

Doctoral School in *Sciences and Technologies of Chemistry and Materials*

XXXV cycle

Curriculum: *Pharmaceutical, Food and Cosmetic Sciences*

***In-silico* methods applied on druggable proteins to identify transient  
pockets: new approaches for studying drug-target molecular mechanisms.**

**A case study on CFTR.**

Doctoral Dissertation of ***Matteo Uggeri***

Supervisor: Prof.ssa Paola Fossa

Co-Supervisor: Dr.ssa Pasqualina D'Ursi

Academic years: 2019-2022



## Preface

I hereby declare that the subjects presented in this thesis dissertation are the results of the research work carried out throughout my Ph.D. course.

The research that will be presented in this Ph.D. thesis concerns three topics, all characterized by a common thread: the study of proteins of therapeutic interest.

In the first part of the thesis will be presented a drug repositioning study to identify drugs able to rescue the mutated cystic fibrosis transmembrane conductance regulator using and manually curated in-house ligand dataset. In the second part of the thesis, in chapter 1, the characterization of the immunoglobulin-like domain 1 of the leucine-rich repeats and immunoglobulin-like domains protein 2 in the homodimerization and the study of its deletion and pathogenic mutations on the dimerization will be presented. While, in chapter 2, the study of the dynamic differences between the wild-type and mutated unstructured N-terminal of the nucleoporin 98 protein will be discussed.

These results were developed over a two-year period at the Department of Pharmacy (DIFAR) in Genova (Italy) and a one-year period as a guest at the National Research Council in the Institute for Biomedical Technologies (ITB-CNR) in Segrate (Italy), from which I benefited of their High-Performance Computing (HPC) infrastructures, under the supervision of Prof.ssa Paola Fossa and the co-supervision of Dott.ssa Pasqualina D'Ursi (ITB-CNR).

Then, due to the covid pandemic and lockdown restrictions limiting the abroad stages, I spent a six-month training at the Department of Molecular and Translational Medicine at the University of Brescia under the supervision of Prof. Arnaldo Caruso, further continuing the study of the p17 matrix protein of the Human Immunodeficiency Virus (HIV), after having completed and published the phylogenetic study of this protein.



## Papers Contribution

### Published articles contribution:

- In the article: Colombo EA, Valiante M, Uggeri M, Orro A, Majore S, Grammatico P, Gentilini D, Gervasini C, Finelli P, D'Ursi P, Larizza L. Germline biallelic Nucleoporin 98 variants in two siblings presenting a Rothmund-Thomson like spectrum: functional changes borne out by protein molecular modelling studies. *International Journal of Molecular Sciences*. 2023, 24(4), 4028. doi: 10.3390/ijms24044028. I contributed by doing the computational analyses, plus writing, reviewing, and editing the paper.
- In the article: Fossa P\$, Uggeri M\$, Orro A, Urbinati C, Rondina A, Milanese M, Pedemonte N, Pesce E, Padoan R, Ford RC, Meng X, Rusnati M, D'Ursi P. (\$ Authors equally contributed to this work) Virtual Drug Repositioning as a Tool to Identify Natural Small Molecules That Synergize with Lumacaftor in F508del-CFTR Binding and Rescuing. *Int J Mol Sci*. 2022 Oct 14;23(20):12274. doi: 10.3390/ijms232012274. PMID: 36293130; PMCID: PMC9602983. I contributed by doing the computational analyses, plus doing the conceptualization, writing, and editing of the paper.
- In the article: Orro A\$, Uggeri M\$, Rusnati M, Urbinati C, Pedemonte N, Pesce E, Moscatelli M, Padoan R, Cichero E, Fossa P, D'Ursi P. (\$ Authors equally contributed to this work) In silico drug repositioning on F508del-CFTR: A proof-of-concept study on the AIFA library. *Eur J Med Chem*. 2021 Mar 5;213:113186. doi: 10.1016/j.ejmech.2021.113186. Epub 2021 Jan 13. PMID: 33472120. I contributed by doing the computational analyses, plus helping in the writing, reviewing, and editing of the paper.
- In the article: Caccuri F, D'Ursi P, Uggeri M, Bugatti A, Mazzuca P, Zani A, Filippini F, Salmona M, Ribatti D, Slevin M, Orro A, Lu W, Liò P, Gallo RC, Caruso A. Evolution toward beta common chain receptor usage links the matrix proteins of HIV-1 and its ancestors to human erythropoietin. *Proc Natl Acad Sci U S A*. 2021 Jan 12;118(2):e2021366118. doi: 10.1073/pnas.2021366118. PMID: 33372148; PMCID: PMC7812818. I contributed by performing the computational analysis, plus helping in the writing of the computational results.
- In the article: Rusnati M, D'Ursi P, Pedemonte N, Urbinati C, Ford RC, Cichero E, Uggeri M, Orro A, Fossa P. Recent Strategic Advances in CFTR Drug Discovery: An Overview. *Int J Mol Sci*. 2020

*Mar 31;21(7):2407. doi: 10.3390/ijms21072407. PMID: 32244346; PMCID: PMC7177952. I*  
contributed by helping with the writing, the review, and the editing of the paper.



Prof. Paola Fossa Department  
of Pharmacy,  
Section of Medicinal Chemistry, School  
of Medical and Pharmaceutical  
Sciences,  
University of Genova,  
Viale Benedetto XV 3, 16132 Genova, Italy.  
paola.fossa@unige.it

Genova, 12th December 2022

We hereby declare that Dott. Matteo Uggeri originally contributed to computational development, writing, editing, and reviewing of those papers that have been inserted partially into his PhD thesis dissertation, as it is reported in the Authors Contribution Section of each paper.

The inclusion of extracts from the above mentioned papers in Matteo Uggeri PhD thesis are in agreement with European Journal of Medicinal Chemistry and International Journal of Molecular Sciences journal policy.

Yours faithfully

Paola Fossa and Pasqualina D'Ursi

## Policy

### European Journal of Medicinal Chemistry:

Can I include/use my article in my thesis/dissertation?

Yes. Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.

<https://www.elsevier.com/about/policies/copyright/permissions>

Articles:

- Orro A.<sup>§</sup>; Uggeri M.<sup>§</sup>; Rusnati M.; Urbinati C.; Pedemonte N.; Pesce E.; Moscatelli M.; Padoan R.; Cichero E.; Fossa P.; D'Ursi P. (§ equally contributed to this work). In silico drug repositioning on F508del-CFTR: a proof-of-concept study on the AIFA library. *European Journal of Medicinal Chemistry* 2021, 213, 113186. DOI: <https://doi.org/10.1016/j.ejmech.2021.113186>

### International Journal of Molecular Sciences

No special permission is required to reuse all or part of article published by MDPI, including figures and tables. For articles published under an open access Creative Common CC BY license, any part of the

article may be reused without permission provided that the original article is clearly cited. Reuse of an article does not imply endorsement by the authors or MDPI.

<https://www.mdpi.com/openaccess>

Articles:

- Colombo EA, Valiante M, Uggeri M, Orro A, Majore S, Grammatico P, Gentilini D, Gervasini C, Finelli P, D'Ursi P, Larizza L. Germline biallelic Nucleoporin 98 variants in two siblings presenting a Rothmund-Thomson like spectrum: functional changes borne out by protein molecular modelling studies. *International Journal of Molecular Sciences*. 2023, 24(4), 4028. DOI: 10.3390/ijms24044028.
- Fossa P.<sup>§</sup>; Uggeri M.<sup>§</sup>; Orro A.; Urbinati A.; Rondina A.; Milanesi M.; Pedemonte N.; Pesce E.; Padoan R.; Ford R.C.; Meng X.; Rusnati M.; D'Ursi P. (§ equally contributed to this work). Virtual drug repositioning as a tool to identify natural small molecules that synergize with lumacaftor



in F508del-CFTR binding and rescuing. *International Journal of Molecular Sciences*. 2022, 14;23(20):12274. DOI: 10.3390/ijms232012274

- M. Rusnati, P. D'Ursi, N. Pedemonte, C. Urbinati, R.C. Ford, E. Cichero, M. Uggeri, A. Orro, P. Fossa. Recent Strategic Advances in CFTR Drug Discovery: An Overview. *International Journal of Molecular Sciences*. 2020, 21(7):2407. DOI: <https://doi.org/10.1073/pnas.2021366118>

## Abstract

Cystic Fibrosis is the most common genetical lethal disorder in Caucasians and it is caused by the mutation of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) protein. Up to now, for the treatment of cystic fibrosis patients carrying at least one copy of CFTR deleted of the phenylalanine 508 (F508del-CFTR), the worldwide most frequent mutation, only four drugs have been approved to be used in combination or alone. All the approved compounds have been developed, studied, and are currently commercialized by Vertex Pharmaceuticals. Despite the benefits of these marketed drugs, they are still too expensive for many countries, and they cannot be prescribed to all patients. Thus remains a pressing need to better understand the CFTR structure-function relationship, and the binding site and molecular mechanism of already approved drugs, to identify other CFTR modulators for the rescue of the mutated protein, in particular, F508del-CFTR.

On these bases my research activity has been focused on a deep study of the protein function, investigating its three-dimensional structure and dynamics in complex with the already approved CF drug lumacaftor and new possible CFTR modulators by means of drug repositioning. An optimized model, obtained before the starting of my PhD, of the F508del-CFTR protein and a library of pockets, in which an interesting large druggable pocket (DP1) was identified using lumacaftor as a template, has been used for the following drug repositioning strategy. An in-house database which included 846 drugs and nutraceuticals approved by AIFA (actually implemented to more than 10000 molecules from AIFA and Drugbank database) was built, drawing their 3D structure with the right protonation state of the drugs, and then screened by docking against F508del-CFTR. Among the best eleven repositioned compounds identified within this procedure, tadalafil was one that has been already taken into consideration for cystic fibrosis therapy, confirming the goodness of this approach. Quercetin emerged as the best ligand among the eleven selected, suggesting that small molecules could give a consistent contribution in the search for new CFTR modulators. Focusing on this concept, the several DP1 sub-pockets surrounding the lumacaftor binding region were explored, searching for the most druggable ones and in the meantime scouting small molecules able to fill the transient druggable DP1 sub-pockets and synergize with lumacaftor. At the end of this procedure, NAM was found as a possible hit.

Moreover, during my PhD project, my computational studies have been also focused on two proteins of therapeutical interest, which mutations are causative of rare genetical disorders: the Leucine-Rich and ImmunoGlobulin-like domains 2 (LRIG2) and the Nucleoporin 98 (NUP98), whose mutations lead to the Urofacial Syndrome and a phenotype resembling the Rothmund-Thomson syndrome, respectively.

The study of LRIG2 involved the investigation of the role of the first immunoglobulin-like domain (Ig1) of the LRIG2 protein, and its deletion and mutations, in the LRIG2 homodimerization. The LRIG2 homodimerization was predicted *in silico* and its dimerization interface was computationally characterized. Then, by means of accelerated molecular dynamic simulations, the central role of the Ig1 domain in the LRIG2 dimerization was furthermore validated by studying the impact of the Ig1 domain mutations, described in the literature as pathogenic, in the context of the monomeric LRIG2. This advanced molecular dynamic technique allowed to clarify the role of these mutations in the impairment of the LRIG2 homodimerization.

Eventually, regarding the study of NUP98, a novel germline alteration (G28D) located in the unstructured N-terminal of the NUP98, which is characterized by phenylalanine-glycine (FG) repeats, was computationally evaluated. Differences in the dynamic behavior between the wild type and G28D mutated protein were observed, which are produced from a dispersion of the intramolecular cohesion elements (FG repeats) leading to more elongated conformational states of the unstructured N-terminal of the NUP98 mutant in comparison to the wild type. Those differences may affect the role of NUP98 as a multi-docking station for RNA and proteins, and its folding process when a specific interaction is required.

## List of abbreviations used

### Common:

MD - Molecular Dynamic

RMSF – Root-Mean Square Fluctuation

RMSD – Root-Mean Square Deviation

### Part 1:

CF – Cystic Fibrosis

CFTR – Cystic Fibrosis Transmembrane conductance Regulator

NBD – Nuclear Binding Domain

MSD – Membrane Spanning Domain

TMD – TransMembrane Domain

R – Regulatory domain

ICL – InterCellular Loop

ECL – ExtraCellular Loop

F508del – Deletion of Phenylalanine 508

AIFA – Agenzia Italiana del FArmaco

SMILES – Simplified Molecular Input Line Entry System

ER – Endoplasmic Reticulum

HBE – Human Bronchial Epithelial

FEV – Forced Expiratory Volume

Cryo-EM – Cryo-Electron Microscopy

DP1 – Druggable Pocket 1

SPR – Surface Plasmon Resonance Analysis

NAM – Niacinamide

NAD – Nicotinamide Adenine Dinucleotide

NADP – Nicotinamide Adenine Dinucleotide Phosphate

MMGBSA - Molecular Mechanics Generalized Born Surface Area

## **Part 2 – Chapter 1:**

LRIG2 – Leucine-Rich and ImmunoGlobulin-Like domains 2

UFS – Urofacial Syndrome

HPSE2 – HeParanaSE 2

Ig1del-LRIG2 – Leucine-Rich and ImmunoGlobulin-Like domains 2 deleted of the first ImmunoGlobulin-Like domain

LRR – LeucineRich Repeats (LRR)

Ig1del – Deletion of the first ImmunoGlobulin-Like domain

Int-1 – Interface-1

Int-2 – Interface-2

aMD – accelerated MD

BS – Beta-Strand

PCA – Principal Component Analysis

HS – Hinman syndrome

VUS – Variant of Uncertain Significance

## **Part 2 – Chapter 2:**

Nup – Nucleoporin

IDP – Intrinsically Disordered Protein

IDPR – Intrinsically Disordered Protein Region

IDRs – Intrinsically Disordered Region

PDB – Protein Data Bank

NPC – Nuclear Pore Complex

FG – Phenylalanine-Glycine

F-F – Phenylalanine- Phenylalanine

NTR – Nuclear Transporter Receptor

PPI – Protein-Protein Interaction

SASA – Solvent Accessible Surface Area



# INDEX

<b>Aim of the project .....</b>	<b>1</b>
---------------------------------	----------

## PART 1

### Chapter 1 - CFTR

<b>1 Introduction .....</b>	<b>3</b>
1.1 Cystic Fibrosis .....	3
1.2 CFTR: maturation, protein structure, and function.....	6
1.3 CFTR mutations.....	8
1.4 Cystic Fibrosis therapies .....	10
1.5 CFTR modulators .....	11
1.5.1 Read-Through agents .....	11
1.5.2 Stabilizers.....	11
1.5.3 Amplifiers.....	12
1.5.4 Potentiators.....	12
1.5.5 Correctors .....	12
1.6 Correctors putative binding pockets .....	14
1.7 Drug Repositioning .....	16
<b>2 Results .....</b>	<b>18</b>
2.1 In-house manually curated ligand dataset .....	18
2.1.1 In-house ligand dataset for computational experiments.....	18
2.2 Searching for molecules able to rescue the F508del-CFTR .....	20
2.2.1 Computational drug repositioning pipeline to identify molecules able to bind in the putative lumacaftor binding pocket and rescue the F508del-CFTR protein.....	20
2.2.2. Identification of transient druggable DP1 sub-pockets.....	28
2.2.3 Complex stability and molecular interactions of the repositioned drugs .....	32
2.2.4 Surface Plasmon Resonance analysis .....	38
2.2.5 Discussion .....	40
2.3 Small molecules able to synergize with lumacaftor for the F508del-CFTR rescue .....	42

2.3.1 Computational drug repositioning pipeline to identify small molecules able to interact in the DP1 sub-pocket synergizing lumacaftor .....	42
2.3.2 Pocket Fitting Evaluation for Lumacaftor and NAM: Which Molecule Binds First? .....	45
2.3.3 Surface Plasmon Resonance analysis .....	50
2.3.4 Identification of the Binding Pocket of NAM to the apo F508del-CFTR .....	51
2.3.5 Effect of NAM-Lumacaftor Co-Treatment on Mutant CFTR Rescue.....	51
2.3.6 Discussion and conclusions .....	52
<b>3 Materials and Methods .....</b>	<b>57</b>
3.1 Ligand dataset preparation .....	57
3.2 Computational and analytical set-up in the searching for approved drug to rescue the F508del-CFTR. ....	57
3.2.1 Drug-repositioning pipeline for the searching of approved drugs to rescue the F508del-CFTR in the DP1.....	57
3.2.2 Preparation of the dataset for drug repositioning analysis.....	58
3.2.3 Drug repositioning using docking experiment .....	58
3.2.4 Molecular dynamics simulations replica of the F508del-CFTR in complex with the repositioned drugs .....	59
3.2.5 MM-GBSA analyses of the F508del-CFTR-ligand complexes.....	59
3.2.6 Fpocket analysis for the identification of DP1 druggable transient sub-pockets.....	59
3.3 Computational and analytical set-up in the searching for small molecules able to rescue the F508del-CFTR together with lumacaftor .....	60
3.3.1 Drug repositioning pipeline update.....	60
3.3.2 Molecular Dynamic Simulations of the NAM in complex with the F508del-CFTR with or without lumacaftor .....	61

## PART 2

### Chapter 1 - LRIG2

<b>4 Introduction .....</b>	<b>62</b>
4.1 The Urofacial Syndrome .....	62
4.2 LRIG2 protein .....	64
<b>5 Results .....</b>	<b>67</b>



5.1 Identification of a novel LRIG2 pathogenic variant .....	67
5.2 Primary sequence analysis of LRIG1 and LRIG2 Ig-like domains .....	67
5.3 Modelling the dimeric forms of LRIG1-3Ig and LRIG2-3Ig .....	68
5.4 LRIG1-3Ig and LRIG2-3Ig dimer characterization by molecular dynamic simulations.....	69
5.5 Characterization of LRIG2 point mutations in the Ig1 domain .....	74
5.5.1 LRIG2 p.H566Y .....	75
5.5.2 LRIG2 p.R550C .....	78
5.5.3 LRIG2 p.S523R.....	81
5.6 Discussion and conclusions .....	83
<b>6 Materials and Methods .....</b>	<b>86</b>
6.1 LRIG1-3Ig and LRIG2-3Ig models, docking, and molecular dynamic simulation .....	86
6.2 Modelling of LRIG2 variants and accelerated molecular dynamic simulation .....	86

## Chapter 2 - NUP98

<b>7 Introduction .....</b>	<b>88</b>
7.1 Intrinsically Disordered Proteins and Intrinsically Disordered Regions .....	88
7.2 Nucleoporin 98 and FG-repeats motif.....	90
<b>8 Results .....</b>	<b>92</b>
8.1 Identification of novel NUP98 pathogenic variants .....	92
8.2 Molecular modeling study of NUP98 variant .....	92
8.3 Compactness differences evaluation between the WT and the G28D .....	94
8.4 G28D is less intramolecularly cohesive than WT NUP98 .....	96
8.5 RNA interaction impairment in the G28D variant .....	100
8.6 Discussion and conclusions .....	105
<b>9 Materials and Methods .....</b>	<b>108</b>
9.1. Molecular modelling simulation and analysis .....	108

## Summary

<b>10. Methodologies Details .....</b>	<b>109</b>
10.1 Homology Modeling .....	109
10.2 Molecular Docking .....	110
10.3 Molecular Dynamic Simulation.....	112
<b>11. Conclusions and Future Prospectives .....</b>	<b>115</b>
<b>12 Bibliography.....</b>	<b>117</b>

## Aim of the project

The aim of my PhD project has been focused on the study of proteins of therapeutic interest, by means of specific computational methodologies, derived from the computational community, or developed in-house.

