccm.201808-1436LE) (pages 65, 101, 113).
- (430) Wong, A. P.; Bear, C. E.; Chin, S.; Pasceri, P.; Thompson, T. O.; Huan, L.-J.; Ratjen, F.; Ellis, J.; Rossant, J. *Nature Biotechnology* **2012**, *30*, 876–882, DOI: [10.1038/nbt.2328](https://doi.org/10.1038/nbt.2328) (page 65).
- (431) Keegan, D. E.; Brewington, J. J. *International Journal of Molecular Sciences* **2021**, *22*, 4448, DOI: [10.3390/ijms22094448](https://doi.org/10.3390/ijms22094448) (page 65).
- (432) Clancy, J. P. et al. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2019**, *18*, 22–34, DOI: [10.1016/j.jcf.2018.05.004](https://doi.org/10.1016/j.jcf.2018.05.004) (pages 65, 100).
- (433) Ciciriello, F.; Bijvelds, M. J. C.; Alghisi, F.; Meijzen, K. F.; Cristiani, L.; Sorio, C.; Melotti, P.; Fiocchi, A. G.; Lucidi, V.; De Jonge, H. R. *Journal of personalized medicine* **2022**, *12*, DOI: [10.3390/jpm12040632](https://doi.org/10.3390/jpm12040632) (page 65).
- (434) Dekkers, J. F. et al. *Nature Medicine* **2013**, *19*, 939–945, DOI: [10.1038/nm.3201](https://doi.org/10.1038/nm.3201) (pages 66, 115).
- (435) Bonora, M.; Paterniani, S.; Rimessi, A.; De Marchi, E.; Suski, J. M.; Bononi, A.; Giorgi, C.; Marchi, S.; Missiroli, S.; Poletti, F.; Wieckowski, M. R.; Pinton, P. *Purinergic Signalling* **2012**, *8*, 343–357, DOI: [10.1007/s11302-012-9305-8](https://doi.org/10.1007/s11302-012-9305-8) (page 66).
- (436) Cui, G.; Cottrill, K. A.; McCarty, N. A. In *Structure and Function of Membrane Proteins*, Schmidt-Krey, I., Gumbart, J. C., Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2021, pp 49–67, DOI: [10.1007/978-1-0716-1394-8_4](https://doi.org/10.1007/978-1-0716-1394-8_4) (page 67).
- (437) Cai, Z.; Sohma, Y.; Bompadre, S. G.; Sheppard, D. N.; Hwang, T. C. *Methods in molecular biology (Clifton, N.J.)* **2011**, *741*, 419–441, DOI: [10.1007/978-1-61779-117-8_27](https://doi.org/10.1007/978-1-61779-117-8_27) (page 67).
- (438) Hoenig, M. P.; Zeidel, M. L. *Clinical journal of the American Society of Nephrology: CJASN* **2014**, *9*, 1272–1281, DOI: [10.2215/CJN.08860813](https://doi.org/10.2215/CJN.08860813) (pages 67, 69).
- (439) BioRender Patch-Clamp Recording Principle <https://app.biorender.com/biorender-templates/figures/all/t-602d84be5d2b0800a60bdf3c-patch-clamp-recording-principle> (accessed 07/30/2022) (page 68).
- (440) Chang, X.-b.; Mengos, A.; Hou, Y.-x.; Cui, L.; Jensen, T. J.; Aleksandrov, A.; Riordan, J. R.; Gentzsch, M. *Journal of Cell Science* **2008**, *121*, 2814–2823, DOI: [10.1242/jcs.028951](https://doi.org/10.1242/jcs.028951) (page 70).
- (441) Lopes-Pacheco, M. *Frontiers in Pharmacology* **2016**, *7* (page 70).
- (442) Lewis, H. A.; Zhao, X.; Wang, C.; Sauder, J. M.; Rooney, I.; Noland, B. W.; Lorimer, D.; Kearins, M. C.; Conners, K.; Condon, B.; Maloney, P. C.; Guggino, W. B.; Hunt, J. F.; Emtage, S. *Journal of Biological Chemistry* **2005**, *280*, 1346–1353, DOI: [10.1074/jbc.M410968200](https://doi.org/10.1074/jbc.M410968200) (page 70).

- (443) Riordan, J. R.; Rommens, J. M.; Kerem, B.; Alon, N.; Rozmahel, R.; Grzelczak, Z.; Zielenski, J.; Lok, S.; Plavsic, N.; Chou, J. L. *Science (New York, N.Y.)* **1989**, *245*, 1066–1073, DOI: [10.1126/science.2475911](https://doi.org/10.1126/science.2475911) (pages 71, 119, 129).
- (444) Saotome, K.; Murthy, S. E.; Kefauver, J. M.; Whitwam, T.; Patapoutian, A.; Ward, A. B. *Nature* **2018**, *554*, 481–486, DOI: [10.1038/nature25453](https://doi.org/10.1038/nature25453) (page 71).
- (445) Zalk, R.; Clarke, O. B.; des Georges, A.; Grassucci, R. A.; Reiken, S.; Mancia, F.; Hendrickson, W. A.; Frank, J.; Marks, A. R. *Nature* **2015**, *517*, 44–49, DOI: [10.1038/nature13950](https://doi.org/10.1038/nature13950) (page 71).
- (446) Chen, X.; Liu, M.; Tian, Y.; Li, J.; Qi, Y.; Zhao, D.; Wu, Z.; Huang, M.; Wong, C. C. L.; Wang, H.-W.; Wang, J.; Yang, H.; Xu, Y. *Cell Research* **2018**, *28*, 518–528, DOI: [10.1038/s41422-018-0029-3](https://doi.org/10.1038/s41422-018-0029-3) (page 71).
- (447) Ho, B.; Baryshnikova, A.; Brown, G. W. *Cell Systems* **2018**, *6*, 192–205.e3, DOI: [10.1016/j.cels.2017.12.004](https://doi.org/10.1016/j.cels.2017.12.004) (page 71).
- (448) Ratjen, F.; Bell, S. C.; Rowe, S. M.; Goss, C. H.; Quittner, A. L.; Bush, A. *Nature reviews. Disease primers* **2015**, *1*, 15010, DOI: [10.1038/NRDP.2015.10](https://doi.org/10.1038/NRDP.2015.10) (pages 73, 115, 119).
- (449) Gadsby, D. C.; Nairn, A. C. *Trends in Biochemical Sciences* **1994**, *19*, 513–518, DOI: [10.1016/0968-0004\(94\)90141-4](https://doi.org/10.1016/0968-0004(94)90141-4) (page 74).
- (450) Fu, J.; Ji, H. L.; Naren, A. P.; Kirk, K. L. *Journal of Physiology* **2001**, *536*, 459–470, DOI: [10.1111/J.1469-7793.2001.0459C.XD](https://doi.org/10.1111/J.1469-7793.2001.0459C.XD) (page 74).
- (451) Jurkuvenaite, A.; Varga, K.; Nowotarski, K.; Kirk, K. L.; Sorscher, E. J.; Li, Y.; Clancy, J. P.; Bebok, Z.; Collawn, J. F. *Journal of Biological Chemistry* **2006**, *281*, 3329–3334, DOI: [10.1074/jbc.M508131200](https://doi.org/10.1074/jbc.M508131200) (pages 74, 87).
- (452) Loo, T. W.; Bartlett, M. C.; Clarke, D. M. *Biochemical Pharmacology* **2013**, *86*, 612–619, DOI: [10.1016/J.BCP.2013.06.028](https://doi.org/10.1016/J.BCP.2013.06.028) (pages 74, 100).
- (453) Ren, H. Y.; Grove, D. E.; De La Rosa, O.; Houck, S. A.; Sophia, P.; Van Goor, F.; Hoffman, B. J.; Cyr, D. M. *Molecular Biology of the Cell* **2013**, *24*, 3016–3024, DOI: [10.1091/MBC.E13-05-0240](https://doi.org/10.1091/MBC.E13-05-0240) (pages 74, 100).
- (454) Loo, T. W.; Clarke, D. M. *Biochemical Pharmacology* **2017**, *136*, 24–31, DOI: [10.1016/J.BCP.2017.03.020](https://doi.org/10.1016/J.BCP.2017.03.020) (page 74).
- (455) Hudson, R. P.; Dawson, J. E.; Chong, P. A.; Yang, Z.; Millen, L.; Thomas, P. J.; Brouillette, C. G.; Forman-Kay, J. D. *Molecular Pharmacology* **2017**, *92*, 124–135, DOI: [10.1124/MOL.117.108373](https://doi.org/10.1124/MOL.117.108373) (page 74).
- (456) Okiyoneda, T.; Veit, G.; Dekkers, J. F.; Bagdany, M.; Soya, N.; Xu, H.; Roldan, A.; Verkman, A. S.; Kurth, M.; Simon, A.; Hegedus, T.; Beekman, J. M.; Lukacs, G. L. *Nature Chemical Biology* **2013**, *9*, 444–454, DOI: [10.1038/NCHEM BIO.1253](https://doi.org/10.1038/NCHEM BIO.1253) (pages 74, 87, 100).
- (457) Berkers, G. et al. *Cell Reports* **2019**, *26*, 1701–1708.e3, DOI: [10.1016/J.CELREP.2019.01.068](https://doi.org/10.1016/J.CELREP.2019.01.068) (pages 74, 101).

- (458) McCarthy, C.; Brewington, J. J.; Harkness, B.; Clancy, J. P.; Trapnell, B. C. *The European Respiratory Journal* **2018**, *51*, 1702457, DOI: [10.1183/13993003.02457-2017](https://doi.org/10.1183/13993003.02457-2017) (pages 74, 101).
- (459) Ramalho, A. S.; Förstová, E.; Vonk, A. M.; Ferrante, M.; Verfailli, C.; Dupont, L.; Boon, M.; Proesmans, M.; Beekma, J. M.; Sarouk, I.; Cordero, C. V.; Vermeule, F.; De Boeck, K. *The European Respiratory Journal* **2021**, *57*, 1902426, DOI: [10.1183/13993003.02426-2019](https://doi.org/10.1183/13993003.02426-2019) (pages 74, 89, 101, 113).
- (460) Sato, T.; Vries, R. G.; Snippert, H. J.; Van De Wetering, M.; Barker, N.; Stange, D. E.; Van Es, J. H.; Abo, A.; Kujala, P.; Peters, P. J.; Clevers, H. *Nature* **2009**, *459*, 262–265, DOI: [10.1038/NATURE07935](https://doi.org/10.1038/NATURE07935) (page 74).
- (461) Awatade, N. T.; Wong, S. L.; Hewson, C. K.; Fawcett, L. K.; Kicic, A.; Jaffe, A.; Waters, S. A. *Frontiers in Pharmacology* **2018**, *9*, DOI: [10.3389/FPHAR.2018.01429](https://doi.org/10.3389/FPHAR.2018.01429) (pages 74, 119).
- (462) Pollard, B. S.; Pollard, H. B. *Pediatric Pulmonology* **2018**, *53*, S12–S29, DOI: [10.1002/PPUL.24118](https://doi.org/10.1002/PPUL.24118) (page 74).
- (463) Dey, I.; Shah, K.; Bradbury, N. A. *Journal of Genetic Syndromes & Gene Therapy* **2016**, *07*, DOI: [10.4172/2157-7412.1000284](https://doi.org/10.4172/2157-7412.1000284) (page 75).
- (464) Clancy, J. P. et al. *PLoS ONE* **2013**, *8*, DOI: [10.1371/JOURNAL.PONE.0073905](https://doi.org/10.1371/JOURNAL.PONE.0073905) (page 75).
- (465) Graeber, S. Y.; Hug, M. J.; Sommerburg, O.; Hirtz, S.; Hentschel, J.; Heinzmamn, A.; Dopfer, C.; Schulz, A.; Mainz, J. G.; Tümmller, B.; Mall, M. A. *American Journal of Respiratory and Critical Care Medicine* **2015**, *192*, 1252–1255, DOI: [10.1164/RCCM.201507-1271LE](https://doi.org/10.1164/RCCM.201507-1271LE) (pages 75, 77).
- (466) Veeze, H. J.; Halley, D. J.; Bijman, J.; De Jongste, J. C.; De Jonge, H. R.; Sinaasappel, M. *Journal of Clinical Investigation* **1994**, *93*, 461–466, DOI: [10.1172/JCI116993](https://doi.org/10.1172/JCI116993) (page 77).
- (467) Dekkers, J. F.; Van Mourik, P.; Vonk, A. M.; Kruisselbrink, E.; Berkers, G.; de Winter-de Groot, K. M.; Janssens, H. M.; Bronsveld, I.; van der Ent, C. K.; de Jonge, H. R.; Beekman, J. M. *Journal of Cystic Fibrosis* **2016**, *15*, 568–578, DOI: [10.1016/j.jcf.2016.04.007](https://doi.org/10.1016/j.jcf.2016.04.007) (pages 78, 89, 105, 107, 113–115).
- (468) Cuyx, S.; Ramalho, A. S.; Corthout, N.; Fieuws, S.; Fürstová, E.; Arnauts, K.; Ferrante, M.; Verfaillie, C.; Munck, S.; Boon, M.; Proesmans, M.; Dupont, L.; De Boeck, K.; Vermeulen, F. *Thorax* **2021**, *76*, 1146–1149, DOI: [10.1136/THORAXJNL-2020-216368](https://doi.org/10.1136/THORAXJNL-2020-216368) (page 78).
- (469) Van Mourik, P.; Beekman, J. M.; Van Der Ent, C. K. *European Respiratory Journal* **2019**, *54*, DOI: [10.1183/13993003.02379-2018](https://doi.org/10.1183/13993003.02379-2018) (page 78).
- (470) Zomer-van Ommen, D. D.; de Poel, E.; Kruisselbrink, E.; Oppelaar, H.; Vonk, A. M.; Janssens, H. M.; van der Ent, C. K.; Hagemeijer, M. C.; Beekman, J. M. *Journal of Cystic Fibrosis* **2018**, *17*, 316–324, DOI: [10.1016/J.JCF.2018.02.007](https://doi.org/10.1016/J.JCF.2018.02.007) (page 78).

- (471) Laselva, O.; Bartlett, C.; Gunawardena, T. N. A.; Ouyang, H.; Eckford, P. D. W.; Moraes, T. J.; Bear, C. E.; Gonska, T. *The European Respiratory Journal* **2021**, *57*, 2002774, DOI: [10.1183/13993003.02774-2020](https://doi.org/10.1183/13993003.02774-2020) (pages 80, 86, 100, 107, 124).
- (472) Shaughnessy, C. A.; Zeitlin, P. L.; Bratcher, P. E. *Scientific Reports* **2021**, *11*, DOI: [10.1038/S41598-021-99184-1](https://doi.org/10.1038/S41598-021-99184-1) (pages 80, 86, 100, 107).
- (473) Veit, G.; Velkov, T.; Xu, H.; Vadeboncoeur, N.; Bilodeau, L.; Matouk, E.; Lukacs, G. L. *Journal of Personalized Medicine* **2021**, *11*, DOI: [10.3390/JPM11070643](https://doi.org/10.3390/JPM11070643) (pages 80, 86, 107).
- (474) Beckerman, M. **2015**, 61–94, DOI: [10.1007/978-3-319-22117-5_3](https://doi.org/10.1007/978-3-319-22117-5_3) (page 83).
- (475) Van Goor, F.; Hadida, S.; Grootenhuis, P. D.; Burton, B.; Stack, J. H.; Straley, K. S.; Decker, C. J.; Miller, M.; McCartney, J.; Olson, E. R.; Wine, J. J.; Frizzell, R. A.; Ashlock, M.; Negulescu, P. A. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 18843–18848, DOI: [10.1073/PNAS.1105787108](https://doi.org/10.1073/PNAS.1105787108) (pages 85, 114).