The targets considered are proteins where mutations can affect different aspects of the protein function determining a pathological condition, see Cystic Fibrosis Transmembrane Conductance Regulator protein (CFTR) where mutations can affect the protein folding and/or function leading to Cystic Fibrosis (CF), Leucine-Rich repeats and ImmunoGlobulin-like domains 2 (LRIG2) where mutations can affect the protein-protein interaction interface leading to the Urofacial Syndrome (UFS), or the nucleoporin 98 (NUP98) where the G28D mutation in the N-terminal unstructured protein region leads to a phenotype resembling the Rothmund-Thomson syndrome.

To fully characterize and understand the 3D structure and function of these targets, molecular dynamic (MD) simulations let me exhaustively explore their conformational spaces, scouting and finding transient pockets involved in protein-ligand and protein-protein interactions.

CFTR deleted of the phenylalanine 508 (F508del), the worldwide most frequent CF mutation, was used as a case study to find molecules able to help the rescue of the F508del-CFTR. In the cases of CFTR the study of a lumacaftor putative pocket led also to the identification of transient pockets and a specifically tailored pipeline, developed in-house, has been applied for a drug repositioning approach using an in-house manually curated ligand dataset.

Successively, in the second part of the thesis, the computational study of two proteins of therapeutic interest: LRIG2 and NUP98, which mutations led to very rare genetic disorders, will be presented. LRIG2 and NUP98 wild-type (WT) and mutated proteins were studied by means of molecular dynamic (MD) simulations, exploring their dynamical behavior to characterize the molecular mechanisms of the protein malfunctioning.

Finally, during my PhD I also completed the phylogenic study of p17, the matrix protein of the Human Immunodeficiency Virus (HIV) virus, identifying the amino acid residues responsible for the switch-on/off of the p17 angiogenic activity.



# PART 1

## Chapter 1 – CFTR

### 1 Introduction

#### 1.1 Cystic Fibrosis

Cystic Fibrosis (CF) is the most common life-limiting and lethal autosomal recessive disorder in Caucasians with an incidence of about 1/2500 live births. It affects about 32000 individuals in Europe and 85000 all over the world [1], but its frequency is highly variable among ethnic groups.

CF was first identified by Dorothy Andersen in 1938 [2]. It is caused by several mutations in the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) protein. CFTR is an ion chloride channel located in the plasma membrane of epithelial cells, contributing to regulating the chloride and bicarbonates ions concentrations inside and outside the cell (Figure 1.1).

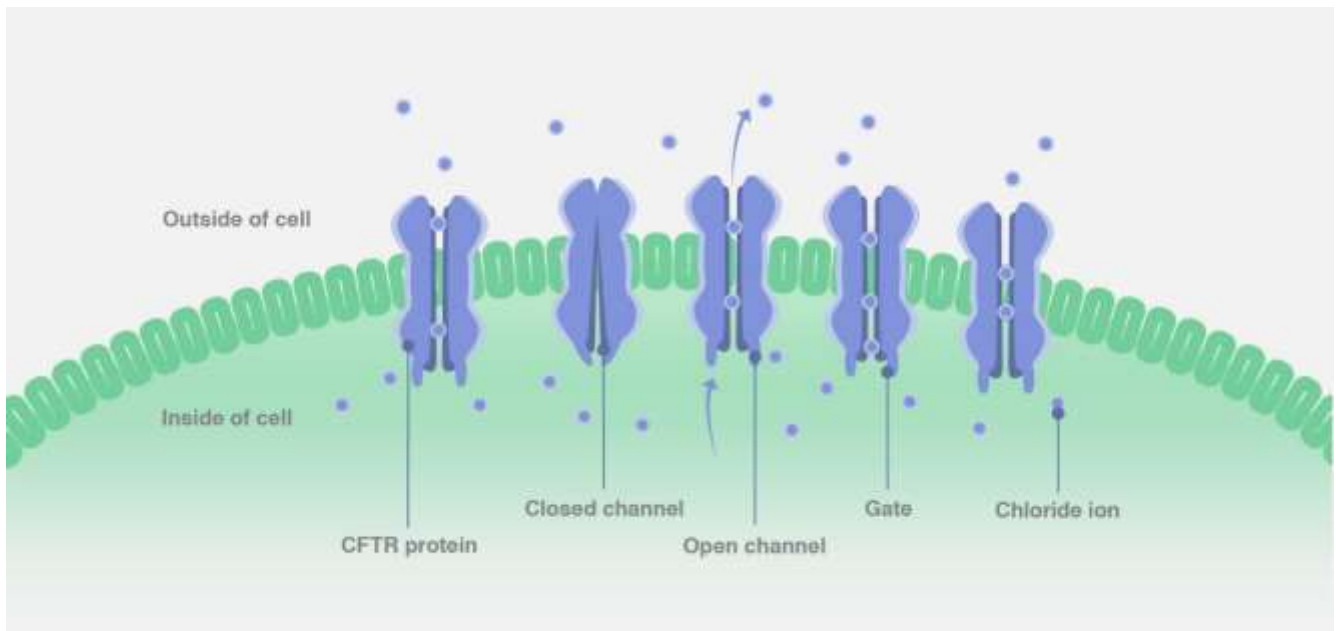


Figure 1.1. A simplified example of the CFTR ion chloride movement regulation inside and outside the cell (Picture from <https://www.cfsource.co.nz/>).

As conceivable, CF is a multi-organ disease due to the loss of the CFTR function in all epithelial tissues. At the time when CF was discovered, the life expectancy of the patients ranged from a few months to a few years. Nowadays, early diagnosis and improved treatments have progressively increased life expectancy up to about 50 years [3]. Despite this, CF patients still have to fight against a large spectrum of complications mainly caused by chloride ion concentration unbalance and chronic obstruction of ducts in multiple organs (Figure 1.2).

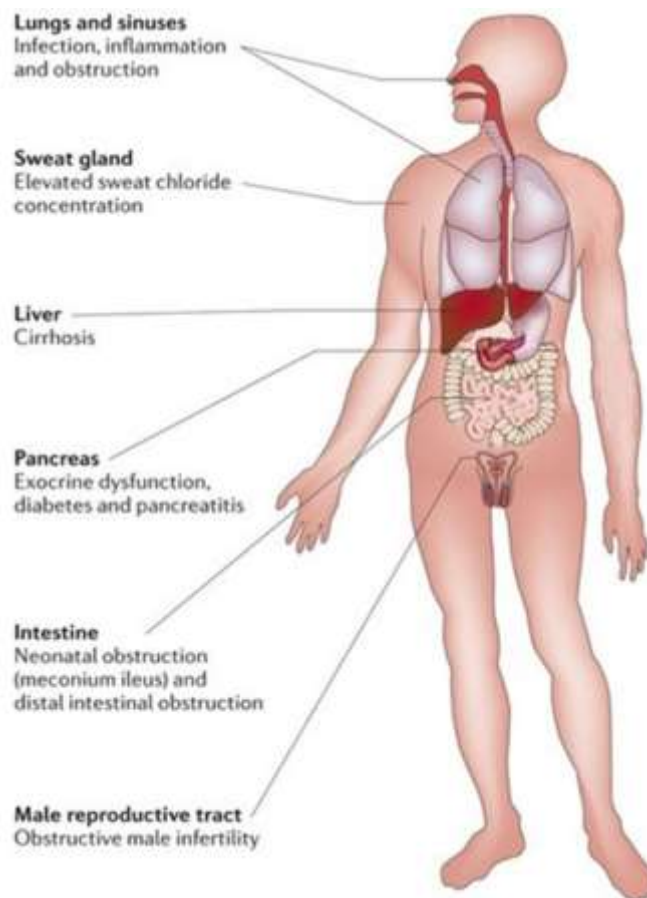


Figure 1.2. The phenotype associated with CF (Picture from Pisi G., *Pharmaceuticals (Basel)*, 2021, 14(9):928).

In the common imaginary, CF is associated with airway misfunction. Besides that, CF is characterized by pancreatic as well as digestive insufficiency, which are often associated with malabsorption, underweight, diabetes, bowel obstruction, and hepatic damage [2]. In addition, CF patients sweat more, with a higher dehydration risk [4], can suffer of eye surface irritation, and male sterility [5,6]. Moreover, new complications, not diagnosable when the life expectancy was lower, are arising. For example, CF patients are highly susceptible to the development of early and aggressive colorectal cancer. 50% of CF patients will develop adenomas by the age of 40, among which 25% are aggressive and advanced adenomas, and some adenocarcinomas [7]. However, CF symptoms among different individuals are significantly variable, which can be a reflection of the multitude of CFTR functions.

To date, the impairment of lung function, and the consequential airway diseases, is the most clinically critical CF phenotype. CF patients produce mucus that can no longer be efficiently cleared off due to the disruption of the extracellular water-salt balance. This led to reduced water content, an accumulation and thickness of the mucus, and a reduced pH. This cascade severely impairs the removal of mucus and bacteria by ciliary beating and the innate defense against bacteria [8] which can find a favorable environment causing chronic infections mainly by *Pseudomonas Aeruginosa*, *Staphylococcus Aureus* and *Haemophilus*, and chronic inflammation, as resumed in Figure 1.3. In the end, the airway obstructions led to an impairment of the lung capacity and to respiratory failure that may require life-saving interventions such as lung transplants (6.1% of living patients) [9,1].

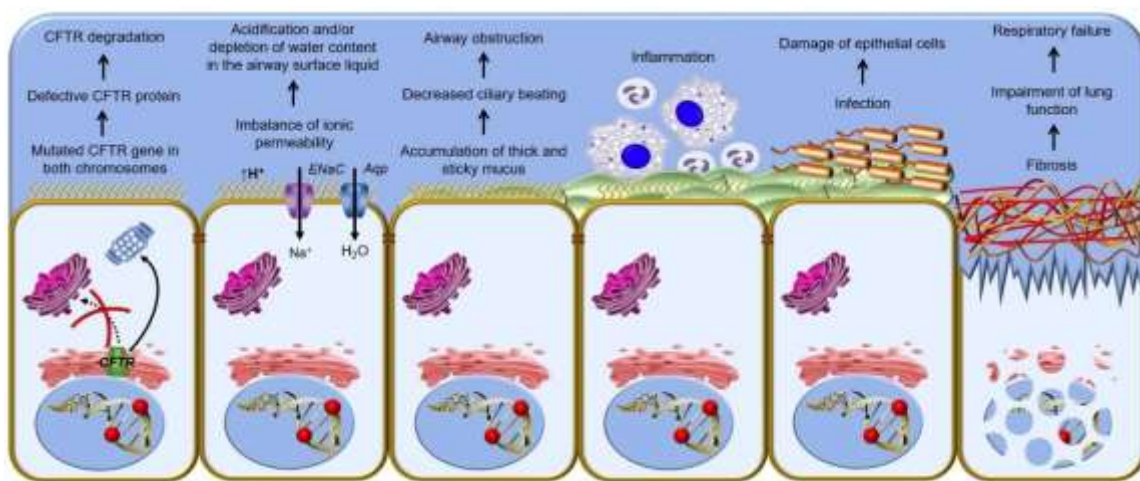


Figure 1.3. Overview of the cascade taking place in the CF patient lung epithelial cells (Picture from Lopes-Pacheco M., *Front. Pharmacol.*, 2016, 7, 275).

## 1.2 CFTR: maturation, protein structure, and function

The gene transcribing the CFTR protein was identified in 1989 [10], marking a milestone in the search for a CF cure.

Over the years, also the protein maturation process was elucidated. Only 20-40% of CFTR reach the plasma membrane in its mature conformation after different steps: post-transcriptional splicing, protein translation, folding at the Endoplasmic Reticulum (ER), glycosylation in the Golgi compartment, trafficking to the apical membrane, endosomal recycling, and retrieval [11,12]. The process is also regulated by multiple quality system control. In particular, the ubiquitin-proteasome, the proteostasis, and the lysosomes are respectively involved misfolded CFTR degradation, to control the protein maturation pathway, and to eliminate non-native protein escaping degradation [11].

The correctly folded protein can reach the plasma surface and carry out its function.

CFTR is a member of the ABC transporter superfamily, it is an active ion channel that requires ATP to allow the ion movement through the cell upon being activated by phosphorylation [13]. CFTR has a similar domain composition as the other ABC transporters. It is a protein of 1480 amino acids composed of two nucleotide-binding domains (NBD1 and NBD2) and two Membrane Spanning Domains (MSD1 and MSD2) resembling a dimer but, unlike other ABC transporters, it possesses a unique long regulatory domain (R) connecting the two parts of the protein (Figure 1.4).

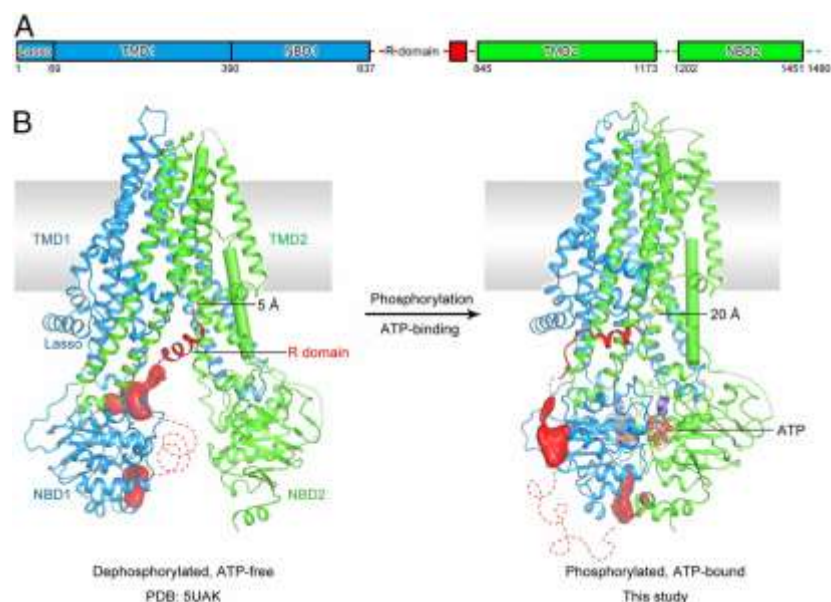




Figure 1.4. Human wild-type CFTR protein structure (*PDB ID code 5UAK*) with the R domain in its dephosphorylate or phosphorylate state. The MSDs domains are here reported as TMDs. Dashed lines indicate regions not resolved in the structure. In red the EM densities, corresponding to unstructured regions within the R domain (*Picture from Zhang Z., PNAS, 2018, 115, 12757-12762*).

The NBDs domains control the gating of the CFTR channel participating in the ATP binding and in its hydrolysis, even if it occurs only at one of the two ATP-binding sites, while the other one is defective in ATP hydrolysis and stably binds a nucleotide [14].

Each MSD domain consists of six membrane-spanning  $\alpha$ -helices called TransMembrane Domains (TMD1 to TMD12) which form the channel pore for the chloride ions transport. The positive charge of arginines and lysines distributed on the TMDs attracts anions into the pore [15].

TMDs are connected to the cytoplasmic ends by long intracellular loops (ICL1-4) predicted to interact and transduce information between TMDs and NBDs and to the extracellular ends by short extracellular loops (ECL1-4) [14,16]. Finally, the R domain which connects the two parts of the protein is disordered and highly charged, displaying multiple PKA phosphorylation sites. Its phosphorylation and dephosphorylation control the opening and closing of the channel [17].

In particular, a single phosphorylated residue is not enough for the activation, in fact, the activity increase with the increasing of the phosphorylated state of the R domain.

CFTR is the only known member of the ABC superfamily not acting as a transporter but as an anion channel. Chloride and bicarbonate ions are transported, but other monovalent anions such as fluoride, iodide, bromide, thiocyanate, formate, and nitrate, can passively permeate. The size of the anions determines their relative permeability [18].

As stated above, CFTR acts as an active ion channel. There are at least two critical steps in the CFTR activation leading to the opening of the pore. The first is the multiple phosphorylations of the R domain [19] which occurs in at least nine sites. When the phosphorylated R domain is displaced from its inhibitor position between the NBDs, it promotes the second step: the binding of the ATP to the NBDs domain, which can now connect each other, together with the ICLs, facilitating the dimerization [6].

Moreover, data suggest that the phosphorylated R domain can also interact with ICLs after NBDs dimerization [20,21].

CFTR activity isn't characterized by a simple open/close channel, but it is defined by different steps (Figure 1.5) and three states called: *open*, *closed*, and *open ready*, where the gate is not opened yet but could rapidly change into the open state [14].

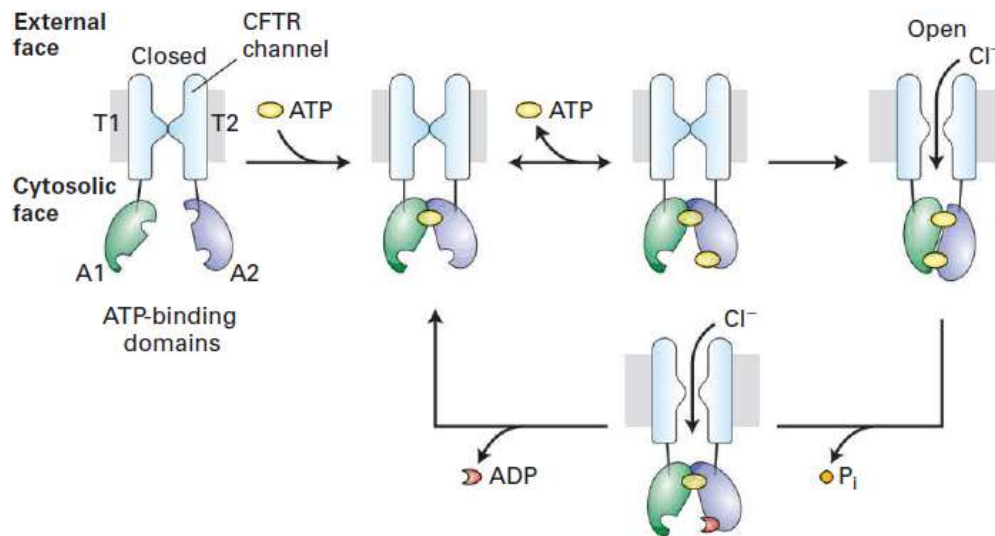


Figure 1.5. Gating mechanism of CFTR (Picture from <https://biologyease.com/cystic-fibrosis/>).