- (476) Gees, M. et al. *Frontiers in Pharmacology* **2018**, *9*, DOI: [10.3389/FPHAR.2018.01221](https://doi.org/10.3389/FPHAR.2018.01221) (page 86).
- (477) Dekkers, J. F. et al. *Science Translational Medicine* **2016**, *8*, DOI: [10.1126/SCITRANSLMED.AAD8278](https://doi.org/10.1126/SCITRANSLMED.AAD8278) (pages 86, 100, 101).
- (478) Phuan, P. W.; Son, J. H.; Tan, J. A.; Li, C.; Musante, I.; Zlock, L.; Nielson, D. W.; Finkbeiner, W. E.; Kurth, M. J.; Galietta, L. J.; Haggie, P. M.; Verkman, A. S. *Journal of Cystic Fibrosis* **2018**, *17*, 595–606, DOI: [10.1016/J.JCF.2018.05.010](https://doi.org/10.1016/J.JCF.2018.05.010) (pages 86, 158, 167).
- (479) Phuan, P. W.; Tan, J. A.; Rivera, A. A.; Zlock, L.; Nielson, D. W.; Finkbeiner, W. E.; Haggie, P. M.; Verkman, A. S. *Scientific Reports* **2019**, *9*, DOI: [10.1038/S41598-019-54158-2](https://doi.org/10.1038/S41598-019-54158-2) (page 86).
- (480) Veit, G.; Da Fonte, D. F.; Avramescu, R. G.; Premchandar, A.; Bagdany, M.; Xu, H.; Bensinger, D.; Stubba, D.; Schmidt, B.; Matouk, E.; Lukacs, G. L. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2020**, *19*, 236–244, DOI: [10.1016/j.jcf.2019.10.011](https://doi.org/10.1016/j.jcf.2019.10.011) (pages 86, 113, 114, 124).
- (481) Nussinov, R.; Tsai, C. J. *Cell* **2013**, *153*, 293–305, DOI: [10.1016/J.CELL.2013.03.034](https://doi.org/10.1016/J.CELL.2013.03.034) (page 86).
- (482) Berkers, G. et al. *Journal of Cystic Fibrosis* **2020**, *19*, 955–961, DOI: [10.1016/J.JCF.2020.04.014](https://doi.org/10.1016/J.JCF.2020.04.014) (page 86).
- (483) McKone, E. F.; Borowitz, D.; Drevinek, P.; Griese, M.; Konstan, M. W.; Wainwright, C.; Ratjen, F.; Sermet-Gaudelus, I.; Plant, B.; Munck, A.; Jiang, Y.; Gilmartin, G.; Davies, J. C. *The Lancet Respiratory Medicine* **2014**, *2*, 902–910, DOI: [10.1016/S2213-2600\(14\)70218-8](https://doi.org/10.1016/S2213-2600(14)70218-8) (page 86).
- (484) Volkova, N.; Moy, K.; Evans, J.; Campbell, D.; Tian, S.; Simard, C.; Higgins, M.; Konstan, M. W.; Sawicki, G. S.; Elbert, A.; Charman, S. C.; Marshall, B. C.; Bilton, D. *Journal of Cystic Fibrosis* **2020**, *19*, 68–79, DOI: [10.1016/J.JCF.2019.05.015](https://doi.org/10.1016/J.JCF.2019.05.015) (page 86).

- (485) Sohma, Y.; Yu, Y.-C.; Hwang, T.-C. *Current Pharmaceutical Design* **2013**, *19*, 3521–3528, DOI: [10.2174/13816128113199990320](https://doi.org/10.2174/13816128113199990320) (page 86).
- (486) Keating, D. et al. *New England Journal of Medicine* **2018**, *379*, 1612–1620, DOI: [10.1056/NEJMoa1807120](https://doi.org/10.1056/NEJMoa1807120) (pages 86, 119, 159).
- (487) Veit, G.; Roldan, A.; Hancock, M. A.; da Fonte, D. F.; Xu, H.; Hussein, M.; Frenkiel, S.; Matouk, E.; Velkov, T.; Lukacs, G. L. *JCI Insight* **2020**, *5*, DOI: [10.1172/JCI.INSIGHT.139983](https://doi.org/10.1172/JCI.INSIGHT.139983) (page 86).
- (488) Veit, G.; Vaccarin, C.; Lukacs, G. L. *Journal of Cystic Fibrosis* **2021**, *20*, 895–898, DOI: [10.1016/J.JCF.2021.03.011](https://doi.org/10.1016/J.JCF.2021.03.011) (pages 86, 100).
- (489) Werlin, S. et al. *Journal of pediatric gastroenterology and nutrition* **2015**, *60*, 675–679, DOI: [10.1097/MPG.0000000000000623](https://doi.org/10.1097/MPG.0000000000000623) (page 87).
- (490) Kleizen, B.; van Willigen, M.; Mijnders, M.; Peters, F.; Grudniewska, M.; Hillegaars, T.; Thomas, A.; Kooijman, L.; Peters, K. W.; Frizzell, R.; van der Sluijs, P.; Braakman, I. *Journal of Molecular Biology* **2021**, *433*, DOI: [10.1016/J.JMB.2021.166955](https://doi.org/10.1016/J.JMB.2021.166955) (pages 88, 159).
- (491) De Jonge, H. R.; Ballmann, M.; Vreeze, H.; Bronsveld, I.; Stanke, F.; Tümmler, B.; Sinaasappel, M. *Journal of Cystic Fibrosis* **2004**, *3*, 159–163, DOI: [10.1016/J.JCF.2004.05.034](https://doi.org/10.1016/J.JCF.2004.05.034) (page 94).
- (492) Derichs, N.; Sanz, J.; Von Kanel, T.; Stolpe, C.; Zapf, A.; Tümmler, B.; Gallati, S.; Ballmann, M. *Thorax* **2010**, *65*, 594–599, DOI: [10.1136/THX.2009.125088](https://doi.org/10.1136/THX.2009.125088) (page 94).
- (493) Li, H.; Sheppard, D. N.; Hug, M. J. *Journal of Cystic Fibrosis* **2004**, *3*, 123–126, DOI: [10.1016/J.JCF.2004.05.026](https://doi.org/10.1016/J.JCF.2004.05.026) (page 94).
- (494) Pankow, S.; Bamberger, C.; Calzolari, D.; Martínez-Bartolomé, S.; Lavallée-Adam, M.; Balch, W. E.; Yates, J. R. *Nature* **2015**, *528*, 510–516, DOI: [10.1038/NATURE15729](https://doi.org/10.1038/NATURE15729) (page 95).
- (495) Van Goor, F.; Yu, H.; Burton, B.; Hoffman, B. J. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2014**, *13*, 29–36, DOI: [10.1016/j.jcf.2013.06.008](https://doi.org/10.1016/j.jcf.2013.06.008) (pages 96, 113, 114, 158, 159, 167).
- (496) Mark, P.; Nilsson, L. *Journal of Physical Chemistry A* **2001**, *105*, 9954–9960, DOI: [10.1021/jp003020w](https://doi.org/10.1021/jp003020w) (page 96).
- (497) Nosé, S.; Klein, M. L. *Molecular Physics* **1983**, *50*, 1055–1076, DOI: [10.1080/00268978300102851](https://doi.org/10.1080/00268978300102851) (page 97).
- (498) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *Journal of Physical Chemistry B* **2010**, *114*, 7830–7843, DOI: [10.1021/JP101759Q](https://doi.org/10.1021/JP101759Q) (page 97).
- (499) Pandit, S. A.; Scott, H. L. *Soft Matter* **2009**, *4*, 1–82, DOI: [10.1002/9783527623372.CH1](https://doi.org/10.1002/9783527623372.CH1) (page 97).
- (500) Buchoux, S. *Bioinformatics (Oxford, England)* **2017**, *33*, 133–134, DOI: [10.1093/BIOINFORMATICS/BTW563](https://doi.org/10.1093/BIOINFORMATICS/BTW563) (page 97).

- (501) Rasa Ruseckaite Tom Ranger, J. D. M. G. S. B. N. B. S. A. *AUSTRALIAN CYSTIC FIBROSIS DATA REGISTRY ANNUAL REPORT 2017*; 2017 (page 100).
- (502) Saint-Criq, V.; Gray, M. A. *Cellular and Molecular Life Sciences* **2017**, *74*, 93–115, DOI: [10.1007/s00018-016-2391-y](https://doi.org/10.1007/s00018-016-2391-y) (page 100).
- (503) Marson, F. A. L.; Bertuzzo, C. S.; Ribeiro, J. D. *The Lancet. Respiratory Medicine* **2016**, *4*, e37–e38, DOI: [10.1016/S2213-2600\(16\)30188-6](https://doi.org/10.1016/S2213-2600(16)30188-6) (page 100).
- (504) Of the Commissioner, O. FDA Approves New Breakthrough Therapy for Cystic Fibrosis FDA, <https://www.fda.gov/news-events/press-announcements/fda-approves-new-breakthrough-therapy-cystic-fibrosis> (accessed 07/05/2022) (page 100).
- (505) FDA Approves SYMDEKO (Tezacaftor/Ivacaftor and Ivacaftor) to Treat the Underlying Cause of Cystic Fibrosis in People Ages 12 and Older with Certain Mutations in the CFTR Gene | Vertex Pharmaceuticals <https://investors.vrtx.com/news-releases/news-release-details/fda-approves-symdekom-tezacaftorivacaftor-and-ivacaftor-treat> (accessed 07/05/2022) (page 100).
- (506) FDA Approves ORKAMBI™ (Lumacaftor/Ivacaftor) - the First Medicine to Treat the Underlying Cause of Cystic Fibrosis for People Ages 12 and Older with Two Copies of the F508del Mutation | Vertex Pharmaceuticals <https://investors.vrtx.com/news-releases/news-release-details/fda-approves-orkambitm-lumacaftorivacaftor-first-medicine-treat> (accessed 07/05/2022) (page 100).
- (507) FDA Approves Expansion of Modulators for People With Certain Rare Mutations | Cystic Fibrosis Foundation <https://www.cff.org/news/2020-12/fda-approves-expansion-modulators-people-certain-rare-mutations> (accessed 07/05/2022) (page 100).
- (508) Bell, S. C. et al. *The Lancet. Respiratory Medicine* **2020**, *8*, 65–124, DOI: [10.1016/S2213-2600\(19\)30337-6](https://doi.org/10.1016/S2213-2600(19)30337-6) (page 100).
- (509) Clarke, L. A.; Sousa, L.; Barreto, C.; Amaral, M. D. *Respiratory Research* **2013**, *14*, 38, DOI: [10.1186/1465-9921-14-38](https://doi.org/10.1186/1465-9921-14-38) (page 100).
- (510) Martinovich, K. M.; Iosifidis, T.; Buckley, A. G.; Looi, K.; Ling, K.-M.; Sutanto, E. N.; Kicic-Starcevich, E.; Garratt, L. W.; Shaw, N. C.; Montgomery, S.; Lannigan, F. J.; Knight, D. A.; Kicic, A.; Stick, S. M. *Scientific Reports* **2017**, *7*, 17971, DOI: [10.1038/s41598-017-17952-4](https://doi.org/10.1038/s41598-017-17952-4) (page 100).
- (511) Pranke, I. M. et al. *Scientific Reports* **2017**, *7*, 7375, DOI: [10.1038/s41598-017-07504-1](https://doi.org/10.1038/s41598-017-07504-1) (pages 101, 102, 113, 114).
- (512) Dekkers, J. F.; van der Ent, C. K.; Beekman, J. M. *Rare Diseases (Austin, Tex.)* **2013**, *1*, e27112, DOI: [10.4161/rdis.27112](https://doi.org/10.4161/rdis.27112) (page 101).
- (513) Guinamard, R.; Akabas, M. H. *Biochemistry* **1999**, *38*, 5528–5537, DOI: [10.1021/bi990155n](https://doi.org/10.1021/bi990155n) (pages 101, 113, 114).
- (514) Han, S. T.; Rab, A.; Pellicore, M. J.; Davis, E. F.; McCague, A. F.; Evans, T. A.; Joynt, A. T.; Lu, Z.; Cai, Z.; Raraigh, K. S.; Hong, J. S.; Sheppard, D. N.; Sorscher, E. J.; Cutting, G. R. *JCI insight* **2018**, *3*, 121159, DOI: [10.1172/jci.insight.121159](https://doi.org/10.1172/jci.insight.121159) (pages 105, 145).

- (515) Van Barneveld, A.; Stanke, F.; Tamm, S.; Siebert, B.; Brandes, G.; Derichs, N.; Ballmann, M.; Junge, S.; Tümmler, B. *Biochimica Et Biophysica Acta* **2010**, *1802*, 1062–1069, DOI: [10.1016/j.bbadi.2010.08.001](https://doi.org/10.1016/j.bbadi.2010.08.001) (page 110).
- (516) The Human Protein Atlas <http://www.proteinatlas.org/> (accessed 07/05/2022) (page 110).
- (517) Uhlén, M. et al. *Science (New York, N.Y.)* **2015**, *347*, 1260419, DOI: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419) (page 110).
- (518) Cui, G.; Zhang, Z.-R. R.; O'Brien, A. R.; Song, B.; McCarty, N. A.; O'Brien, A. R. W.; Song, B.; McCarty, N. A.; O'Brien, A. R.; Song, B.; McCarty, N. A. *The Journal of Membrane Biology* **2008**, *222*, 91–106, DOI: [10.1007/s00232-008-9105-9](https://doi.org/10.1007/s00232-008-9105-9) (pages 110, 114, 131, 141).
- (519) Sorum, B.; Czégé, D.; Csanády, L. *Cell* **2015**, *163*, 724–733, DOI: [10.1016/j.cell.2015.09.052](https://doi.org/10.1016/j.cell.2015.09.052) (pages 110, 142).
- (520) Gentzsch, M.; Boyles, S. E.; Cheluvaraju, C.; Chaudhry, I. G.; Quinney, N. L.; Cho, C.; Dang, H.; Liu, X.; Schlegel, R.; Randell, S. H. *American Journal of Respiratory Cell and Molecular Biology* **2017**, *56*, 568–574, DOI: [10.1165/rcmb.2016-0276MA](https://doi.org/10.1165/rcmb.2016-0276MA) (pages 113, 115).
- (521) Saint-Criq, V.; Delpiano, L.; Casement, J.; Onuora, J. C.; Lin, J.; Gray, M. A. *Cells* **2020**, *9*, 2137, DOI: [10.3390/cells9092137](https://doi.org/10.3390/cells9092137) (page 113).
- (522) Matthes, E.; Goepp, J.; Martini, C.; Shan, J.; Liao, J.; Thomas, D. Y.; Hanrahan, J. W. *Frontiers in Pharmacology* **2018**, *9*, 1490, DOI: [10.3389/fphar.2018.01490](https://doi.org/10.3389/fphar.2018.01490) (pages 113, 159).
- (523) Brewington, J. J.; Filbrandt, E. T.; LaRosa, F. J.; Moncivaiz, J. D.; Ostmann, A. J.; Strecker, L. M.; Clancy, J. P. *JCI insight* **2018**, *3*, 99385, DOI: [10.1172/jci.insight.99385](https://doi.org/10.1172/jci.insight.99385) (page 113).
- (524) Kerem, E. et al. *Annals of the American Thoracic Society* **2021**, *18*, 433–441, DOI: [10.1513/AnnalsATS.202006-6590C](https://doi.org/10.1513/AnnalsATS.202006-6590C) (page 114).