### 1.3 CFTR mutations

Mutations in the CFTR gene cause CF. To date, more than 2100 variants have been described, mostly presumed to cause the disease and most of them affect the NDB1 domain. All the mutations are predicted to impair the CFTR function, but the mechanism by which they cause the disease can be very different. Up to now, seven different classes of mutation affecting the CFTR protein expression or function have been described (Figure 1.6) [6]

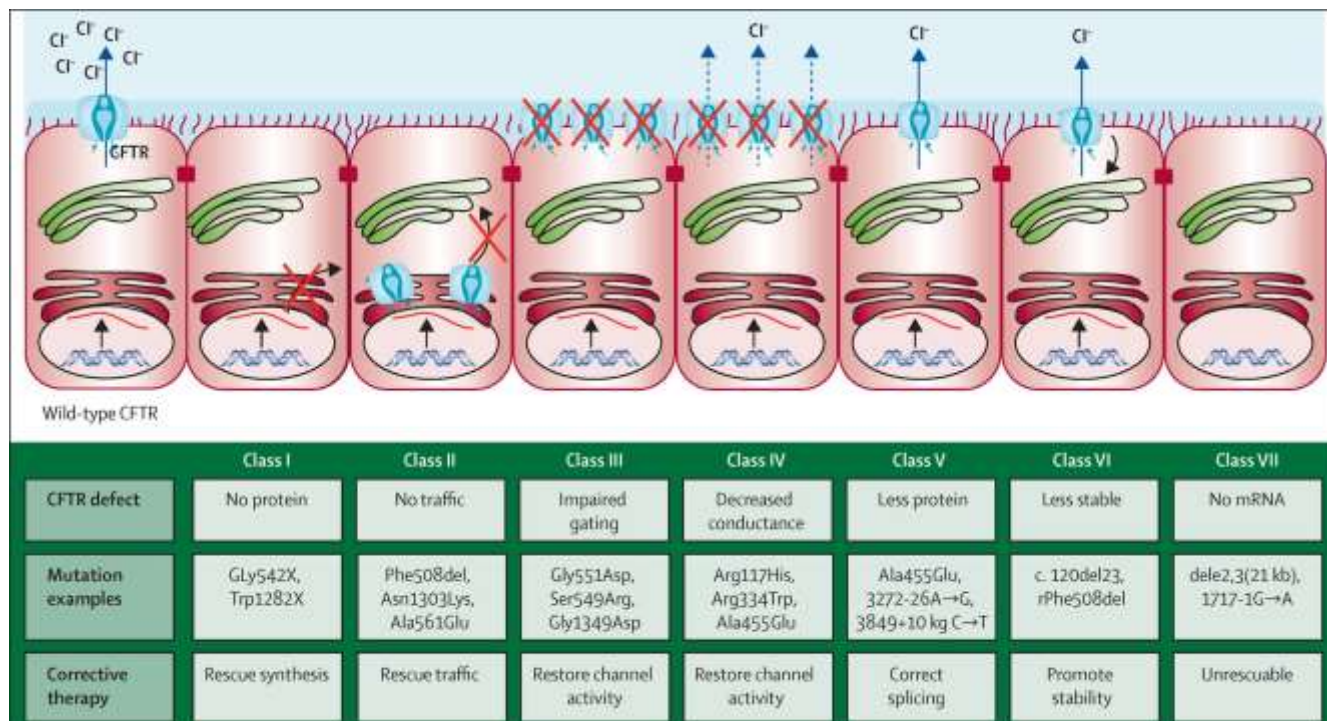


Figure 1.6. The seven classes of CFTR mutations (Picture from De Boeck K., *Acta Paediatrica*, 2020, 109, 893-899).

As shown by the figure above, every mutation class presents a different CFTR impairment.

Class I mutations result in the absence of the protein. The insertion of a premature terminal signal caused by an abnormal splice site or insertion/deletion led to a truncated mRNA and consequent lack of protein synthesis [22].

Class II mutations result in an increase in protein turnover. A misfolded protein is generated, retained in the ER, and subsequently degraded in the proteasome before arriving at the cell surface. Nevertheless, a small CFTR amount can escape and arrive at the plasma membrane, however, the protein is not fully functional and stable. Mutations at the ICLs:NBDs interface are particularly abundant in this class. 70% of CFTR mutations are represented by a mutation occurring in this class: the deletion of the phenylalanine 508 (F508del) [23]

Class III mutations result in opening gate defects. CFTR affected by these mutations arrives at the cell surface, but it is unable to open the gate for ion transport. The G551D is the most common mutation of this class, and it is the third most common CFTR mutation (about 3-4%), affecting ATP binding and hydrolyzation [24].

Class IV mutations result in reduced conduction and decreased flow of ions, occurring in residues that contribute to anion selectivity [6]. Similar to class III, the mutated protein arrives at the plasma membrane, but it is defective in the ion conductance, decreasing the number of ions able to pass through the channel.

Class V and class VI mutations both result in the reduction of the CFTR protein amount at the cell surface. They increase the protein turnover due to partial mRNA incorrect alternative splicing and due to less stability when incorporated into the membrane, respectively [6,25].

Class VII mutations result in the absence of the full length of the mature mRNA [26].

Even if this classification is helpful, several CFTR mutations show common features to different classes of mutations. For example, F508del (Class II) can also be included in Class III and Class VI mutations.

#### **1.4 Cystic Fibrosis therapies**

Heterozygote CFTR people, carrying a mutation in only one CFTR allele, also called carriers, show a less severe disease in comparison to patients carrying both mutated alleles, but they are not risk-free. About 70% of them are heterozygotes for the deletion of the F508, and a recent study showed that carriers have a higher risk of developing various CF-related problematic in different organ systems, in comparison to the healthy population [27]. USA carriers have a significantly high probability to suffer of chronic pancreatitis, bronchiectasis, infections, and male infertility. They also risk suffering of type I/II diabetes, failure to thrive, cholelithiasis, constipation, short stature, scoliosis, and jaundice. Despite these findings, the individual-level risk remained low for most conditions. However, another study found that CFTR carriers have been hospitalized for respiratory infections more than healthy people and, in turn, they are more likely to be prescribed antimicrobials to treat a respiratory infection [28]. However, it is essential to note that homozygous CFTR people with a loss of CFTR function have nearly no level of functional CFTR protein, and they show a severe phenotype of the disease.

Current CF therapies are aimed at preventing a severe clinical outcome, but symptomatic treatment is needed. Aggressive antibiotic strategies are used to prevent and treat lung infections as well as anti-inflammatory and mucus thinning drugs, oxygen therapy, bronchodilators, osmotic agent, and pancreatic enzymes, until getting to bowel surgery and ultimately lung transplant [29].

These therapies have significantly enhanced the mean survival age of patients [30] but the burden of CF care remains very high, requiring the development of alternative strategies. In particular, class II and class III CFTR mutations were found to be a possible target for small molecules called CFTR modulators [31].

### **1.5 CFTR modulators**

Up to now, modulators are a group of five drug types that modulate the CFTR function. They are read-through agents, stabilizers, amplifiers, potentiators, and correctors which: promote the ribosomal read-through of nonsense mutation, stabilize the protein at the plasma membrane, stabilize/increase the mRNA of the mutated CFTR, restore the gating activity, and restore the trafficking, respectively [29]. Heavy financial investments have led to the pre-clinical evaluation of thousands of these synthetic drugs. Over the years, three correctors (lumacaftor – VX809, tezacaftor – VX661, and elexacaftor – VX445) and one potentiator (ivacaftor – VX770) were approved, and some other compounds are being clinically evaluated. Approved modulators are currently used for administration to CF patients in four different combinations: Kalydeco (VX770), Orkambi (VX809/VX770), Symdeko (VX661/VX770), and Trikafta (VX445/VX661/VX770).

#### **1.5.1 Read-Through agents**

Class I mutations need compounds able to promote the ribosomal read-through allowing the synthesis of the full-length CFTR protein. These compounds can inhibit the ribosomal proof-reading allowing the translation to continue to the end of the gene, by reducing the codon-anticodon pairing adhesion [29]. Unfortunately, these agents have the ability to a read-through stop codon in different genes, thus generating toxic aggregates [32]. Up to now, molecules belonging to this modulator class are not approved yet.

#### **1.5.2 Stabilizers**

Class IV mutations need compounds able to correct the instability of the CFTR channel. These compounds are able to stabilize the ion channel at the plasma membrane, decreasing the protein turnover. Up to now, no stabilizer drugs have been approved for therapeutic use [12].

### **1.5.3 Amplifiers**

Molecules belonging to this class of modulators do not target a specific CF class mutation, working independently from the CFTR mutation. They are able to increase the immature CFTR protein expression stabilizing its mRNA. So, amplifiers can act together with potentiators and correctors increasing the CFTR protein expression [33]. Also, for this class of modulators, no approved drugs are available.

### **1.5.4 Potentiators**

Class III or class IV mutations benefit from the use of this class of modulators. CFTR protein with mutations belonging to this class arrives at the cell surface, but its channel functionality needs to be restored. Potentiators increase the ion chloride movement throughout the CFTR channel by helping the gating activity.

Ivacaftor is the only approved drug for this class of modulators. It was first approved by the Food and Drug Administration in 2012 to treat the G511D CFTR mutation, which is the most common class III mutation. Now its use has been extended also to Class III and Class IV mutations, accounting for 8% of CF patients [34]. Also, its therapeutical efficacy against the most common CFTR mutation, the F508del (class II mutation) was evaluated, but no benefits were found when only ivacaftor was used. Its combination with a corrector is required for the trafficking of the mutated CFTR to the plasma membrane, where the ivacaftor can exploit its function [35]. However, its effectiveness on Class II mutation, also when administrated in combination with correctors is still debated but could be beneficial when the potentiator is administered at a low concentration:  $\leq 1 \mu\text{M}$  [36-39].

### **1.5.5 Correctors**

Class II mutations, which need to restore the trafficking of the mutated protein at the cell surface, need “correctors”. F508del is the primary target mutation of this class, being the first common CFTR mutation, with a frequency of about 70%. F508del-CFTR does not correctly fold in the early stage of its maturation, it is so retained in the ER and then degraded [40]. Correctors can be divided into two groups, regarding how they execute their function: indirectly modulating the components of the

cellular quality control machinery or directly interacting with CFTR, promoting the correct protein folding [41]. Compounds that directly interact with CFTR are the most studied. The first approved corrector was Lumacaftor, in 2015 (Figure 1.7).

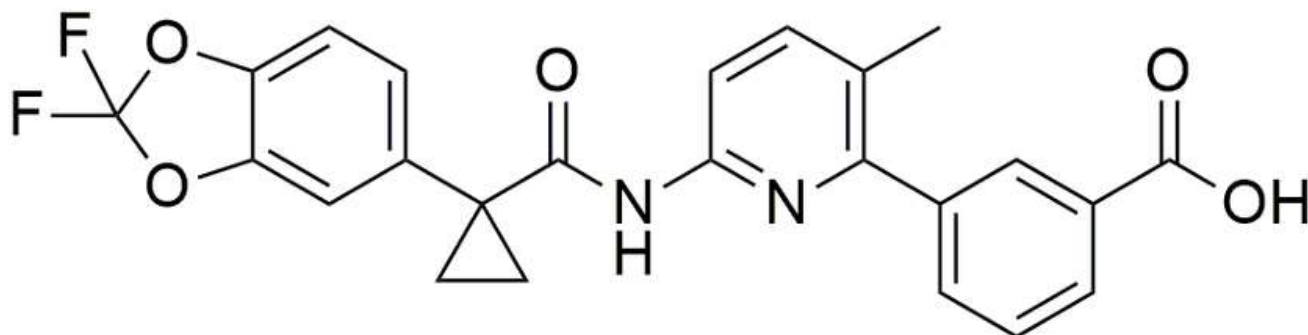


Figure 1.7. Lumacaftor structure.

It is administered in combination with Ivacaftor for the treatment of the F508del-CFTR homozygous. Lumacaftor restores the F508del-CFTR expression in Human Bronchial Epithelial (HBE) cells, acting on the early stage of protein folding [42]. When administered alone in humans it showed a significant decrease in the sweat chloride levels and minimal improvement in Forced Expiratory Volume (FEV) [43]. The F508del-CFTR mutation, besides the processing and trafficking defect, is also defecting in gating and protein stability when localized on the cell surface [34]. So, to overcome the lumacaftor limitation, its combination with the potentiator Ivacaftor was approved, leading to about a 3% of FEV improvement, as well as an overall better outcome of the disease complication.

Another corrector, approved in 2018, is Tezacaftor. Its monotherapy showed slightly higher improvement than Lumacaftor alone and its combination with Ivacaftor showed further improvement in the state of health of the patients.

Elexacaftor is the fourth and last approved CF drug, in 2019. It is administered in combination with Tezacaftor and Ivacaftor, showing a 15.1% increase in the FEV as well as weight increasing, less chloride sweating, and an overall rapid clinical improvement of the patient [44]. The summary of the CF approved drugs and combinations is reported in Table 1.1 below.

Table 1.1. Modulators approved for CF treatment.

<b>Modulators</b>	<b>Commercial Name</b>	<b>Approval Year</b>	<b>Responsive Mutations</b>	<b>Approved Ages</b>
<b>Ivacaftor</b>	Kalydeco (EU/USA)	2012	G551D, S549N, G1244E, G178R, S1251N, G551S, G1349D, S1255P, R117H, E56K, K1060T, P67L, E193K, A1067T, R74W, L206W, G1069R, D110E, R347H, D579G, R1070Q, D1270N, D110H, R352Q, S945L, R1070W, R117C, A455E, S977F, F1074L, F1052V, D115H; 3849+10 kb C>T, 2789+5G>A, 3273-26A>G, 711+3A>G E831X	≥4 months
<b>Lumacaftor- Ivacaftor</b>	Orkambi (EU/USA)	2015	Two copies of F508del	≥2 years
<b>Tezacaftor- Ivacaftor</b>	Symkevi® (EU) Symdeko® (USA)	2018	Two copies of F508del or One copy of F508del in association with E56K, K1060T, P67L, E193K, A1067T, R74W, L206W, D110E, D110H, R347H, D579G, R1070Q, D1270N, R352Q, S945L, R1070W, R117C, A455E, S977F, F1074L, F1052V, D1152H, 3849+10 kb OST, 2789+5G>A, 327326A>G, 711+3A>G	≥6 years
<b>Elexacaftor- Tezacaftor- Ivacaftor</b>	Kaftrio® (EU) Trikafta® (USA)	2020 (EU) 2019 (USA)	At least one copy of F508del	≥12 years

### 1.6 Correctors putative binding pockets

As introduced in the previous chapter, approved and marketed correctors are compounds able to rescue the folding of the mutated CFTR, helping the protein to escape from degradation and be trafficked to the plasma membrane of the cells [45]. Correctors bind to CFTR when the protein is still in



the endoplasmic reticulum but, while lumacaftor and tezacaftor (type I correctors) were found to enhance the protein conformational stability in the early stage of the protein biogenesis making it less sensitive to the ER degradation, elxacaftor (type III corrector) is believed to exert its function in a later phase, when the protein is already stabilized and can escape from the proteosome degradation [46-48].

Despite the efforts in the last years to develop and market new CF drugs, their binding site and molecular mechanism of action have not been yet fully elucidated. Regarding lumacaftor, the first marketed corrector and the most studied one, several studies have suggested its potential binding site and mechanism of action at different CFTR domains. In fact, some papers suggested a putative lumacaftor binding site at the interface between the NBD1:MSD1/2 interface [49] or, more specifically, the NBD1:ICL4 interface [50]. Another study on the F508del-CFTR model computationally showed that lumacaftor could also bind to the NBD1, partially flanking ICL4 and ICL1, stabilizing the interface [51]. Other authors identified the binding site of lumacaftor at the NBD1 domain, with an allosteric effect on the NBD1:ICL4 interfaces [52]. On the other hand, literature data also highlighted a possible lumacaftor binding on MSD1 [53-55] or at the NBD1:ICL1 interface [56]. A recent study using docking experiments and molecular dynamics on the wt-CFTR plus site direct mutagenesis on the F508del-CFTR highlighted a lumacaftor putative binding site at the MSD1, finding a stabilization of the NBD1:ICL4 domain interface through an allosteric mechanism [57].

So, while the identification of a single lumacaftor putative binding pocket is important for the understanding of its mechanism of action, the possibility of multiple binding sites still remains a debated question that needs to be further investigated.

Recently, Cryo-Electron Microscopy (Cryo-EM), for the first time ever, allows the unveiling of the protein structure of the human wild-type CFTR in its de-phosphorylated and phosphorylated conformations [58,59]. Successively, in 2019 the first complex between ivacaftor and the wild type CFTR was published [60]. During the present year, the structure of the wild-type CFTR in combination with lumacaftor was published for the first time, also coupled with site direct mutagenesis experiments on the F508del-CFTR. These data highlighted a lumacaftor binding site in the MSD1, in agreement with the previous literature findings above cited. Lumacaftor was reported as able to induce the stabilization of this domain in the early stages of protein biogenesis, making CFTR less sensitive to ER degradation and

restoring the tertiary structure of the misfolded protein [61]. Furthermore, in late 2022, the structure of F508del-CFTR in combination with elexacaftor and lumacaftor, with or without ivacaftor, was published by the same authors [47]. However, up to now, the structure of the F508del-CFTR complex with only lumacaftor is not yet available.

### **1.7 Drug Repositioning**

Nowadays, a new drug reaches the market about 13-15 years after its discovery, with an average cost of 2-3 billion US dollars [62]. Moreover, since 1950 the number of new drugs approved per billion US dollars spent on research and development is continuously decreasing [62]. This could make the pharmaceutical industry a less desirable choice for investors, and, in turn, it would have disadvantages for the whole pharmaceutical R&D.

On the other hand, drug repositioning (or repurposing) is a strategy aimed at identifying new uses for already approved or investigational drugs, outside their original therapeutic target [63]. It brings many advantages: less risk of failure because the molecule has been already studied in pre-clinical and clinical stages (and maybe marketed), a development time of about 6 years, and less investment needed of about 300 million US dollars. Furthermore, repositioned drugs can reveal new targets and pathways not already exploited [64]. The drug repositioning approach exploits two basic principles: the interdependence between diseases can make a drug works on more than one single disease, or better on more than one target, and the absence of an “absolute” selectivity of the drug towards one single target, makes it active on more than one target and pathway (see unwanted side effects). Drug repositioning studies can follow two strategies: drug-based, where the discovery of a new application for the drug is based on the knowledge related to the drug, and disease-based, where the discovery of a new application for a known drug is deriving from the study of the disease [65].

Drug repositioning is a great opportunity for rare diseases. In fact, there are more than 7 thousand rare diseases, but only 5% of them have an approved drug. Computational techniques can be used to find repositioned drugs, offering a relatively quick method to identify and test them, and for some rare diseases, it can be the more reasonable approach.

For what concern the application of drug repositioning to cystic fibrosis, it has been poorly taken into consideration. Studies highlighted the possibility of using this methodology to pulmonary deliver drugs,

for enhancing mucociliary clearance, or antibiotics, both by inhalation [66,67]. However, its use to explore new possible effective drugs against the mutated CFTR has not been fully exploited yet.

In the first part of the thesis, the results of the application of this strategy against the F508del-CFTR to find a new application for already approved drugs, using an in-house manually curated ligand database ready to use for computational experiments and *ad hoc* developed computational pipelines, will be presented, and discussed.

## **2 Results**

My research activity on F508del-CFTR started from a library of pockets already created in the research group in which I spent my PhD. The library has been derived from molecular dynamic simulations on F508del-CFTR and collects several putative druggable pockets, among which the one putative of lumacaftor. The library represented a useful tool for the subsequent drug repurposing studies, aiming at identifying not only interesting druggable pockets on the mutated protein but also putative drugs able to rescue the misfolded protein. Initially, the repositioning involved the AIFA database. All drugs and nutraceuticals included in the AIFA database were thus sketched as 3D structures and constituted the first core of a manually curated database that was then developed and enlarged during the PhD.

### **2.1 In-house manually curated ligand dataset**

#### **2.1.1 In-house ligand dataset for computational experiments**

The two pipelines that will be presented in this first part of the thesis are based on an in-house ligand dataset manually curated and gradually expanded during my PhD years. Up to now, the dataset contains approved drugs from the Italian medicine agency (Agenzia Italiana del FARMaco – AIFA - <https://www.aifa.gov.it/en/web/guest/home>) and the DrugBank database [68], plus experimental and investigational drugs from the DrugBank. The molecules were retrieved from the above-cited dataset and filtered excluding the not appropriate ones, such as hormones, inorganic salts, contrast agents, peptides, and proteins (Figure 2.1).

A

B

C

Principio Attivo	Active Ingredient	Type	DrugBank
abacavir	Abacavir	Drug	DB01048
abatacept	Abatacept	Protein	DB01281
abciximab	Abciximab	mAb	DB00054
abiraterone	Abiraterone	Drug	DB05812
acamprosato	Acamprosate	Drug	DB00659
acarbosio	Acarbose	Drug	DB00284
acebutololo	Acebutolol	Drug	DB01193
aceclofenac	Aceclofenac	Drug	DB06736
acenocumarolo	Acenocoumarol	Drug	DB01418
acetazolamide	Acetazolamide	Drug	DB00819

Figure 2.1. A) AIFA and B) Drugbank web site from which is possible to get drug information. C) Part of the manually curated ligand library before the discharge of not appropriate molecules.

The molecules were collected in the Simplified Molecular Input Line Entry System (SMILES) format, then converted in their 3D model and further charged at physiological pH 7.4 to obtain the right protonation state (Figure 2.2) using the Ligprep and Epik software by Schrodinger [69,70].

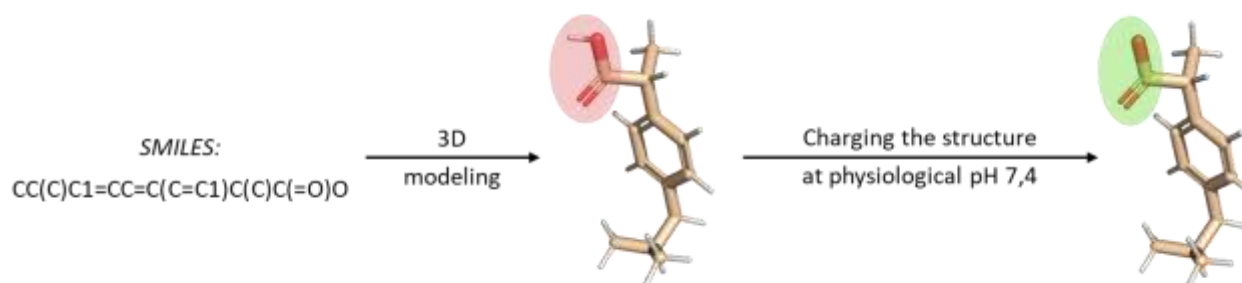


Figure 2.2. From SMILE to a correct 3D charged structure. Ibuprofen is used here as an example.

Up to now, the dataset contains a total of 10288 molecules, subdivided in 2295 approved, 2296 investigational, 5525 experimental, 74 nutraceutical, 30 illicit, and 68 withdrawn molecules. Furthermore, several descriptors for the molecular characteristics of all the molecules are available.

This developed database can include all molecules for which a SMILE (or a 2D/3D structure) is available, allowing the creation of a 3D structure (eventually charged at physiological pH) ready to be used for drug repurposing.

## **2.2 Searching for molecules able to rescue the F508del-CFTR**

The following presented results are from the paper: Orro A<sup>§</sup>, Uggeri M<sup>§</sup>, Rusnati M, Urbinati C, Pedemonte N, Pesce E, Moscatelli M, Padoan R, Cichero E, Fossa P, D'Ursi P. (§ Authors equally contributed to this work) *In silico drug repositioning on F508del-CFTR: A proof-of-concept study on the AIFA library. Eur J Med Chem. 2021 Mar 5;213:113186. doi: 10.1016/j.ejmech.2021.113186. Epub 2021 Jan 13. PMID: 33472120.*

### **2.2.1 Computational drug repositioning pipeline to identify molecules able to bind in the putative lumacaftor binding pocket and rescue the F508del-CFTR protein**

Preliminary findings by the team in which I have been carrying out my research activity, using an F508del-CFTR 3D model, identified a putative binding site for lumacaftor (called Druggable Pocket 1 or DP1) located between the NBD1-NBD2 interface, partially flanking ICL4 [71]. On these data, a computational pipeline based on the repositioning of 846 AIFA-approved drugs was developed by me during my PhD first year (Figure 2.3).

The pipeline is divided into two parts. The first part processes an initial apo F508del-CFTR model embedded in the DOPC lipid bilayer in order to find a set of MDs frames and the DP1 pocket of lumacaftor [71]. The second part is the one which was used for the repositioning studies described in this thesis. Figure 2.3 depicts the complete pipeline.

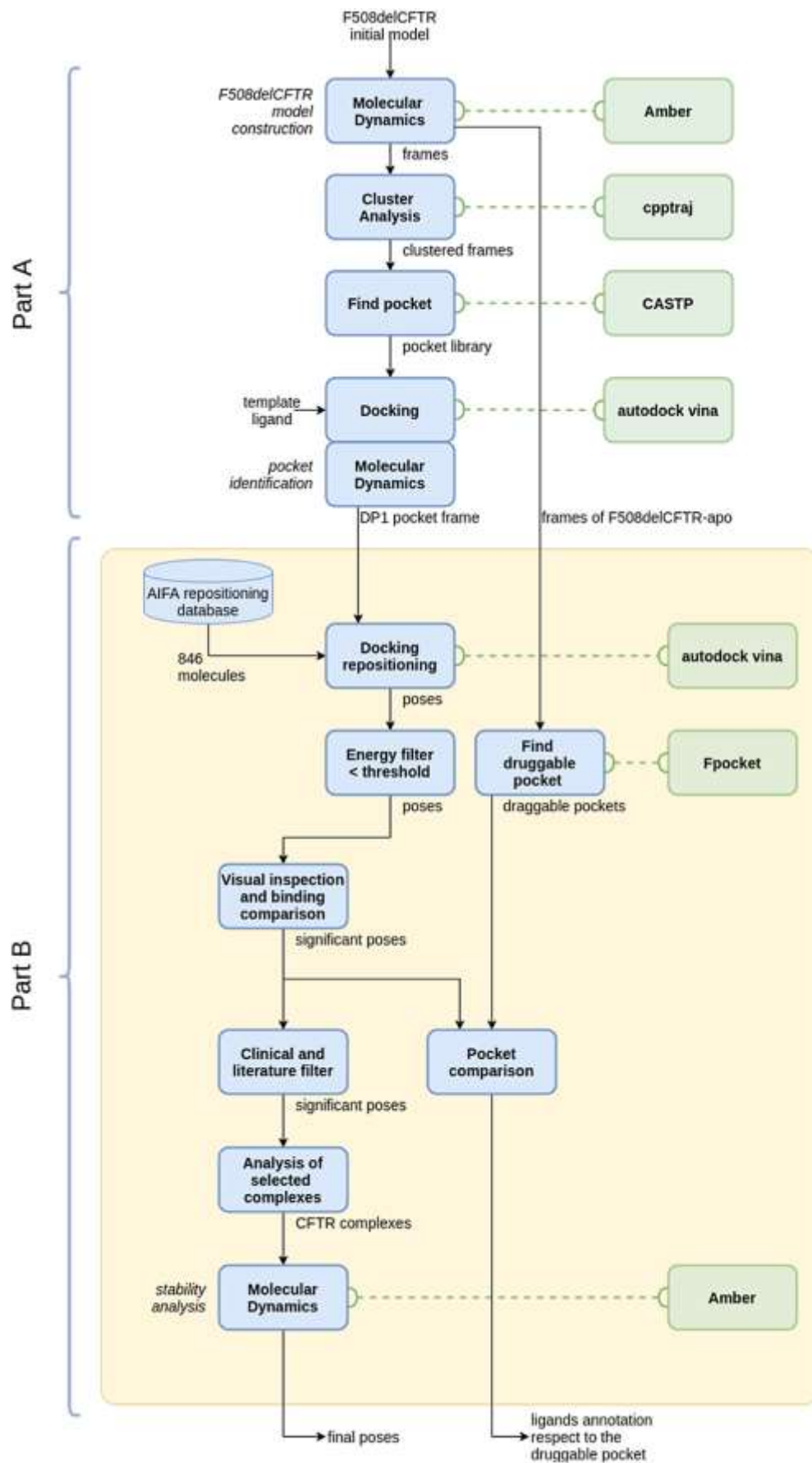


Figure 2.3. Schematic representation of the computational drug repositioning pipeline. Part A of the pipeline is the depiction of studies performed to define the pockets library and DP1 [71]. Part B, is the depiction of the drug repurposing procedure, from the selected pocket DP1 to docking simulations against the AIFA drug database.

Docking repositioning results (Table 2.1) were firstly filtered by energy, taking in consideration only those poses with docking energies like that of lumacaftor previously found (-11.5 Kcal/mol) [71].

Table 2.1. Docking results of the repositioned drugs from AIFA database with a docking score equal or lower that the one found for lumacaftor.

n°	Molecule	Docking Score
1	Rutin	-13,1
2	Telmisartan	-12,4
3	Eltrombopag	-12,3
4	Zafirlukast	-12,1
5	Nilotinib	-12,0
6	Tadalafil	-11,8
7	Imatinib	-11,8
8	Doxorubicin	-11,8
9	Olaparib	-11,7
10	Tolvaptan	-11,6
11	Gliquidone	-11,5

Then a visual inspection (H-bonds) of the 3D poses interaction pattern and, again, a comparison with the lumacaftor binding mode allowed the selection of 11 drugs repositioned on F508del-CFTR (Figure 2.4).



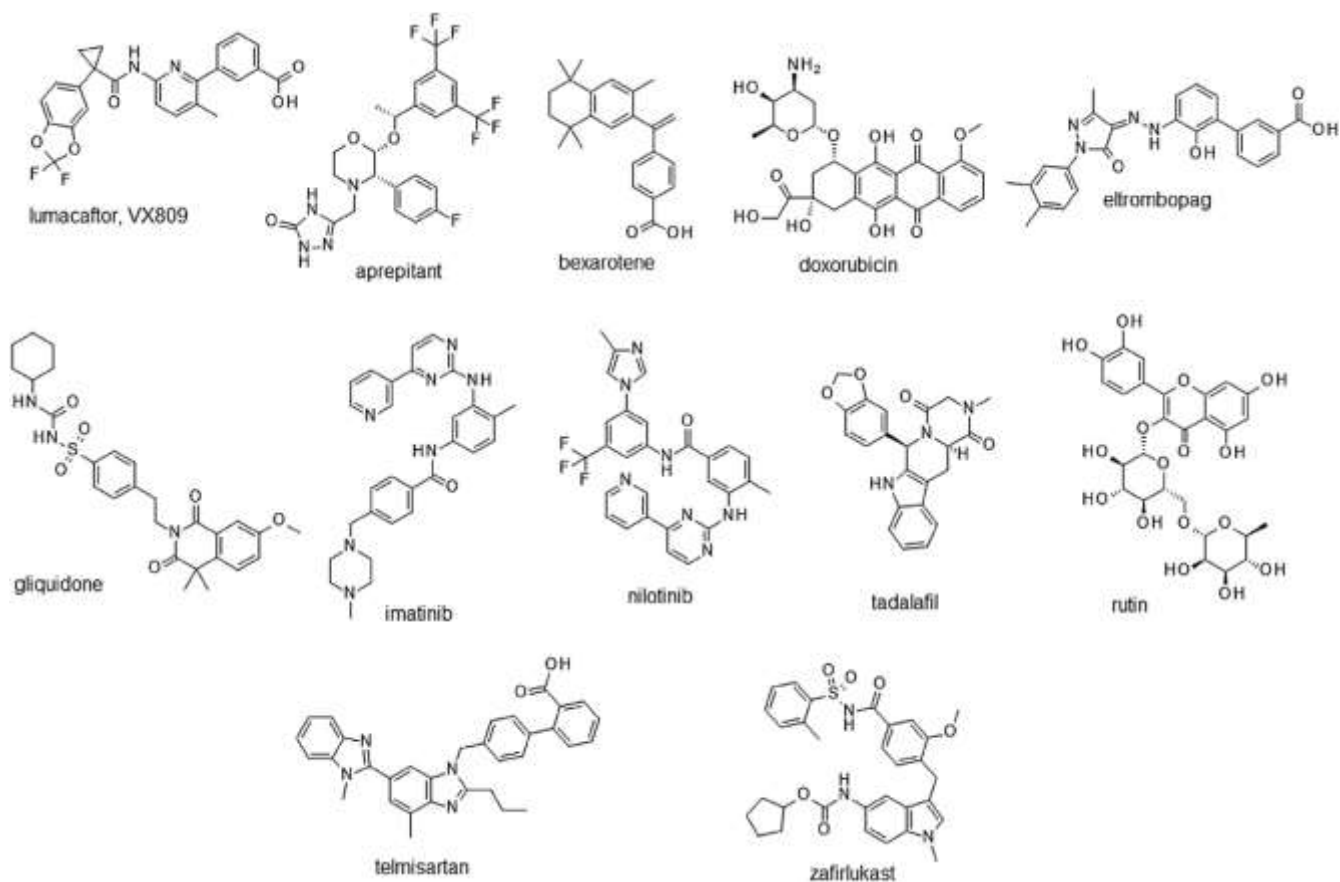


Figure 2.4. Chemical structures of the repositioned drugs. In figure are reported the 2D structure of the 11 drugs repositioned: lumacaftor (template), aprepitant, bexarotene, doxorubicin, eltrombopag, gliquidone, imatinib, nilotinib, rutin tadalafil, telmisartan and zafirlukast.

The 11 repositioned drugs have different chemical scaffolds and include a natural compound, rutin, the glycoside of the flavonol quercetin, and synthetic drugs such as imatinib, nilotinib, doxorubicin, bexarotene, aprepitant, eltrombopag, gliquidone, tadalafil, telmisartan, and zafirlukast. H-bond interactions occur with S492, S495, W496 of NBD1, W1063 of ICL4 and C1344 of NBD2. Lipophilic interactions occur instead with D173 of ICL1, W401, F490, F494 and W496 of NBD1, T1064 of ICL4 and V1345 of NBD2. Finally, a  $\pi$ - $\pi$  staking occurs with W1063 of ICL4 and a salt bridge with R560 of NBD1 (Table 2.2).

Table 2.2. Binding energy ranking of the repositioned drugs. The residues are reported as one-letter code. The interaction pattern is listed. Apex numbers indicate the number of interactions displayed by different ligand atoms with the same amino acid residue, while \* indicates interaction residues common with lumacaftor.