- (525) Cui, G.; Freeman, C. S.; Knotts, T.; Prince, C. Z.; Kuang, C.; McCarty, N. A. *The Journal of Biological Chemistry* **2013**, *288*, 20758–20767, DOI: [10.1074/jbc.M113.476226](https://doi.org/10.1074/jbc.M113.476226) (page 114).
- (526) Zhang, J.; Hwang, T.-C. *The Journal of General Physiology* **2017**, *149*, 355–372, DOI: [10.1085/jgp.201611664](https://doi.org/10.1085/jgp.201611664) (pages 114, 144, 168).
- (527) Jih, K.-Y.; Lin, W.-Y.; Sohma, Y.; Hwang, T.-C. *Current Opinion in Pharmacology* **2017**, *34*, 98–104, DOI: [10.1016/j.coph.2017.09.015](https://doi.org/10.1016/j.coph.2017.09.015) (page 114).
- (528) Silva, I. A. L.; Railean, V.; Duarte, A.; Amaral, M. D. *Journal of Personalized Medicine* **2021**, *11*, 421, DOI: [10.3390/jpm11050421](https://doi.org/10.3390/jpm11050421) (page 115).
- (529) Jurkuvenaite, A.; Chen, L.; Bartoszewski, R.; Goldstein, R.; Bebok, Z.; Matalon, S.; Collawn, J. F. *American Journal of Respiratory Cell and Molecular Biology* **2010**, *42*, 363–372, DOI: [10.1165/rcmb.2008-04340C](https://doi.org/10.1165/rcmb.2008-04340C) (page 115).
- (530) Gentzsch, M.; Chang, X.-B.; Cui, L.; Wu, Y.; Ozols, V. V.; Choudhury, A.; Pagano, R. E.; Riordan, J. R. *Molecular Biology of the Cell* **2004**, *15*, 2684–2696, DOI: [10.1091/mbc.e04-03-0176](https://doi.org/10.1091/mbc.e04-03-0176) (page 115).

- (531) Cui, G. et al. *Developmental Cell* **2019**, *51*, 421–430.e3, DOI: [10.1016/j.devcel.2019.09.017](https://doi.org/10.1016/j.devcel.2019.09.017) (page 115).
- (532) Blouquit, S.; Morel, H.; Hinnrasky, J.; Naline, E.; Puchelle, E.; Chinet, T. *American Journal of Respiratory Cell and Molecular Biology* **2002**, *27*, 503–510, DOI: [10.1165/rccb.4869](https://doi.org/10.1165/rccb.4869) (page 115).
- (533) Smith, J. J.; Welsh, M. J. *The Journal of Clinical Investigation* **1993**, *91*, 1590–1597, DOI: [10.1172/JCI116365](https://doi.org/10.1172/JCI116365) (page 115).
- (534) Avramescu, R. G.; Kai, Y.; Xu, H.; Bidaud-Meynard, A.; Schnúr, A.; Frenkiel, S.; Matouk, E.; Veit, G.; Lukacs, G. L. *Human Molecular Genetics* **2017**, *26*, 4873–4885, DOI: [10.1093/hmg/ddx367](https://doi.org/10.1093/hmg/ddx367) (page 115).
- (535) Callebaut, I.; Hoffmann, B.; Lehn, P.; Mornon, J. P. *Cellular and molecular life sciences: CMLS* **2017**, *74*, 3–22, DOI: [10.1007/s00018-016-2385-9](https://doi.org/10.1007/s00018-016-2385-9) (page 115).
- (536) Molinski, S. V. et al. *EMBO molecular medicine* **2017**, *9*, 1224–1243, DOI: [10.15252/emmm.201607137](https://doi.org/10.15252/emmm.201607137) (page 116).
- (537) Mall, M. A.; Hartl, D. *The European Respiratory Journal* **2014**, *44*, 1042–1054, DOI: [10.1183/09031936.00228013](https://doi.org/10.1183/09031936.00228013) (page 119).
- (538) Elborn, J. S. *Lancet (London, England)* **2016**, *388*, 2519–2531, DOI: [10.1016/S0140-6736\(16\)00576-6](https://doi.org/10.1016/S0140-6736(16)00576-6) (page 119).
- (539) Rowe, S. M.; Miller, S.; Sorscher, E. J. *New England Journal of Medicine* **2005**, *352*, 1992–2001, DOI: [10.1056/NEJMra043184](https://doi.org/10.1056/NEJMra043184) (page 119).
- (540) De Boeck, K.; Amaral, M. D. *The Lancet. Respiratory Medicine* **2016**, *4*, 662–674, DOI: [10.1016/S2213-2600\(16\)00023-0](https://doi.org/10.1016/S2213-2600(16)00023-0) (page 119).
- (541) Bonadia, L. C.; de Lima Marson, F. A.; Ribeiro, J. D.; Paschoal, I. A.; Pereira, M. C.; Ribeiro, A. F.; Bertuzzo, C. S. *Gene* **2014**, *540*, 183–190, DOI: [10.1016/j.gene.2014.02.040](https://doi.org/10.1016/j.gene.2014.02.040) (page 119).
- (542) Accurso, F. J. et al. *The New England Journal of Medicine* **2010**, *363*, 1991–2003, DOI: [10.1056/NEJMoa0909825](https://doi.org/10.1056/NEJMoa0909825) (page 119).
- (543) Yu, H.; Burton, B.; Huang, C.-J.; Worley, J.; Cao, D.; Johnson, J. P.; Urrutia, A.; Joubran, J.; Seepersaud, S.; Sussky, K.; Hoffman, B. J.; Van Goor, F. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2012**, *11*, 237–245, DOI: [10.1016/j.jcf.2011.12.005](https://doi.org/10.1016/j.jcf.2011.12.005) (page 119).
- (544) Rosenfeld, M.; Wainwright, C. E.; Higgins, M.; Wang, L. T.; McKee, C.; Campbell, D.; Tian, S.; Schneider, J.; Cunningham, S.; Davies, J. C.; ARRIVAL study group *The Lancet. Respiratory Medicine* **2018**, *6*, 545–553, DOI: [10.1016/S2213-2600\(18\)30202-9](https://doi.org/10.1016/S2213-2600(18)30202-9) (page 119).
- (545) Lopes-Pacheco, M.; Boinot, C.; Sabirzhanova, I.; Rapino, D.; Cebotaru, L. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* **2017**, *41*, 2194–2210, DOI: [10.1159/000475578](https://doi.org/10.1159/000475578) (page 119).

- (546) Dekkers, J. F.; Gondra, R. A. G.; Kruisselbrink, E.; Vonk, A. M.; Janssens, H. M.; Groot, K. M. d. W.-d.; van der Ent, C. K.; Beekman, J. M. *European Respiratory Journal* **2016**, *48*, 451–458, DOI: [10.1183/13993003.01192-2015](https://doi.org/10.1183/13993003.01192-2015) (page 119).
- (547) Wainwright, C. E. et al. *The New England Journal of Medicine* **2015**, *373*, 220–231, DOI: [10.1056/NEJMoa1409547](https://doi.org/10.1056/NEJMoa1409547) (pages 119, 124).
- (548) Rowe, S. M.; Daines, C.; Ringshausen, F. C.; Kerem, E.; Wilson, J.; Tullis, E.; Nair, N.; Simard, C.; Han, L.; Ingenito, E. P.; McKee, C.; Lekstrom-Himes, J.; Davies, J. C. *New England Journal of Medicine* **2017**, *377*, 2024–2035, DOI: [10.1056/NEJMoa1709847](https://doi.org/10.1056/NEJMoa1709847) (pages 119, 120).
- (549) Middleton, P. G. et al. *New England Journal of Medicine* **2019**, *381*, 1809–1819, DOI: [10.1056/NEJMoa1908639](https://doi.org/10.1056/NEJMoa1908639) (page 119).
- (550) Boyle, M. P.; Bell, S. C.; Konstan, M. W.; McColley, S. A.; Rowe, S. M.; Rietschel, E.; Huang, X.; Waltz, D.; Patel, N. R.; Rodman, D.; VX09-809-102 study group *The Lancet. Respiratory Medicine* **2014**, *2*, 527–538, DOI: [10.1016/S2213-2600\(14\)70132-8](https://doi.org/10.1016/S2213-2600(14)70132-8) (pages 119, 124, 159).
- (551) Hamosh, A.; Rosenstein, B. J.; Cutting, G. R. *Human Molecular Genetics* **1992**, *1*, 542–544, DOI: [10.1093/hmg/1.7.542](https://doi.org/10.1093/hmg/1.7.542) (page 120).
- (552) Valley, H. C.; Bukis, K. M.; Bell, A.; Cheng, Y.; Wong, E.; Jordan, N. J.; Allaire, N. E.; Sivachenko, A.; Liang, F.; Bihler, H.; Thomas, P. J.; Mahiou, J.; Mense, M. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2019**, *18*, 476–483, DOI: [10.1016/j.jcf.2018.12.001](https://doi.org/10.1016/j.jcf.2018.12.001) (page 120).
- (553) Seibert, F. S.; Linsdell, P.; Loo, T. W.; Hanrahan, J. W.; Riordan, J. R.; Clarke, D. M. *Journal of Biological Chemistry* **1996**, *271*, 27493–27499, DOI: [10.1074/jbc.271.44.27493](https://doi.org/10.1074/jbc.271.44.27493) (pages 120, 124).
- (554) Larsson, P.; Kneislz, R. C.; Marklund, E. G. *Journal of Computational Chemistry* **2020**, *41*, 1564–1569, DOI: [10.1002/jcc.26198](https://doi.org/10.1002/jcc.26198) (page 127).
- (555) Olesen, K.; Awasthi, N.; Bruhn, D. S.; Pezeshkian, W.; Khandelia, H. *Journal of Chemical Theory and Computation* **2018**, *14*, 3342–3350, DOI: [10.1021/acs.jctc.8b00267](https://doi.org/10.1021/acs.jctc.8b00267) (page 127).
- (556) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. *Computer Physics Communications* **2009**, *180*, 1961–1972, DOI: [10.1016/j.cpc.2009.05.011](https://doi.org/10.1016/j.cpc.2009.05.011) (page 127).
- (557) Bonomi, M. et al. *Nature Methods* **2019**, *16*, 670–673, DOI: [10.1038/s41592-019-0506-8](https://doi.org/10.1038/s41592-019-0506-8) (page 127).
- (558) Grossfield, A. WHAM: The Weighted Histogram Analysis Method, version 2.2, 2012 (pages 127, 152).
- (559) Kim, J.-B. *Korean Journal of Pediatrics* **2014**, *57*, 1–18, DOI: [10.3345/kjp.2014.57.1.1](https://doi.org/10.3345/kjp.2014.57.1.1) (page 129).
- (560) El Hiani, Y.; Linsdell, P. *The Journal of Biological Chemistry* **2010**, *285*, 32126–32140, DOI: [10.1074/jbc.M110.113332](https://doi.org/10.1074/jbc.M110.113332) (page 129).

- (561) Lee, J. H.; Richter, W.; Namkung, W.; Kim, K. H.; Kim, E.; Conti, M.; Lee, M. G. *Journal of Biological Chemistry* **2007**, *282*, 10414–10422, DOI: [10.1074/jbc.M610857200](https://doi.org/10.1074/jbc.M610857200) (page 131).
- (562) Sheppard, D. N.; Welsh, M. J. *Physiological Reviews* **1999**, *79*, S23–S45, DOI: [10.1152/physrev.1999.79.1.S23](https://doi.org/10.1152/physrev.1999.79.1.S23) (page 131).
- (563) Gao, L.; Kim, K. J.; Yankaskas, J. R.; Forman, H. J. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **1999**, *277*, L113–L118, DOI: [10.1152/ajplung.1999.277.1.L113](https://doi.org/10.1152/ajplung.1999.277.1.L113) (page 131).
- (564) LaRusch, J. et al. *PLoS genetics* **2014**, *10*, e1004376, DOI: [10.1371/journal.pgen.1004376](https://doi.org/10.1371/journal.pgen.1004376) (page 131).
- (565) Negoda, A.; Cowley, E. A.; El Hiani, Y.; Linsdell, P. *Cellular and Molecular Life Sciences* **2018**, *75*, 3027–3038, DOI: [10.1007/s00018-018-2777-0](https://doi.org/10.1007/s00018-018-2777-0) (pages 131, 132, 141).
- (566) Cui, G.; Rahman, K. S.; Infield, D. T.; Kuang, C.; Prince, C. Z.; McCarty, N. A. *Journal of General Physiology* **2014**, *144*, 159–179, DOI: [10.1085/jgp.201311122](https://doi.org/10.1085/jgp.201311122) (pages 131, 135, 142, 143).
- (567) Simon, M. A.; Csanády, L. *eLife* **2021**, *10*, ed. by Jara-Oseguera, A.; Aldrich, R. W.; Jara-Oseguera, A.; Hwang, T.-C., e74693, DOI: [10.7554/eLife.74693](https://doi.org/10.7554/eLife.74693) (pages 131, 142, 143).
- (568) Black, K. A.; He, S.; Jin, R.; Miller, D. M.; Bolla, J. R.; Clarke, O. B.; Johnson, P.; Windley, M.; Burns, C. J.; Hill, A. P.; Laver, D.; Robinson, C. V.; Smith, B. J.; Gulbis, J. M. *Nature Communications* **2020**, *11*, 3024–3024, DOI: [10.1038/s41467-020-16842-0](https://doi.org/10.1038/s41467-020-16842-0) (page 132).
- (569) Bergh, C.; Heusser, S. A.; Howard, R.; Lindahl, E. *eLife* **2021**, *10*, ed. by Allen, T. W.; Faraldo-Gómez, J. D.; Poitevin, F., e68369, DOI: [10.7554/eLife.68369](https://doi.org/10.7554/eLife.68369) (pages 132, 139).
- (570) McComas, S. E.; Mitrovic, D.; Alleva, C.; Bonaccorsi, M.; Drew, D.; Delemotte, L. Determinants of Sugar-Induced Influx in the Mammalian Fructose Transporter GLUT5, 2022, DOI: [10.1101/2022.06.17.495601](https://doi.org/10.1101/2022.06.17.495601) (page 132).
- (571) Pearson, K. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1901**, *2*, 559–572, DOI: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720) (page 132).
- (572) Hotelling, H. *Biometrika* **1936**, *28*, 321, DOI: [10.2307/2333955](https://doi.org/10.2307/2333955) (page 132).
- (573) Smart, O. S.; Neduvvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. P. *Journal of Molecular Graphics* **1996**, *14*, 354–360, DOI: [10.1016/S0263-7855\(97\)00009-X](https://doi.org/10.1016/S0263-7855(97)00009-X) (page 135).
- (574) Wang, G.; Linsley, R.; Norimatsu, Y. *The FEBS Journal* **2016**, *283*, 2458–2475, DOI: [10.1111/febs.13752](https://doi.org/10.1111/febs.13752) (pages 135, 143).
- (575) Zhang, Z. R.; Cui, G.; Liu, X.; Song, B.; Dawson, D. C.; McCarty, N. A. *Journal of Biological Chemistry* **2005**, *280*, 458–468, DOI: [10.1074/jbc.M409626200](https://doi.org/10.1074/jbc.M409626200) (pages 135, 143, 144).

- (576) Josephson, R. K.; Schwab, W. E. *Journal of General Physiology* **1979**, *74*, 213–236, DOI: [10.1085/jgp.74.2.213](https://doi.org/10.1085/jgp.74.2.213) (page 137).
- (577) Kuyucak, S.; Andersen, O. S.; Chung, S.-H. *Reports on Progress in Physics* **2001**, *64*, 1427–1472, DOI: [10.1088/0034-4885/64/11/202](https://doi.org/10.1088/0034-4885/64/11/202) (page 139).