Compound	H-bond interactions		Hydrophobic interactions		$\pi$ - $\pi$ staking interactions		$\pi$ -cation interactions		Binding energy
	Docking	MD	docking	MD	docking	MD	docking	MD	
rutin	//	E267							-13.1
	//	S492*							
	W496 <sup>3*</sup>	W496*							
	//	I497 <sup>2</sup>							
	//	G544							
	R560 <sup>2</sup>	//	I266	//					
	//	D572	F490*	//					
	V1056	//	W496*	W496*	//	F494			
	T1057 <sup>2</sup>	T1057	T1064*	T1064*					
	K1060	//	//	//					
	T1064	T1064 <sup>2</sup>							
	//	V1293							
	D1341	//							
	G1342	//							
	K1351	K1351							
telmisartan			L172 <sup>3</sup>	//					-12.4
			T465	//					
	//	S169	//	F494 <sup>2</sup>					
	W1063*	W1063 <sup>2*</sup>	W496 <sup>2*</sup>	W496*					
			D572	//					
			L801	//					

			K1060 // T1064*	// W1063 //				
<b>eltrombopag</b>	Q1071 K1351		E201 F494 <sup>2</sup> W496 <sup>2*</sup> L804 <sup>2</sup> K1060 V1293 F1294 K1351		F494 W496 <sup>2</sup>		K1060 K1351	-12,3
<b>zafirlukast</b>	W1063* T1064 K1351		L172 <sup>2</sup> D173* I177 W496* L804 K1060 <sup>2</sup> W1063 F1068		F494		K1351	-12.1
<b>nilotinib</b>	E 201 K1060 Q1071 K1351		E201 W496 <sup>2*</sup> T1064* F1068 <sup>2</sup> V1293 F1294 K1351		W496		K1351	-12.0

<b>doxorubicin</b>	S492* R560 K1060 V1293 K1351		W496 <sup>2*</sup> L804 W1063 <sup>2</sup> T1064* F1068						-11.8
<b>imatinib</b>	S492* K1351		T465 L468 W496* D572 W1063 T1064*		W1063 <sup>2*</sup>		K1351		-11.8
<b>tadalafil</b>	// // W1063* //	S492* K1060 // K1351	W496* L804 T1057 K1060 W1063 //	W496* // // // // T1064*	W496 W1063*	// //	//	K1351	-11.8
<b>lumacaftor (template)</b>	S492 <sup>2*</sup> S495 W496* W1063* C1344		D173* W401 F490* F494 <sup>3*</sup> W496* T1064* V1345		W1063*				-11,6
<b>gliquidone</b>	S492* F494 K1060 K1351		T465 L468 F490* F494 <sup>2*</sup> W496 <sup>2*</sup>		F494 W496 <sup>2</sup>				-11.5

			D572					
<b>aprepitant</b>	K1060 T1064		F490* F494 <sup>2</sup> * W496* W1063 T1064* F1068 V1293 K1351		F494		K1351	-11.3
<b>bexarotene</b>	K1351 <sup>2</sup>		F494 <sup>5</sup> * W496* T1064* V1293 F1294 K1351				K1351	-11.1

Imatinib, rutin, doxorubicin, tadalafil, telmisartan, zafirlukast, and gliquidone show a common H-bonding pattern with lumacaftor (one or more interactions with S492, W496, and W1063). Also, all the repositioned drugs show two or more lipophilic interactions involving common residues with lumacaftor.  $\pi$ - $\pi$  stacking with W1063 has been observed only for imatinib, nilotinib, eltrombopag, tadalafil, telmisartan, aprepitant and gliquidone. Imatinib, nilotinib, eltrombopag, tadalafil, telmisartan, aprepitant, zafirlukast and bexarotene show also a p-cation interaction involving K1351. None of them is able to display the salt bridge with R560 shown by the lumacaftor. The comparison of the binding modes of the repositioned drugs with lumacaftor indicated that, even if sharing the above reported common interactions with the template, they display different abilities in occupying the large DP1, and interacting with the surrounding amino acidic residues from NBD1, NBD2, and ICL4. Accordingly, the selected drugs are almost unable to fully overlap lumacaftor, except for telmisartan, zafirlukast, and

tadalafil. Most of them instead limit their overlapping on lumacaftor to that area defined by the aromatic ring bringing the carboxylic group and the adjacent carbons of the pyridinic ring. In addition, lipophilic interactions seem to play a predominant role over polar interactions in anchoring the drugs to residues in NBD1, NBD2, and ICL4.

### 2.2.2. Identification of transient druggable DP1 sub-pockets

In order to better ascertain the probability of these residues being located into the best druggable pocket along the Molecular Dynamics (MDs), Fpocket software has been applied [72] on the F508del CFTR apo form MDs to monitor the residues located in a range of 4 Å in the lumacaftor complex. As shown in Figure 2.5, Fpocket often assigns to these amino acids a very high probability to be part of a druggable pocket (druggability score = 70-80%).

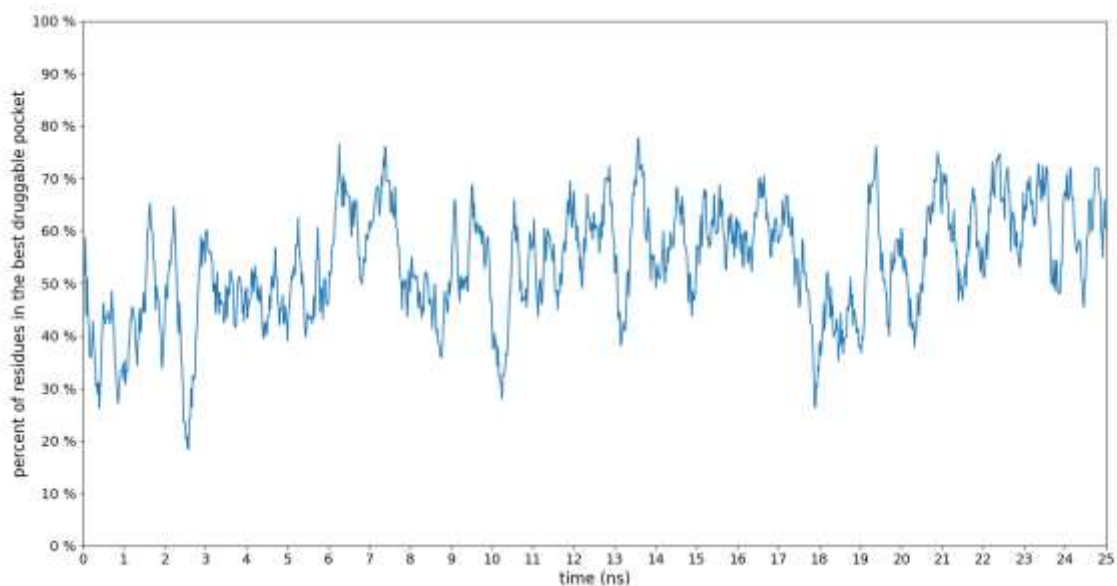


Figure 2.5. Fpocket analysis on F508del CFTR apo form MDs to monitor residues surrounding lumacaftor. Druggability score (%) of the residues close to lumacaftor along the MDs.

This probability has a periodic shape, proving that these residues during the MDs periodically can interact with a ligand in an efficient manner. In conclusion, contextualizing the docking results inside

DP1, three typologies of drugs can be defined depending on the location of residues involved in the binding:

(1) the most populated group corresponding to those drugs able to bridge NBD1 and NBD2, binding also to residues of the ICL4 domain. These are lumacaftor itself, doxorubicin, nilotinib, aprepitant, bexarotene, eltrombopag, zafirlukast and rutin.

(2) drugs that mostly occupy DP1 taking close interactions with residues in NBD1 and ICL4. These correspond to imatinib, tadalafil, and telmisartan.

(3) gliquidone is the only member of the third group, displaying interactions only with the apical region of NBD1.

Interestingly, superimposing one ligand for each group to lumacaftor, it is possible to appreciate three different space sub-regions around the aromatic acid portion of the template, which are not occupied by lumacaftor itself but could be possibly occupied by other drugs (Figure 2.6), and that deserve further investigation (see 2.3).

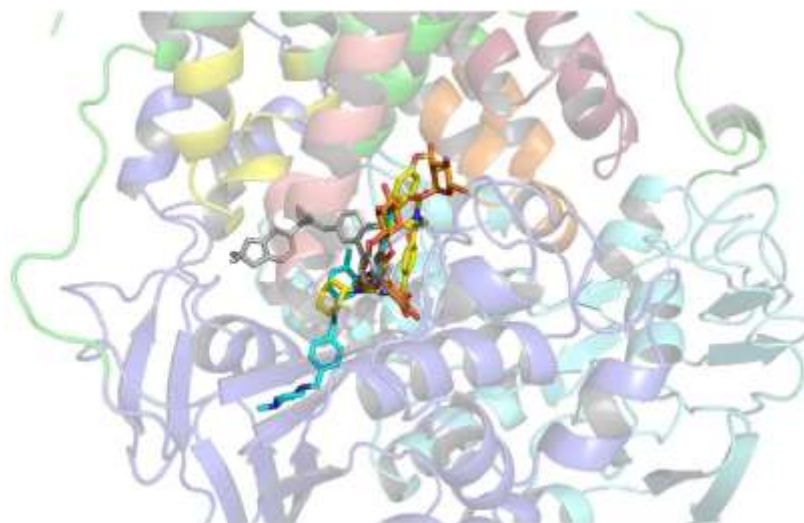


Figure 2.6. Superimposition of imatinib, gliquidone and rutin on lumacaftor. The four drugs are shown as colored sticks: lumacaftor in grey, gliquidone in yellow, rutin in orange, and imatinib in cyan; while F508del-CFTR in cartoon: NBD1 in slate blue, NBD2 in cyan, ICL1 in yellow, ICL2 in orange, ICL3 in raspberry red, ICL4 in salmon pink, and TMs in green.

Fpocket analysis has been thus performed in the same way as described above, on rutin, imatinib, and gliquidone (representatives for the three groups discussed above). The corresponding plots are here reported in Figures 2.7, 2.8, and 2.9. From the plots is evident that in the apo form of F508del CFTR, residues surrounding imatinib (Figure 2.8) appear endowed with a high probability/druggability score (70-80%) along all the MDs, comparable to that of lumacaftor (Figure 5). Residues located around rutin (Figure 2.7) and gliquidone (Figure 2.9) show a high average score only in the latest part of the MDs. For all compounds, the plot has a periodic shape, indirectly suggesting that F508del CFTR periodically exposes selected residues for interaction with a ligand. It is possible to infer that in the case of imatinib and lumacaftor, the binding pocket residues have a high possibility to be available for interaction.

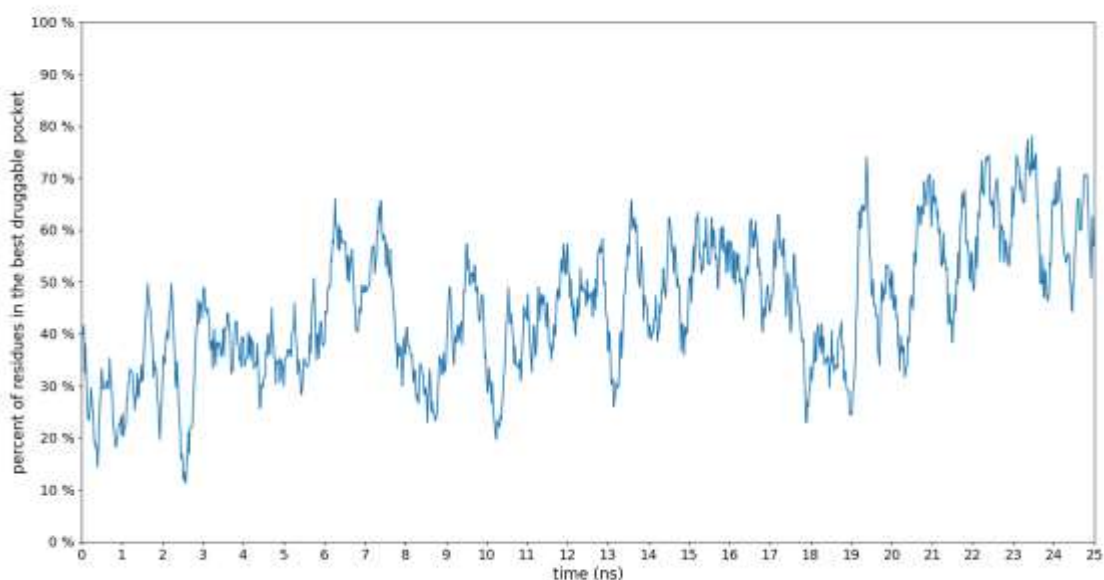


Figure 2.7. Fpocket analysis on F508del CFTR apo form MDs to monitor residues close to rutin. Druggability score (%) of the residues surrounding rutin along the MDs.



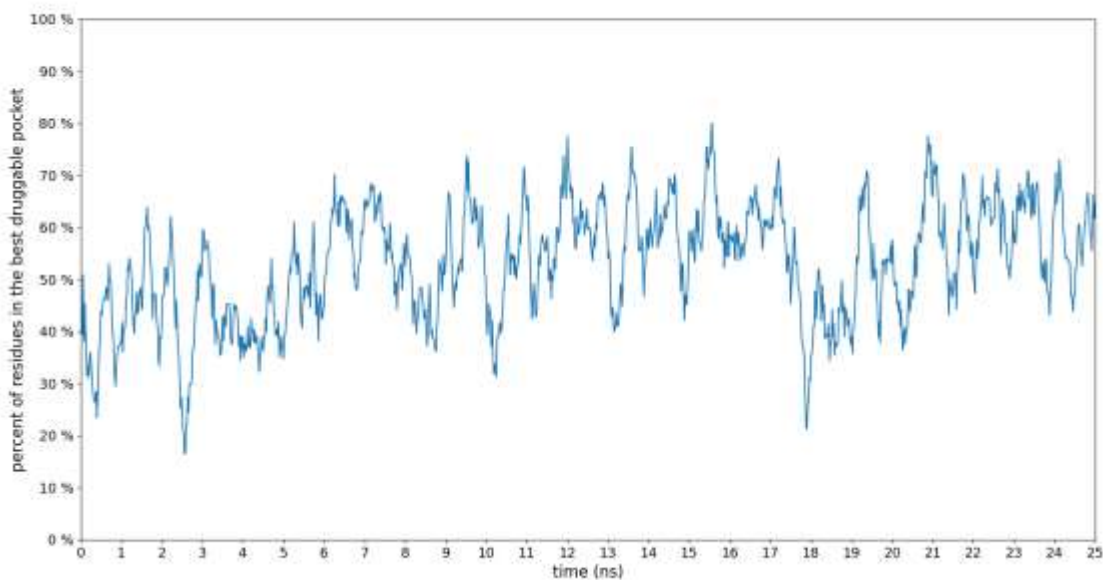


Figure 2.8. Fpocket analysis on F508del CFTR apo form MDs to monitor residue close to imatinib. Druggability score (%) of the residues surrounding imatinib along the MDs.

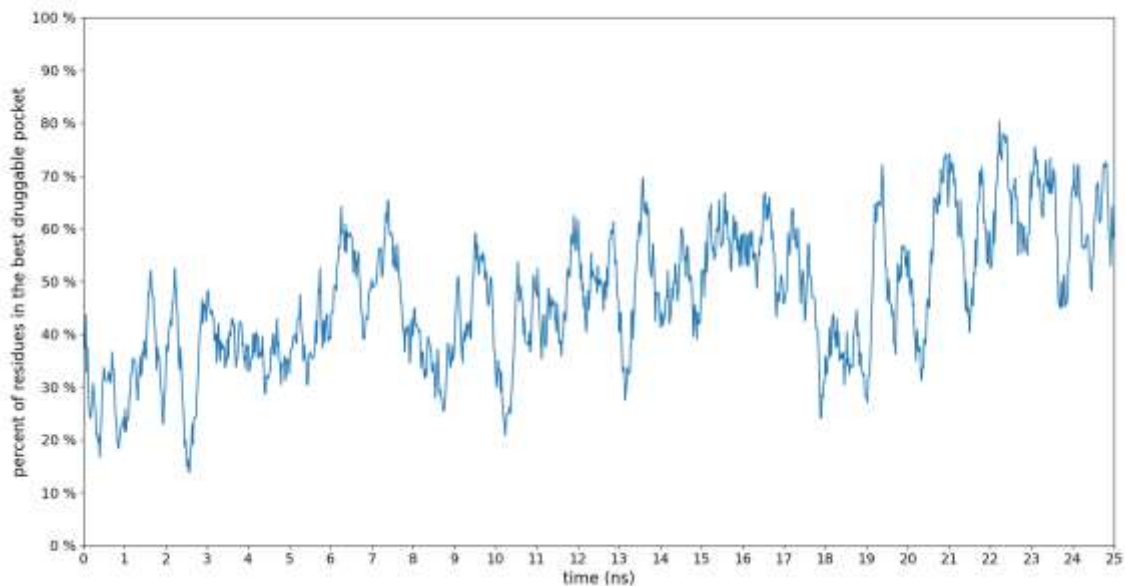


Figure 2.9. Fpocket analysis on F508del CFTR apo form MDs to monitor residues close to gliquidone. Druggability score (%) of the residues surrounding gliquidone along the MDs.

### **2.2.3 Complex stability and molecular interactions of the repositioned drugs**

Then, these repositioned approved drugs obtained from docking experiments have been evaluated by a clinical consultant [73]. The expert found that the most populated group of repositioned drugs was represented by antineoplastic agents: doxorubicin, nilotinib, imatinib, and bexarotene. These are also known as hazardous drugs and have common side effects: nausea, vomiting, diarrhea, alopecia, stomatitis, and, overall, bone marrow depression with a consequent increased risk of Infections. These undesirable contraindications are extremely dangerous in cystic fibrosis patients, which are already at risk of malnutrition and respiratory infections by bacterial, viral, and fungal. Aprepitant is an antiemetic, but, as well as the above-discussed class of drugs, can increase the risk of infections. Zafirlukast, now withdrawn, is a leukotriene receptor antagonist used to alleviate the symptoms of asthma. However, it had been classified as a second-choice drug due to its not yet completely clarified side-effects, among which are breathing difficulties, jaundice, nausea, headache, or excessive tiredness, as well as important psychological problems, which makes unlikely its use on CF patients. Eltrombopag leads to severe hepatotoxicity and its thrombotic or thromboembolic complications make it unsuitable for CF patients. Gliquidone is also not recommended as it is contraindicated in type I diabetes, and CF patients are at risk of developing diabetes. Telmisartan is an angiotensin II receptor antagonist used to treat hypertension in adults and to reduce cardiovascular morbidity in patients with type II diabetes, making it an interesting drug. However, it should not be administered to patients with biliary tract obstructions or liver failure as it is mainly eliminated with bile. Tadalafil is a phosphodiesterase 5 inhibitor that has anti-inflammatory effects. Anti-phosphodiesterase 5 drugs were found to activate the chloride channel in homozygous mice F508del-CFTR, making it an interesting molecule in CF therapy. Finally, rutin is a nutraceutical glycoside, composed of the flavonoid quercetin and rutinose, which is bio-converted in quercetin. Both rutin and quercetin are endowed with a broad spectrum of biological effects among which an anti-inflammatory potential that seems to be the most interesting aspect for CF patients, as inflammation is the initial actor of lung damage at an early age.

Combining the information derived by the binding energy ranking of the repositioned drugs and their clinical assessment, the study continued focusing on telmisartan, tadalafil, and rutin for further computational and experimental evaluation.