- (578) Moradi, M.; Enkavi, G.; Tajkhorshid, E. *Nature Communications* **2015**, *6*, 8393, DOI: [10.1038/ncomms9393](https://doi.org/10.1038/ncomms9393) (pages 139, 165).
- (579) Grus, J., *Data Science from Scratch: First Principles with Python*; O'Reilly Media, Inc, USA: Beijing Köln, 2015; 330 pp. (pages 139, 146).
- (580) Miller, C. C. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1924**, *106*, 724–749, DOI: [10.1098/rspa.1924.0100](https://doi.org/10.1098/rspa.1924.0100) (page 141).
- (581) Wang, W.; Linsdell, P. *Journal of Biological Chemistry* **2012**, *287*, 32136–32146, DOI: [10.1074/jbc.M112.385096](https://doi.org/10.1074/jbc.M112.385096) (pages 141, 143).
- (582) Hutter, C. A. J.; Timachi, M. H.; Hürlimann, L. M.; Zimmermann, I.; Egloff, P.; Göddeke, H.; Kucher, S.; Štefanić, S.; Karttunen, M.; Schäfer, L. V.; Bordignon, E.; Seeger, M. A. *Nature Communications* **2019**, *10*, 2260, DOI: [10.1038/s41467-019-10989-6](https://doi.org/10.1038/s41467-019-10989-6) (page 141).
- (583) Invernizzi, M.; Parrinello, M. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2731–2736, DOI: [10.1021/acs.jpclett.0c00497](https://doi.org/10.1021/acs.jpclett.0c00497) (pages 142, 148, 149).
- (584) Invernizzi, M.; Parrinello, M. *Journal of Chemical Theory and Computation* **2022**, *acs.jctc.2c00152*, DOI: [10.1021/acs.jctc.2c00152](https://doi.org/10.1021/acs.jctc.2c00152) (page 142).
- (585) Smith, Z.; Pramanik, D.; Tsai, S.-T.; Tiwary, P. *The Journal of Chemical Physics* **2018**, *149*, 234105, DOI: [10.1063/1.5064856](https://doi.org/10.1063/1.5064856) (page 142).
- (586) Tiwary, P.; Berne, B. J. *Proceedings of the National Academy of Sciences* **2016**, *113*, 2839–2844, DOI: [10.1073/pnas.1600917113](https://doi.org/10.1073/pnas.1600917113) (page 142).
- (587) Gong, X.; Linsdell, P. *Journal of General Physiology* **2003**, *122*, 673–687, DOI: [10.1085/jgp.200308889](https://doi.org/10.1085/jgp.200308889) (pages 142, 144).
- (588) Gong, X.; Linsdell, P. *Journal of Physiology* **2003**, *549*, 375–385, DOI: [10.1113/jphysiol.2002.038216](https://doi.org/10.1113/jphysiol.2002.038216) (page 142).
- (589) Tabcharani, J.; Rommens, J.; Hou, Y.-X.; Chang, X.-B.; Tsui, L.-C.; Riordan, J.; Hanrahan, J. *Nature* **1993**, *366*, 79–82, DOI: [10.1038/366079a0](https://doi.org/10.1038/366079a0) (page 142).
- (590) Zhou, Z.; Hu, S.; Hwang, T.-C. *Journal of General Physiology* **2002**, *120*, 647–662, DOI: [10.1085/jgp.20028685](https://doi.org/10.1085/jgp.20028685) (page 142).
- (591) Yeh, H.-I.; Yeh, J.-T.; Hwang, T.-C. *The Journal of General Physiology* **2015**, *145*, 47–60, DOI: [10.1085/jgp.201411272](https://doi.org/10.1085/jgp.201411272) (page 142).
- (592) Cottrill, K. A.; Farinha, C. M.; McCarty, N. A. *Communications Biology* **2020**, *3*, 179–179, DOI: [10.1038/s42003-020-0909-1](https://doi.org/10.1038/s42003-020-0909-1) (page 142).
- (593) Lin, Y.; Buyan, A.; Corry, B. *The Journal of General Physiology* **2022**, *154*, e202113064, DOI: [10.1085/jgp.202113064](https://doi.org/10.1085/jgp.202113064) (page 142).
- (594) Kapoor, K.; Pant, S.; Tajkhorshid, E. *Chemical Science* **2021**, *12*, 6293–6306, DOI: [10.1039/DOSC06288J](https://doi.org/10.1039/DOSC06288J) (page 142).

- (595) Farinha, C. M.; Miller, E.; McCarty, N. *Journal of Cystic Fibrosis* **2018**, *17*, S9–S13, DOI: [10.1016/j.jcf.2017.08.014](https://doi.org/10.1016/j.jcf.2017.08.014) (page 142).
- (596) Cui, G.; Cottrill, K. A.; Strickland, K. M.; Imhoff, B. R.; McCarty, N. A. *Biophysical Journal* **2020**, *118*, 588a, DOI: [10.1016/j.bpj.2019.11.3187](https://doi.org/10.1016/j.bpj.2019.11.3187) (page 142).
- (597) Chin, S.; Ramjeesingh, M.; Hung, M.; Ereño-Oreba, J.; Ing, C.; Zeng, Z. W.; Cui, H.; Pomès, R.; Julien, J.-P.; Bear, C. E. Lipid Interactions Enhance Activation and Potentiation of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), 2018, DOI: [10.1101/495010](https://doi.org/10.1101/495010) (page 142).
- (598) Hickey, K. D.; Buhr, M. M. *Journal of Lipids* **2011**, *2011*, 208457, DOI: [10.1155/2011/208457](https://doi.org/10.1155/2011/208457) (page 142).
- (599) Aureli, M.; Schiumarini, D.; Loberto, N.; Bassi, R.; Tamanini, A.; Mancini, G.; Tironi, M.; Munari, S.; Cabrini, G.; Dechechchi, M. C.; Sonnino, S. *Chemistry and Physics of Lipids* **2016**, *200*, 94–103, DOI: [10.1016/j.chmplyslip.2016.08.002](https://doi.org/10.1016/j.chmplyslip.2016.08.002) (page 142).
- (600) Wilson, K. A.; Fairweather, S. J.; MacDermott-Opeskin, H. I.; Wang, L.; Morris, R. A.; O’Mara, M. L. *The Journal of Chemical Physics* **2021**, *154*, 095101, DOI: [10.1063/5.0040887](https://doi.org/10.1063/5.0040887) (page 143).
- (601) Zhang, Z. R.; Song, B.; McCarty, N. A. *Journal of Biological Chemistry* **2005**, *280*, 41997–42003, DOI: [10.1074/jbc.M510242200](https://doi.org/10.1074/jbc.M510242200) (pages 143, 144).
- (602) Csanády, L. *Journal of General Physiology* **2017**, *149*, 413–416, DOI: [10.1085/jgp.201711777](https://doi.org/10.1085/jgp.201711777) (pages 144, 165).
- (603) Van Willigen, M.; Vonk, A. M.; Yeoh, H. Y.; Kruisselbrink, E.; Kleizen, B.; Van Der Ent, C. K.; Egmond, M. R.; De Jonge, H. R.; Braakman, I.; Beekman, J. M.; Van Der Sluijs, P. *Life Science Alliance* **2019**, *2*, DOI: [10.26508/lса.201800172](https://doi.org/10.26508/lsa.201800172) (pages 145, 154).
- (604) Universitaire Ziekenhuizen Leuven *Response to CFTR-modulators in Intestinal Organoids of Patients With CF Having at Least One R334W Mutation*; Clinical trial registration NCT04254705; clinicaltrials.gov, 2020 (page 145).
- (605) Bock, L. V.; Grubmüller, H. *Nature Communications* **2022**, *13*, 1709, DOI: [10.1038/s41467-022-29332-2](https://doi.org/10.1038/s41467-022-29332-2) (page 145).
- (606) Kim, Y.; Chen, J. *Science* **2018**, *359*, 915–919, DOI: [10.1126/science.aar7389](https://doi.org/10.1126/science.aar7389) (page 145).