Tadalafil, telmisartan, and rutin have been then submitted to three independent MDs with different initial velocities to test the stability of their complexes with F508del-CFTR and to fix their binding pose into the protein [74-76]. Structural analysis of all replicas shows that the initial binding mode of each ligand is kept (RMSD lower than 2 Å from the starting docking pose) except for a single rutin run. Subsequently, for each ligand, for the poses with RMSD lower than 2 Å, binding free energy calculations were performed and the pose with the lowest energy was selected as the correct binding mode (Figure 2.10).

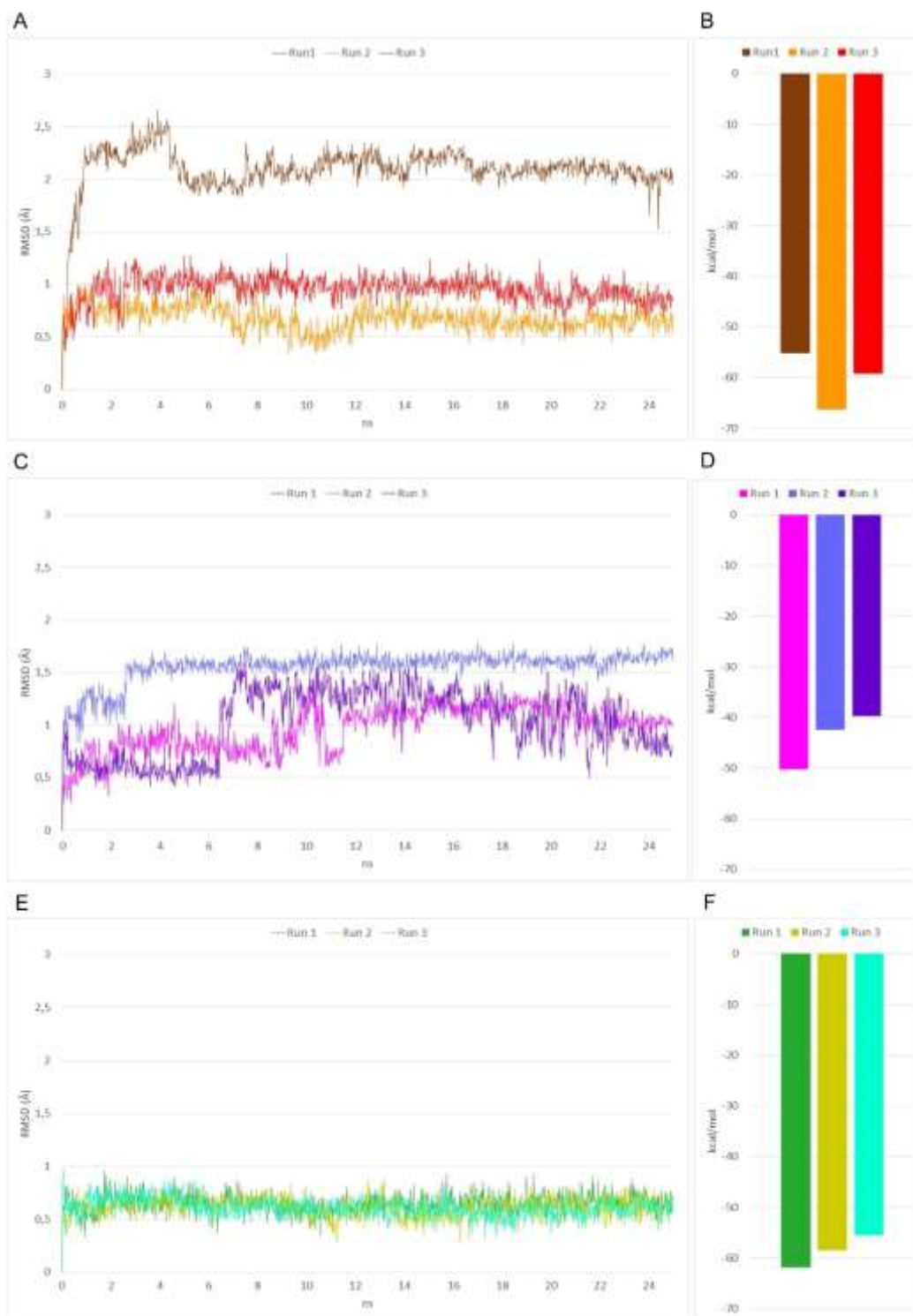


Figure 2.10. RMSD and MMGBSA analyses of the three independent MDs runs. RMSD and MM-GBSA analysis results of ligands during the three independent MDs runs of F508del-CFTR in complex with A-B) rutin C-D) telmisartan E-F) tadalafil.

Related RMSD analysis of the complexes showed good stability of the whole system, with rutin, telmisartan, and tadalafil reaching stability after 5 ns, 12 ns, and 2 ns, respectively. RMSF comparison of the three complexes with the apo F508del-CFTR shows a general lower or equal fluctuation of complexes. In particular, tadalafil, telmisartan, and rutin (Figure 2.11A, 2.12A, and 2.13A) stabilize part of the NBD1 domain, while part of NBD2 is stabilized only by rutin and tadalafil. Moreover, rutin is the only compound able to stabilize the ICL4 domain, while telmisartan produces a higher ICL4 fluctuation when compared to the apo form.

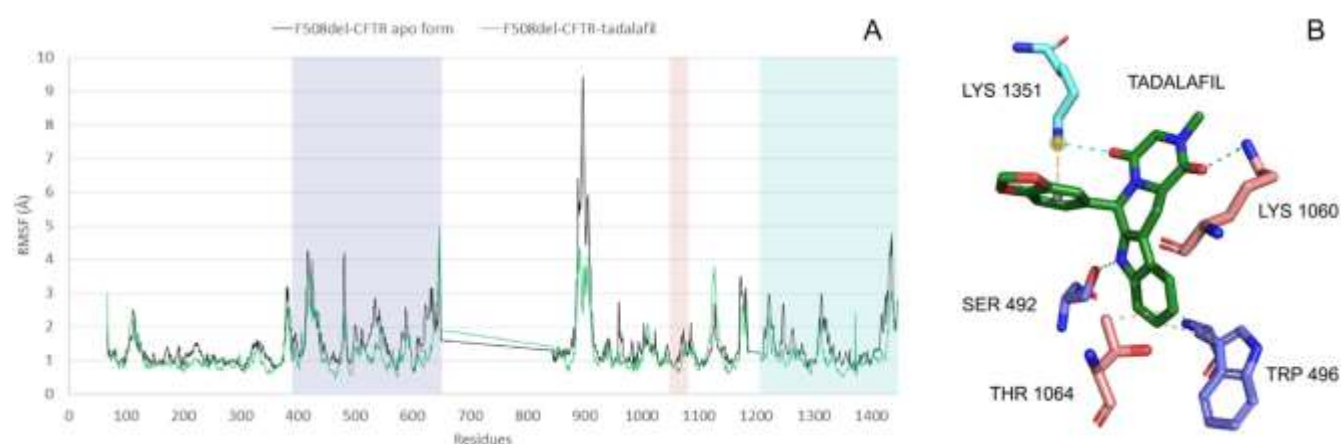


Figure 2.11. RMSF and binding mode interactions for tadalafil complex. A)  $\alpha$  comparison between F508del-CFTR apo form and complexed with tadalafil. F508del-CFTR apo form is colored in black and tadalafil in green. The colored boxes represented the NBD1, ICL4, and NBD2 colored as slate blue, salmon pink, and cyan, respectively. B) The binding mode of tadalafil with F508del-CFTR after MDs. H-bonds and hydrophobic interactions are colored as cyan and ochre dashed dots, respectively, while pi-cation interactions are shown as yellow spheres. F508del-CFTR residues are shown as sticks and colored as follows: NBD1 slate blue, ICL4 salmon pink, and NBD2 cyan.

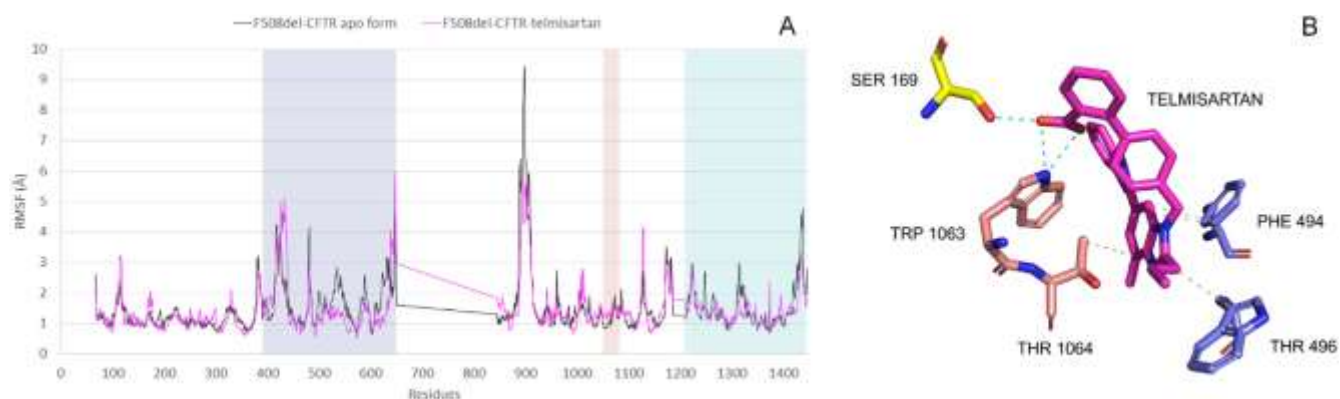


Figure 2.12. RMSF and binding mode interactions for the telmisartan complex. A)  $\alpha$  comparison between F508del-CFTR apo form and complexed with telmisartan. F508del-CFTR apo form is colored in black and telmisartan in magenta. The colored boxes represented the NBD1, ICL4, and NBD2 colored as slate blue, salmon pink, and cyan, respectively. B) The binding mode of telmisartan with F508del-CFTR after MDs. H-bonds and hydrophobic interactions are colored as cyan and ochre dashed dots, respectively. F508del-CFTR residues are shown as sticks and colored as follows: ICL1 yellow, NBD1 slate blue, ICL4 salmon pink

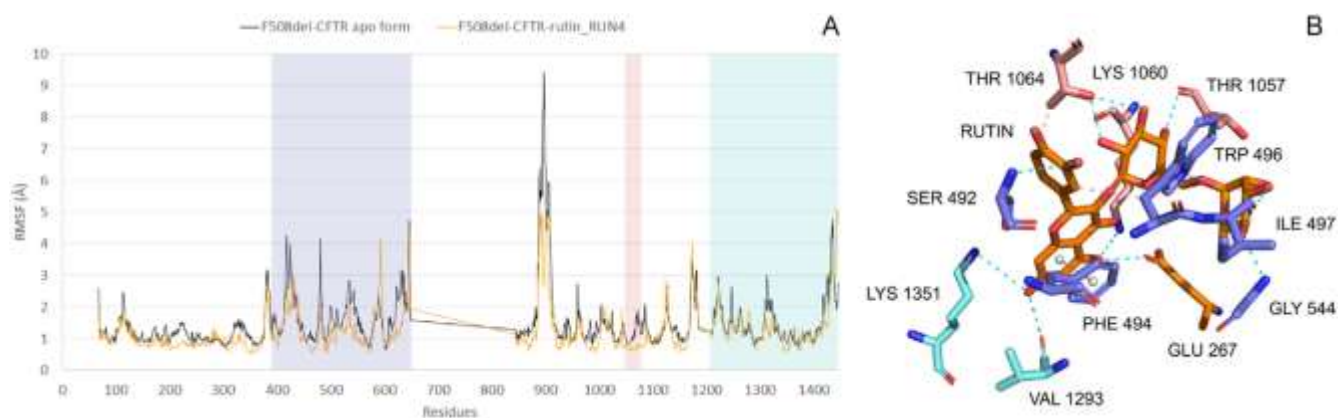


Figure 2.13. RMSF and binding mode interactions for rutin complex. A)  $\alpha$  comparison between F508del-CFTR apo form and complexed with rutin. F508del-CFTR apo form is colored in black, rutin in orange. The colored boxes represented the NBD1, ICL4, and NBD2 colored as slate blue, salmon pink, and cyan, respectively. B) The binding mode of rutin with F508del-CFTR after MDs. H-bonds and hydrophobic interactions are colored as cyan and ochre dashed dots, respectively, while p-cation interactions are shown as white spheres. F508del-CFTR residues are shown as sticks and colored as follows: ICL1 orange, NBD1 slate blue, ICL4 salmon pink, and NBD2 in cyan.

Then, the comparison between rutin and lumacaftor RMSF analyses shows how the two drugs have a comparable stabilizing capability over F508del-CFTR. Rutin efficiently stabilizes NBD1, ICL4, and NBD2, the main regions involved in CFTR rescue. In the case of lumacaftor, a lower stabilizing effect of the drug over F508del-CFTR can be observed limited to the N-terminal part of ICL4 (Figure 2.14).

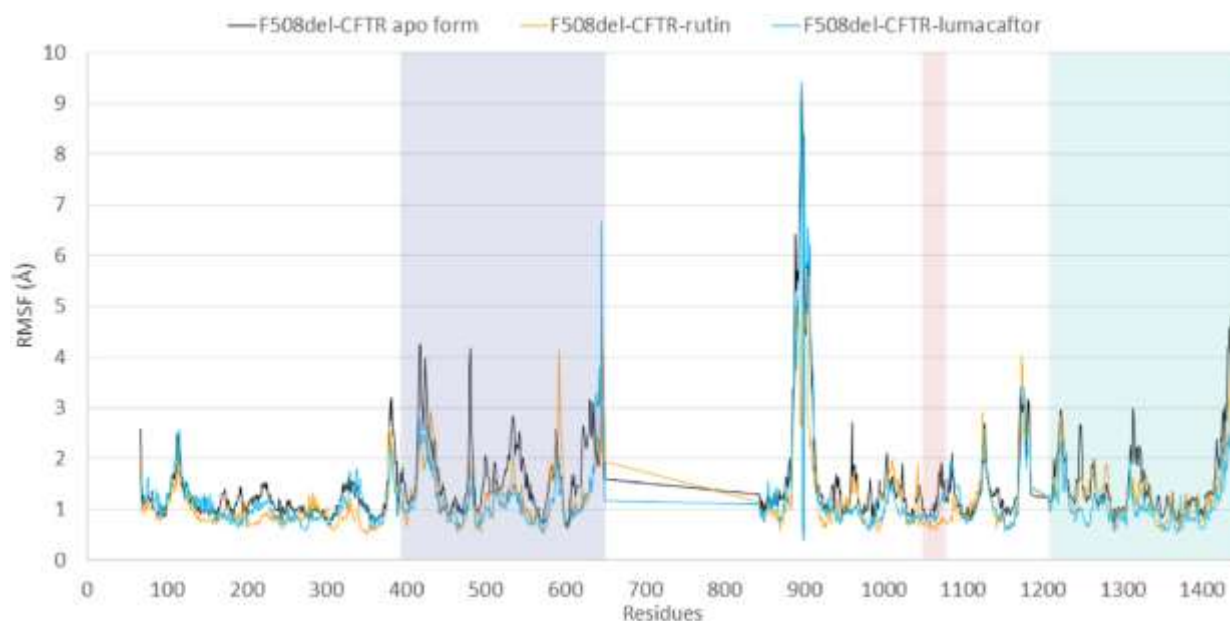


Figure 2.14. RMSF comparison of rutin and lumacaftor complexes. Comparison between F508del-CFTR apo form and complexed with rutin and lumacaftor. F508del-CFTR apo form is colored in black, rutin in orange, and lumacaftor in cyan. The colored boxes represented the NBD1, ICL4, and NBD2 colored as slate blue, salmon pink, and cyan, respectively.

Interactions after MDs for the studied compounds are detailed in the Table 2.1 shown above. Tadalafil slightly moves towards the NBD2 domain and increases the number of its H-bond with the protein. It loses the interaction with W1063 observed in the lumacaftor complex but engages interactions with residues S492, K1060, and K1351 (Figure 2.11B). The number of lipophilic interactions is decreased in comparison with the docking pose, however, the interaction with W496 is maintained and the interaction with T1064 is gained. Moreover, a  $\pi$ -cationic interaction is displayed with K1351. Telmisartan increases its H-bonds interaction by involving S169 of ICL1, plus W1063 (shared with lumacaftor) (Figure 2.12B), and it durably gains these two interactions after about 12 ns. Lipophilic

interactions decrease, with only those with W496 and T1064 maintained and reinforced by F494. At variance, the  $\pi$ -cation interaction with K1351 of NBD2 is lost. Rutin increases the number of H-bonds, thanks to its sugar moieties (Figure 2.13B). It gains interactions with I497, G544, and D572, while keeping those with W496 (shared with lumacaftor), T1057, and T1064. During MDs rutin slightly moves back to the NBD1 domain, thus losing the interaction with D1341 of NBD2. No additional lipophilic interactions were observed, while one  $\pi$ - $\pi$  stacking interaction is displayed with F494. As written above, the three ligands partially superimpose to lumacaftor in the binding pocket, with tadalafil and telmisartan performing better than rutin, whose sugar moieties are directed to a region of DP1 not occupied by any of the other three molecules.

#### **2.2.4 Surface Plasmon Resonance analysis**

In collaboration with Prof. Marco Rusnati from the University of Brescia, Surface Plasmon Resonance (SPR) experiments were performed to validate the binding capacity of these compounds to the F508del-CFTR. The molecules were injected at increasing concentrations onto the biosensor containing F508del-CFTR [71], providing dose dependent binding curve. Results highlighted that tadalafil and telmisartan had a lower binding affinity than lumacaftor, while rutin showed an almost comparable affinity to the CF drug (Table 2.3) [73]. For this reason, tadalafil and telmisartan were not successively studied, while rutin was taken into consideration for further analysis. However, an *in vivo* interaction between Rutin and CFTR is unlikely, because rutin is bioconverted in the flavonol quercetin by the human gut bacteria [77,78] (Figure 2.15) calling for computational and experimental studies aimed at evaluating if quercetin retains F508del-CFTR-binding capacity.