- (607) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2015**, *11*, 1864–1874, DOI: [10.1021/ct5010406](https://doi.org/10.1021/ct5010406) (page 148).
- (608) Tiwary, P.; Parrinello, M. *The Journal of Physical Chemistry B* **2015**, *119*, 736–742, DOI: [10.1021/jp504920s](https://doi.org/10.1021/jp504920s) (page 149).
- (609) Marcus, D. Ouroboros, 2021 (page 153).
- (610) Kim, J.; Davies, Z.; Dunn, C.; Wine, J. J.; Milla, C. *Journal of Cystic Fibrosis* **2018**, *17*, 179–185, DOI: [10.1016/j.jcf.2017.12.005](https://doi.org/10.1016/j.jcf.2017.12.005) (page 154).

- (611) Trikafta FDA Information https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/212273s008lbl.pdf (accessed 09/07/2022) (pages 155, 159, 161).
- (612) Bompadre, S. G.; Sohma, Y.; Li, M.; Hwang, T.-C. *The Journal of General Physiology* **2007**, *129*, 285–298, DOI: [10.1085/jgp.200609667](https://doi.org/10.1085/jgp.200609667) (pages 156, 158).
- (613) Wang, W.; Fu, L.; Liu, Z.; Wen, H.; Rab, A.; Hong, J. S.; Kirk, K. L.; Rowe, S. M. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **2020**, *319*, L770–L785, DOI: [10.1152/AJPLUNG.00262.2019](https://doi.org/10.1152/AJPLUNG.00262.2019) (page 156).
- (614) Li, C.; Ramjeesingh, M.; Wang, W.; Garami, E.; Hewryk, M.; Lee, D.; Rommens, J. M.; Galley, K.; Bear, C. E. *Journal of Biological Chemistry* **1996**, *271*, 28463–28468, DOI: [10.1074/jbc.271.45.28463](https://doi.org/10.1074/jbc.271.45.28463) (page 156).
- (615) Bompadre, S. G.; Li, M.; Hwang, T.-C. *Journal of Biological Chemistry* **2008**, *283*, 5364–5369, DOI: [10.1074/jbc.M709417200](https://doi.org/10.1074/jbc.M709417200) (page 156).
- (616) Awatade, N. T.; Ramalho, S.; Silva, I. A. L.; Felício, V.; Botelho, H. M.; de Poel, E.; Vonk, A.; Beekman, J. M.; Farinha, C. M.; Amaral, M. D. *Journal of Cystic Fibrosis: Official Journal of the European Cystic Fibrosis Society* **2019**, *18*, 182–189, DOI: [10.1016/j.jcf.2018.07.001](https://doi.org/10.1016/j.jcf.2018.07.001) (page 157).
- (617) Lopes-Pacheco, M.; Sabirzhanova, I.; Rapino, D.; Morales, M. M.; Guggino, W. B.; Cebotaru, L. *Chembiochem : a European journal of chemical biology* **2016**, *17*, 493–505, DOI: [10.1002/cbic.201500620](https://doi.org/10.1002/cbic.201500620) (page 157).
- (618) Casals, T et al. *Human Mutation* **1997**, *10*, 387–392, DOI: [10.1002/\(SICI\)1098-1004\(1997\)10:5<387::AID-HUMU9>3.0.CO;2-C](https://doi.org/10.1002/(SICI)1098-1004(1997)10:5<387::AID-HUMU9>3.0.CO;2-C) (page 157).
- (619) Cotten, J. F.; Ostedigaard, L. S.; Carson, M. R.; Welsh, M. J. *Journal of Biological Chemistry* **1996**, *271*, 21279–21284, DOI: [10.1074/jbc.271.35.21279](https://doi.org/10.1074/jbc.271.35.21279) (page 157).
- (620) Penmatsa, H.; Frederick, C. A.; Nekkalapu, S.; Conoley, V. G.; Zhang, W.; Li, C.; Kappes, J.; Stokes, D. C.; Naren, A. P. *Pediatric Pulmonology* **2009**, *44*, 1003–1009, DOI: [10.1002/ppul.21092](https://doi.org/10.1002/ppul.21092) (page 157).
- (621) Ensinck, M. M.; Keersmaecker, L. D.; Ramalho, A. S.; Cuyx, S.; Biervliet, S. V.; Dupont, L.; Christ, F.; Debyser, Z.; Vermeulen, F.; Carlon, M. S. *ERJ Open Research* **2022**, DOI: [10.1183/23120541.00716-2021](https://doi.org/10.1183/23120541.00716-2021) (page 158).
- (622) Kleizen, B.; Hunt, J. F.; Callebaut, I.; Hwang, T. C.; Sermet-Gaudelus, I.; Hafkemeyer, S.; Sheppard, D. N. *Journal of Cystic Fibrosis* **2020**, *19*, S19–S24, DOI: [10.1016/j.jcf.2019.10.021](https://doi.org/10.1016/j.jcf.2019.10.021) (page 159).
- (623) Padanyi, R.; Farkas, B.; Tordai, H.; Kiss, B.; Grubmuller, H.; Soya, N.; Gergely L. Lukacs; Kellermayer, M.; Hegedus, T. *Computational and Structural Biotechnology Journal* **2022**, *20*, 2587–2599, DOI: [10.1016/j.csbj.2022.05.036](https://doi.org/10.1016/j.csbj.2022.05.036) (page 159).
- (624) Kalydeco FDA Approval (pages 159, 161).

- (625) Donaldson, S. H.; Pilewski, J. M.; Grieser, M.; Cooke, J.; Viswanathan, L.; Tullis, E.; Davies, J. C.; Lekstrom-Himes, J. A.; Wang, L. T. *American Journal of Respiratory and Critical Care Medicine* **2018**, *197*, 214–224, DOI: [10.1164/rccm.201704-07170C](https://doi.org/10.1164/rccm.201704-07170C) (page 159).
- (626) Casalino, L. et al. *The International Journal of High Performance Computing Applications* **2021**, *35*, 432–451, DOI: [10.1177/10943420211006452](https://doi.org/10.1177/10943420211006452) (page 165).
- (627) Hofmann, S.; Januliene, D.; Mehdipour, A. R.; Thomas, C.; Stefan, E.; Brüchert, S.; Kuhn, B. T.; Geertsma, E. R.; Hummer, G.; Tampé, R.; Moeller, A. *Nature* **2019**, *571*, 580–583, DOI: [10.1038/s41586-019-1391-0](https://doi.org/10.1038/s41586-019-1391-0) (page 165).
- (628) Orellana, L.; Yoluk, O.; Carrillo, O.; Orozco, M.; Lindahl, E. *Nature Communications* **2016**, *7*, 12575, DOI: [10.1038/ncomms12575](https://doi.org/10.1038/ncomms12575) (page 165).
- (629) Melo, M. C. R.; Bernardi, R. C.; de la Fuente-Nunez, C.; Luthey-Schulten, Z. *The Journal of Chemical Physics* **2020**, *153*, 134104, DOI: [10.1063/5.0018980](https://doi.org/10.1063/5.0018980) (page 166).
- (630) Bones Don't Rust, in collab. with The Mountain Goats., 2022 (page 169).
- (631) Caterina, M. J.; Schumacher, M. A.; Tominaga, M.; Rosen, T. A.; Levine, J. D.; Julius, D. *Nature* **1997**, *389*, 816–824, DOI: [10.1038/39807](https://doi.org/10.1038/39807) (page 170).
- (632) Yu, H.; Yang, Z.; Li, F.; Xu, L.; Sun, Y. *Drug Delivery* **2020**, *27*, 1425–1437, DOI: [10.1080/10717544.2020.1831103](https://doi.org/10.1080/10717544.2020.1831103) (page 172).
- (633) Baym, M.; Lieberman, T. D.; Kelsic, E. D.; Chait, R.; Gross, R.; Yelin, I.; Kishony, R. *Science (New York, N.Y.)* **2016**, *353*, 1147–1151, DOI: [10.1126/science.aag0822](https://doi.org/10.1126/science.aag0822) (page 172).