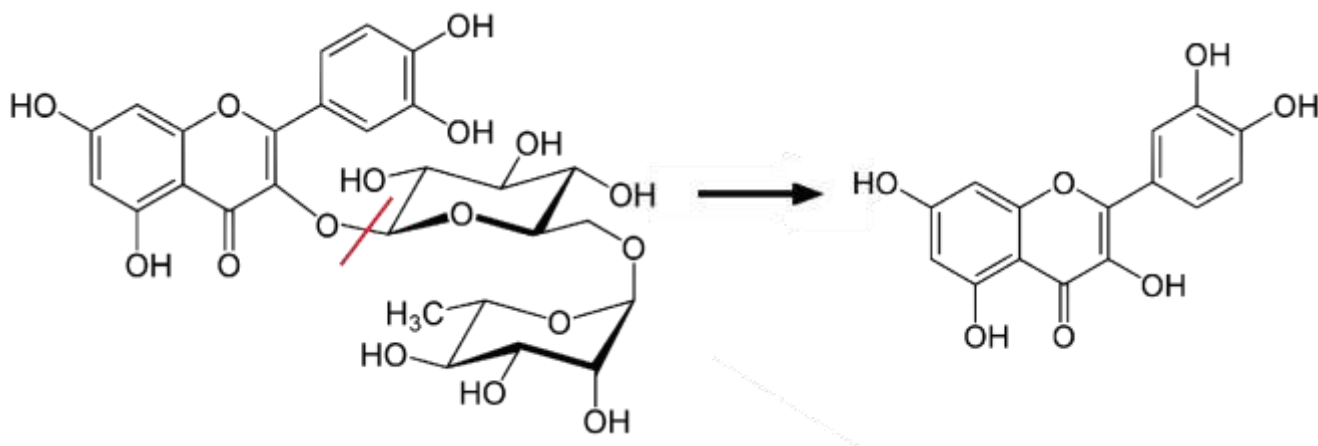


Figure 2.15. 2D chemical structure of rutin and its bioconversion to quercetin.

This is also in light of the fact that quercetin has already been demonstrated to activate CFTR in gut and airway epithelial cells [79,80]. To evaluate if quercetin retains its CFTR-binding capacity docking simulations were performed. Results confirmed that quercetin is able to fit in the DP1 almost occupying the same region already involved in the binding by the aglyconic portion of rutin with a shift of the isoflavonolic scaffold in comparison with rutin. The molecule appears to be superimposed with lumacaftor limited to the aromatic ring substituted with the carboxylic function. Interestingly, SPR analysis demonstrated that quercetin binds F508del-CFTR in a dose-dependent and saturable manner with an affinity that is comparable to that of lumacaftor itself (Table 2.3).

Table 2.3. Binding of the compounds to sensorchip-immobilized F508del-CFTR.  $K_d$  values are reported in  $\mu\text{M}$  + S.E.M. The number of repeated independent calculations are indicated in brackets.

Compound	$K_d$ ( $\mu\text{M}$ )
lumacaftor (template)	$47.8 \pm 18.36$ (5)
telmisartan	$189.3 \pm 48.5$ (3)
tadalafil	$175.0 \pm 50.4$ (4)
rutin	$65.8 \pm 27.3$ (7)
quercetin	$25.6 \pm 10.2$ (3)

The fact that the repositioned drugs were selected on their predicted capacity to dock into the lumacaftor region in DP1 a competition binding assays between the lumacaftor and rutin or quercetin was carried out. As shown in Figure 2.16, at the doses tested, rutin partially inhibits lumacaftor binding to F508del-CFTR while quercetin shows instead an additive effect, suggesting that, due to its small dimension, it succeeds in fitting in the DP1 of F508del-CFTR together with lumacaftor.

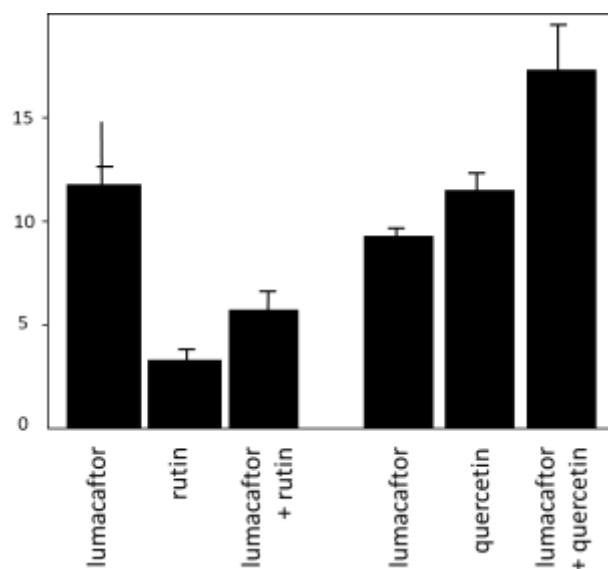


Figure 2.16. SPR competition experiments: lumacaftor was injected into the F508del-CFTR biosensor in the absence or the presence of rutin or quercetin. The values of RU bound at equilibrium reported are the means  $\pm$  S.E.M of 3-10 independent experiments.

### 2.2.5 Discussion

This work was carried out during my first year of PhD and presents a computational drug repositioning approach for the identification of new drugs to treat CF. Here an appropriate drug database ready to be used for computational analysis was built and used together with an optimal computational model of the F508del-CFTR in a lipid environment [73] to study the binding modes of putative F508del-CFTR-rescuing drugs, providing predictions in consensus with the SPR experimental data.

This first work is a “proof-of-concept” example, where computational and biological methods, and their use as a multidisciplinary approach, were applied to reposition the AIFA database on CFTR.

Interestingly, among the repositioned compounds, tadalafil has been already taken into consideration for CF therapy [81]. The most populated group of repositioned drugs identified is the one of the antineoplastic agents, which have been excluded from further investigations due to their severe side effects. However, since these side effects are manifested at doses high enough to exert the antineoplastic effect, the same drugs may likely require lower (and thus safer) doses to exert their CFTR-rescuing activity.

Interesting is also the case of imatinib, which has been already repositioned in the past: developed and approved to target BCR-ABL in chronic myeloid leukemia, some years later was repositioned, still in the oncological area, to target KIT in gastrointestinal stromal tumors [82]. Also worth of note is the observation that bexarotene, nilotinib, and telmisartan have been selected in a drug repositioning program for Alzheimer disease [83] and, again, telmisartan in another repositioning program for trans-sialidase inhibitors of *Trypanosoma cruzi* [84], suggesting that some compounds are more prone than other to emerge as hits in any drug repositioning program, possibly due to their peculiar chemical structure.

During the study, starting from 11 promising repositioned drugs, they were reduced to 3 after a clinical assessment. The binding mode of the 11 repositioned drugs highlights how the DP1 occupied in F508del-CFTR by lumacaftor (here used as a template) is a very large region that can be occupied more stably by very large molecules as antineoplastic agents that, making multiple interactions, display lower complex energy and are consequently top-scored in the energy ranking. Accordingly, the polycyclic natural compound rutin shows the lowest energy complex with F508del-CFTR thanks to its large dimension made up of sugar moieties and a flavonic portion. Rutin is widely reported in the literature as a very interesting molecule, endowed with diverse therapeutic effects, such as anti-cancer, anti-oxidant, adipogenesis suppressant, and neuroprotective. However, *in vivo* rutin is rapidly converted in the flavonol quercetin by human gut bacteria, as above mentioned. The removal of the disaccharide moiety infers the possibility that the beneficial effects of rutin may be actually mediated by quercetin. Also, compared to rutin, quercetin and its derivatives are beneficial because of their more efficient absorption [85]. It should also be noted that animal and human studies revealed that quercetin may alter the bioavailability of certain drugs [86], calling for further evaluation of its use in combination with drugs already administered to CF patients.

Interestingly, however, in the study here presented SPR data showed that quercetin does not interfere with the binding of lumacaftor to F508del-CFTR. Accordingly, quercetin has been already reported to stimulate F508del-CFTR activity, in a dose-dependent manner, in FRT and CFBE41o cells, kept at low temperature to allow the rescue of the trafficking defect of mutant CFTR [80]. Computational studies suggest that quercetin binds F508del-CFTR and that its binding region corresponds to the DP1. Then, in agreement with computational predictions, SPR analysis, demonstrates that quercetin directly binds F508del-CFTR with an affinity that is comparable to that of lumacaftor. More intriguingly, competition experiments indicated that while rutin competes with lumacaftor for the binding to F508del-CFTR, quercetin seems to succeed in interacting with F508del-CFTR along with lumacaftor. A possible interpretation of this behavior is that quercetin binds to the flexible binding cleft spanning from NBD1 to ICL4 and NBD2 in such a way to produce a conformation of DP1 suitable for better interaction of the corrector lumacaftor.

So, starting from the results obtained from rutin and quercetin, a second analysis of the already repositioned drugs inside DP1 was carried out in order to find small molecules able to synergize with lumacaftor in the same binding pocket, for the F508del-CFTR rescue.

### **2.3 Small molecules able to synergize with lumacaftor for the F508del-CFTR rescue**

This further analysis led to the publication of a second paper, from which the following results are presented: *Fossa P\$, Uggeri M\$, Orro A, Urbinati C, Rondina A, Milanesi M, Pedemonte N, Pesce E, Padoan R, Ford RC, Meng X, Rusnati M, D'Ursi P. (\$ Authors equally contributed to this work) Virtual Drug Repositioning as a Tool to Identify Natural Small Molecules That Synergize with Lumacaftor in F508del-CFTR Binding and Rescuing. Int J Mol Sci. 2022 Oct 14;23(20):12274. doi: 10.3390/ijms232012274. PMID: 36293130; PMCID: PMC9602983*

#### **2.3.1 Computational drug repositioning pipeline to identify small molecules able to interact in the DP1 sub-pocket synergizing lumacaftor**

As discussed above, the drug repositioning analysis on the putative lumacaftor binding pocket DP1 of the F508del-CFTR, highlighted that DP1 is a very large pocket that is not fully occupied by lumacaftor

alone, or by one of the selected compounds (tadalafil, telmisartan, rutin), all of them with a high molecular weight. This observation, together with the finding of quercetin, prompted the exploration of the DP1 sub-pockets surrounding the lumacaftor binding region, searching for small molecules able to fill the DP1 sub-pockets and synergize with lumacaftor for the F508del-CFTR rescue. To this aim, the previously describe *ad hoc* drug repositioning pipeline was updated for the searching of small molecules (Figure 2.17) which were re-filtered based on their binding into the druggable DP1 sub-regions and ranked based on binding energy value and MW. Further details are reported in Material and Methods, section 3.2.1.

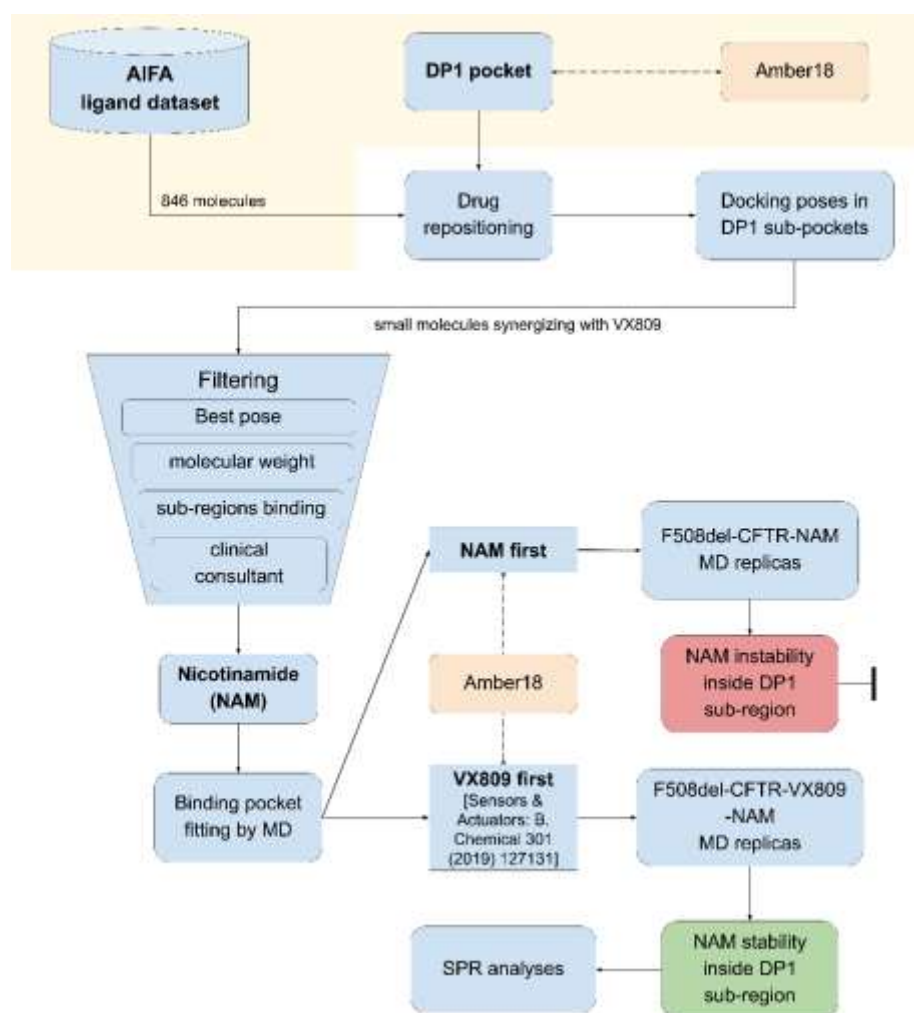

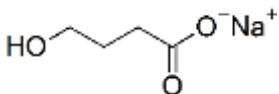
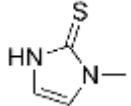
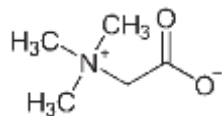
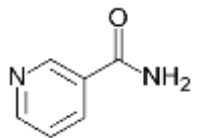


Figure 2.17. Schematic representation of the drug repositioning pipeline adopted. The previously used pipeline (Figure 2.3) was implemented to adapt it for the searching of small molecules able to combine with lumacaftor inside the DP1.

The first five ranked compounds (Table 2.4), being not part of the coded therapies for the treatment of cystic fibrosis patients, have been evaluated by an external clinical consultant on the basis of the AIFA and EMA documents ([www.farmaci.agenziafarmaco.gov.it](http://www.farmaci.agenziafarmaco.gov.it) and [www.ema.europa.eu](http://www.ema.europa.eu)) to discharge those with side effects contraindicated for cystic fibrosis patients. Also, to highlight whether they could have any favorable effect on the treatment of cystic fibrosis, research on PubMed was performed searching: “cystic fibrosis” AND “compound name”.

Table 2.4. Small molecules resulting from the drug repositioning

Molecule Name	Molecular weight (g/mol)	Docking Score (kcal/mol)	2D structure
Cysteamine	77.15	-2.7	
Sodium oxybate	104.17	-4.4	
Methimazole	114.17	-3.6	
Glycine Betaine	118.16	-3.9	
Nicotinamide	122.13	-5.4	

Sodium oxybate, methimazole, and glycine betaine were discarded by the clinical consultant due to their possible side effect on CF patients. Sodium oxybate is used to prevent cataplexy, narcolepsy, and

in alcohol withdrawal syndrome. Its common adverse effects are enuresis and sleepwalking, while less common are lack of appetite, suicidal thoughts, trouble sleeping, unusual weight gain or loss, and changes in behavior. Methimazole is useful in treating hyperthyroidism, but it can have particularly serious side effects, such as agranulocytosis, leukopenia, thrombocytopenia, aplastic anemia, and hypoprothrombinemia and bleeding. Finally, glycine betaine is administered in patients with inborn errors of methionine metabolism. Its adverse effects are mainly related to the gastrointestinal system which could already be compromised in CF patients. On the contrary, cysteamine and nicotinamide (NAM) were found to be usable to treat the pathology, being more compliant for CF patients. Cysteamine was originally prescribed for nephropathic cystinosis, with known adverse effects limited to myopathy, diabetes, and hypothyroidism. However, it was already extensively studied in the CF field finding good and promising results even in patients [87-93], but it was already demonstrated to lack the F508del-CFTR-rescuing activity [94,95]. So, on the basis of the novelty of the possible use of nicotinamide for the rescue of the CFTR function, it was taken into account for further investigation. NAM (or niacinamide) is a water-soluble vitamin, the amide of vitamin B3 also known as Vitamin PP (Pellagra Preventing). It is a constituent of two coenzymes [nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP)] which act as hydrogen and electron carriers through oxidation and reversible reductions and play a vital role in cellular metabolism. Therefore, NAM is an important precursor of NADH and NADPH. Importantly, there are no special warnings and precautions for its use, and no significant pharmacokinetic or pharmacodynamic interactions with other medicines emerged from the analysis of the clinical consultant. Interestingly, it has been suggested to exert a protective effect on acute lung damage caused by ischemia, endotoxin, or oxidative stress. Altogether these features make NAM usable in patients with cystic fibrosis.

### **2.3.2 Pocket Fitting Evaluation for Lumacaftor and NAM: Which Molecule Binds First?**

As described above, NAM was selected as a probable hit, and the docking pose obtained from drug repositioning located the molecule in a sub-region of DP1, in the interface between NBD1 and ICL4 (Figure 2.18). It is close to the hydrophobic cluster between F508, W496, F1068, and, partially, with F1074, which is disrupted by the F508 deletion [96]. This made NAM very interesting for the F508del-

CFTR rescue activity because it could be able to perform recovery by itself or in combination with lumacaftor.

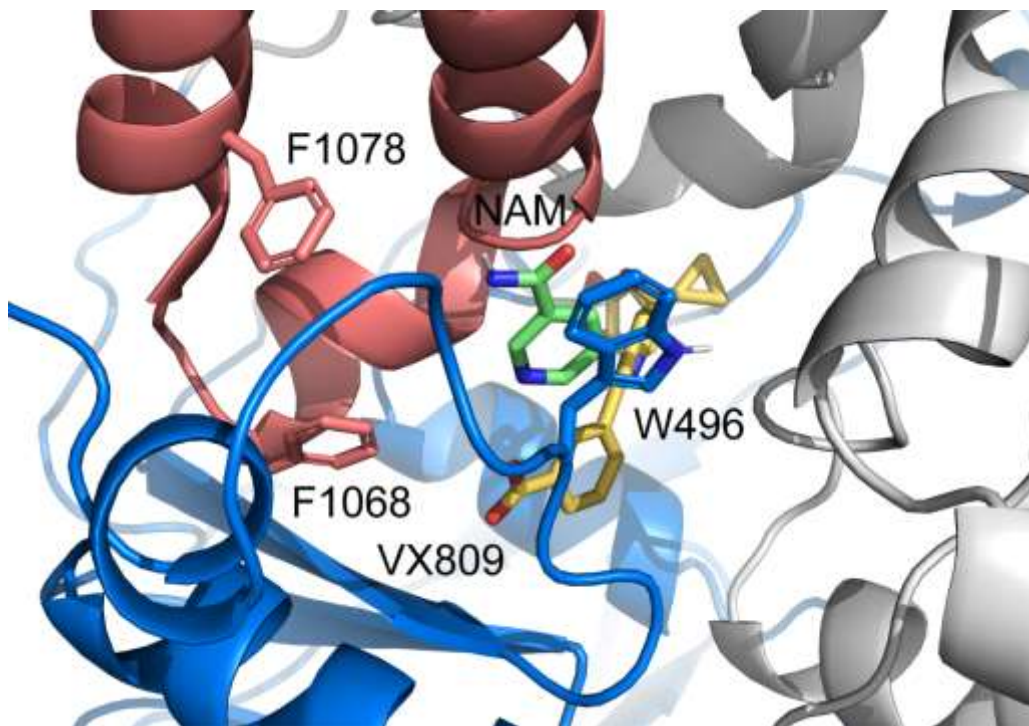


Figure 2.18. Docking pose of NAM in combination with the F508del-CFTR-lumacaftor structure. In blue is the NBD1 domain and in red is the ICL4.

To investigate this second possibility, NAM was evaluated by MDs in combination with lumacaftor considering the protein dynamic behavior at the DP1. With this aim, the fitting determined by the binding of the first ligand, which induces modifications on DP1, so as to allow the binding of the second one, was evaluated. Two different MDs experiments of three replicas were set up. The first was to evaluate the fitting of lumacaftor [71] or NAM in the apo F508del-CFTR, and the latter to evaluate the stability of NAM in complex with the representative conformation of F508del-CFTR-lumacaftor and, vice versa, the stability of lumacaftor in complex with F508del-CFTR after NAM binding. The analyses of the three 50 ns MDs replicas of the F508del-CFTR-NAM complex highlighted that the NAM pose was unstable since the molecule was localized in three different binding poses. As a consequence, was hypothesized that was lumacaftor the one to produce a pocket fitting for the binding of NAM into DP1. Thus, three 50 ns MDs replicas of the complex including both molecules were carried out. The analyses



of the complex demonstrated an overall good binding pose of NAM and lumacaftor inside the DP1. RMSD analyses of F508del-CFTR, lumacaftor, and NAM showed an overall stability of the system (Figure 2.19). Moreover, the two drugs showed an RMSD value  $< 2 \text{ \AA}$  in all three replicas from the starting conformation which is an indication of good ligand stability [75].

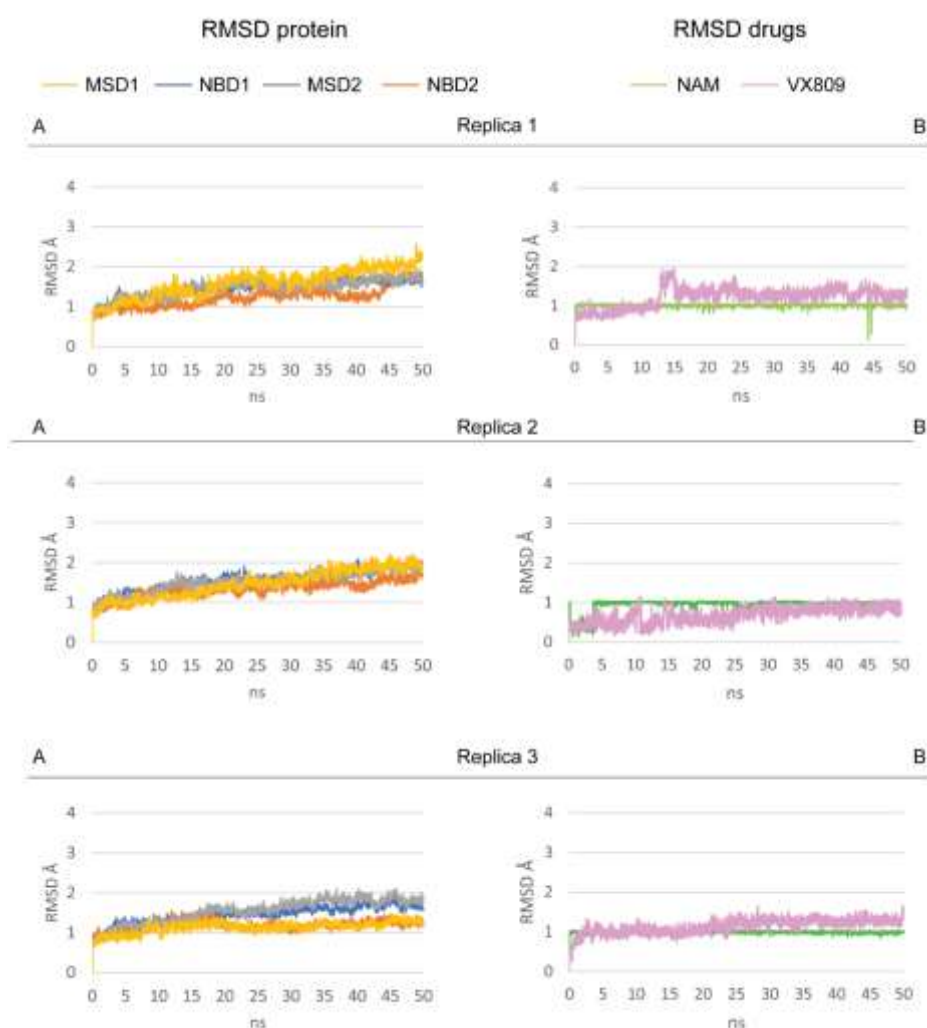


Figure 2.19. RMSD analysis of F508del-CFTR four domains A), and NAM and lumacaftor B) for the three replicas.

Then, to better study the binding pose of the drugs, a hierarchical cluster analysis for each replica was performed. A total of four representative conformations of the F508del-CFTR-lumacaftor-NAM complex were obtained: one for replica 1 representative of 76%, two for replica 2 representative of

47% and 33%, respectively, and one for replica 3 representative of 70% of clustered conformations. Conformation analysis indicated that lumacaftor maintains a similar binding pose in all replicas, while slight differences were found for NAM (Figure 2.20).

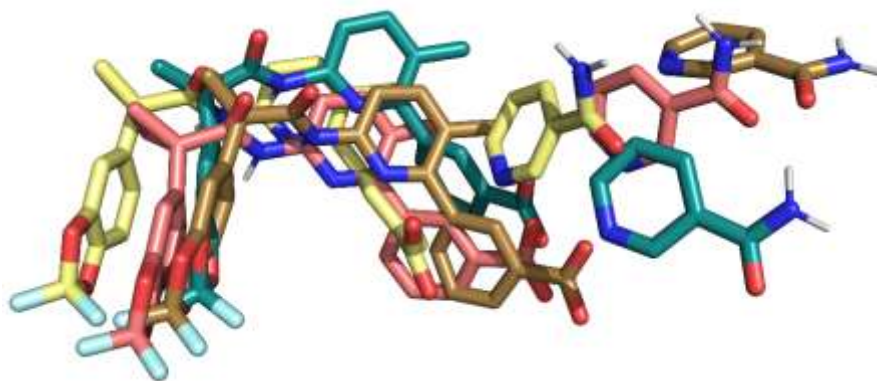


Figure 2.20. Binding poses of NAM and lumacaftor in the representative conformations of the three replicas. Conformation 1 of replica 1 in teal, conformation 1 of replica 2 in coral, conformation 2 of replica 2 in brown, and conformation 1 of replica 3 in yellow.

Only in replica 3 lumacaftor showed a different orientation of its benzoic acid moiety, due to a peculiar positioning of NAM inside the DP1 sub-region. H-bond analysis along MDs of lumacaftor showed differences in its interaction in comparison to its complex with the apo F508del-CFTR [71]. In the combined analyses, due to the presence of NAM, lumacaftor only interacted with W1063 (83.7%, 79.5%, and 84.1% for replica 1, 2, and 3, respectively). The variation of lumacaftor hydrophobic interactions with the representative conformation complex was evaluated by PLIP [97] and along the replicas' trajectories. To calculate these values, the distances between the hydrophobic portion of the protein residues and the interacting hydrophobic portion of lumacaftor among the trajectories were analyzed, with 6 Å as the cut-off value for the distance between the center of mass of the hydrophobic atoms. The analysis highlighted that lumacaftor interacted with F494, K1060, and W1063, in agreement with the previous study [71]. Moreover, lumacaftor created stable hydrophobic interactions with L172, D173, and I177 of NBD1, and D1341 of NBD2 (Table 2.5).

Table 2.5. Average distances of the three replicas between the hydrophobic portion of lumacaftor and NAM with the interacting F508del-CFTR hydrophobic portion residues.

	L171	L173	I177	F494	K1060	K1063	D1341
Lumacaftor	(ICL1)	(ICL1)	(ICL1)	(NBD1)	(ICL4)	(ICL4)	(NBD2)
	5.05 Å	3.95 Å	4.92 Å	5.55 Å	4.45 Å	4.11 Å	5.14 Å
	W496	L1065					
NAM	(NBD1)	(ICL4)					
	5.31 Å	4.53 Å					

Concerning the binding pose of NAM inside the DP1 sub-region, its molecular interactions with the protein were also analyzed (Figure 2.21).

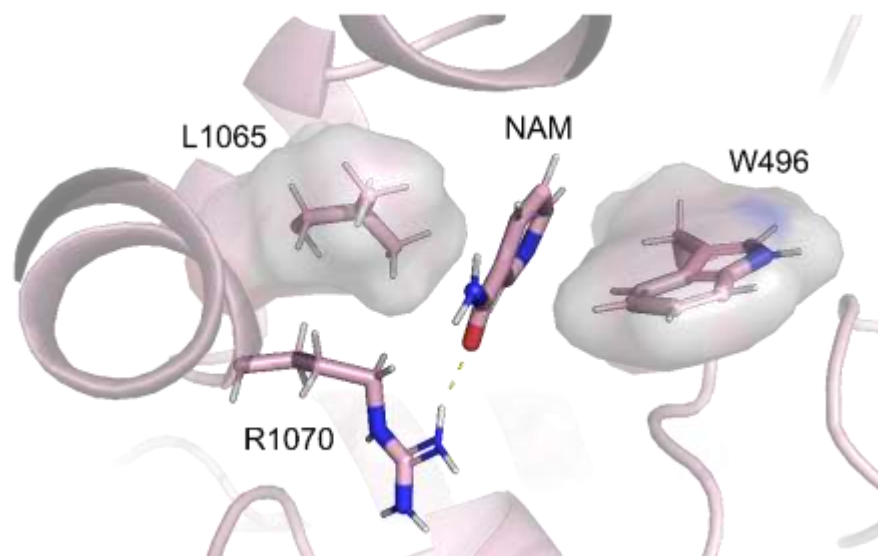


Figure 2.21. NAM binding pose in the F508del-CFTR-lumacaftor-NAM complex. The H-bond is shown in yellow dotted lines. Residues involved in hydrophobic interactions with NAM are depicted with their hydrophobic surfaces.

H-bond analysis highlighted a strong interaction between NAM and R1070 of ICL4 (66,4%, 77,4%, and 83.6% for replica 1, 2, and 3, respectively). A hydrophobic interactions analysis, performed as described above for lumacaftor, indicated the interaction of NAM with W496 of NBD1 and L1065 of ICL4 (Table 2.5). Visual inspection of NAM binding poses along the three replicas showed that the molecule is anchored inside the DP1 sub-pocket, with a single H-bond with an R1070 side chain. Comparing NAM representative conformations obtained from replicas was observed that the side chain of R1070 anchored NAM by an H-bond in a spatial restraint around 6.13 Å. Despite this defined spatial movement determined by the R1070 side chain, the NAM hydrophobic network is preserved thanks to the NAM pyridinic ring, which is steadily interacting with residues L1065 and W496. These results highlighted the ability of NAM to occupy a very specific DP1 sub-region inside the F508del-CFTR, exploiting an H-bonding and several hydrophobic interactions (when lumacaftor is bound). F508del mutation involves the deletion of phenylalanine in NBD1 which contributes to generating molecular contacts at the ICL4/NBD1 interface. The elimination of F508 causes a disruption of a hydrophobic cluster formed by residues F508, W496, F1068, and F1074, located at the interface between NBD1 and ICL4. The hydrophobic interactions between the NAM's pyridine moiety and residues W496 and L1065, both located in the ICL4 loop, might help in restoring this compromised interaction between ICL4 and NBD1. According to calculations, NAM could restate that interaction between W496 and F508, which is lost after mutation, and in addition is able to interact with L1065, a residue recently defined as able to revert F508del mutation effects on CFTR [98].

### **2.3.3 Surface Plasmon Resonance analysis**

SPR was implemented also for the continuous of the CFTR study, thanks to Prof. Marco Rusnati from the University of Brescia. The already used biosensor [71,73] was here utilized to demonstrate the effective capacity of NAM alone to bind to F508del-CFTR. Results showed that NAM binds the mutated protein in a dose-dependent and saturating manner with a  $K_d$  of  $2.5 \pm 1.3 \mu\text{M}$ . Interestingly, in the same experimental conditions, lumacaftor binds F508del-CFTR with a significantly higher  $K_d$  (20–30 times) than NAM [71,99].

### 2.3.4 Identification of the Binding Pocket of NAM to the apo F508del-CFTR

SPR analysis demonstrated the interaction of NAM also to the apo F508del-CFTR, so its binding pose was computationally predicted. Due to the instability of NAM alone inside the DP1 sub-region, its binding site outside the DP1 was computationally evaluated. NAM was docked against the apo F508del-CFTR pocket library obtained in previous work [71]. The best docking pose (-6.1 Kcal/mol) located in the ICL2:NBD2 interface was selected for further analysis. The complex between the docking pose of NAM and the apo F508del-CFTR was evaluated by 50 ns MDs and its stability was confirmed by the RMSD analysis. Furthermore, the binding of NAM in this protein region was characterized by H-bond interactions with A274 (ICL2) and S1359 (NBD2) for 72.6% and 99.5% of lifetime along the trajectory, respectively, were observed (Figure 2.22).

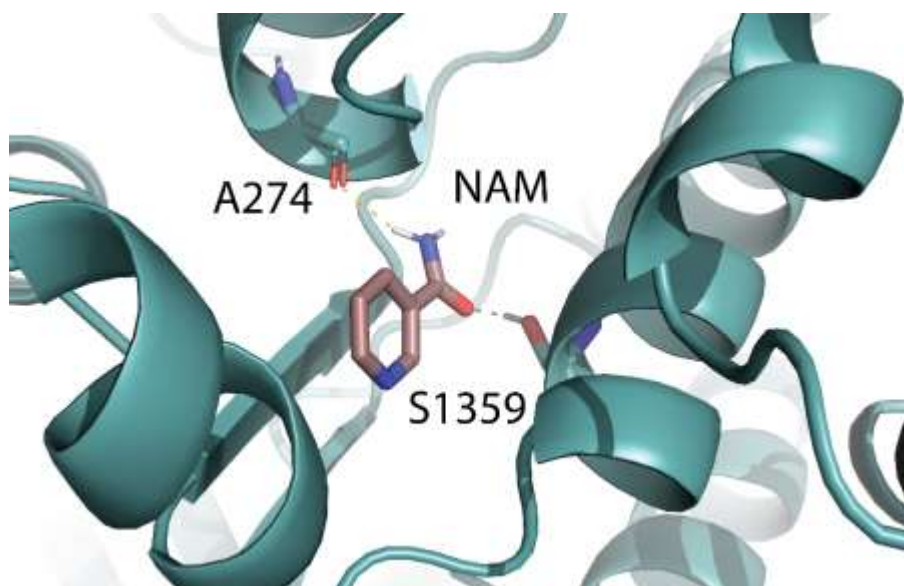


Figure 2.22. Binding pose of NAM in the ICL2:NBD2 interface when it binds to the apo F508del-CFTR. H-bond in yellow.

### 2.3.5 Effect of NAM-Lumacaftor Co-Treatment on Mutant CFTR Rescue

To be sure of the putative NAM capability in the rescue of the F508del-CFTR when given in combination with lumacaftor, Dott.ssa Nicoletta Pedemonte from Istituto Giannina Gaslini, tested the effect of NAM alone and in co-treatment with lumacaftor on F508del-CFTR activity by using the Halide-Sensitive Yellow Fluorescent Protein assay [100]. Cells were treated with lumacaftor in the absence or in the

presence of increasing concentrations of NAM (from 6,25 to 100  $\mu\text{M}$ ). Results showed that a bell-shaped dose response is observable, with a modest but significant increase in the rate of YFP quenching observed with NAM at 50  $\mu\text{M}$  suggesting an augmented CFTR rescue and expression at the plasma membrane, when cells were co-treated with lumacaftor and increasing concentrations of NAM (Figure 2.23). Besides these first results, further analyses are needed to investigate the biological significance of the improved rescue in more relevant cystic fibrosis cell models.

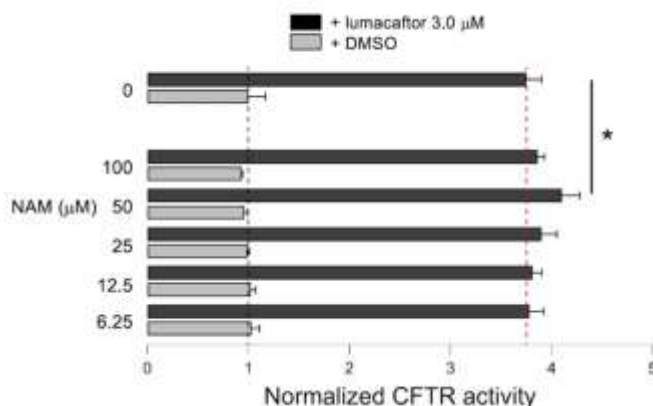


Figure 2.23. Effect of NAM-lumacaftor co-treatment on mutant F508del-CFTR rescue. The bar graphs show F508del-CFTR activity in CFBE41o-cells stably expressing the HS-YFP. CFTR activity was determined as a function of the YFP quenching rate following iodide influx in cells treated for 24 h with DMSO in the absence (vehicle) or in the presence of lumacaftor (3.0  $\mu\text{M}$ ) as a single agent or combined with the indicated concentrations of NAM. \*  $p < 0.05$  (Picture from Fossa P., Uggeri M., *Int J Mol Sci*, 2022, 23(20):12274).

### 2.3.6 Discussion and conclusions

Concerning my research activity on CFTR, I focused on the search for already approved drugs for the rescue of the F508del-CFTR mutated protein by means of an *ad hoc* drug repositioning pipeline. Then, due to the finding of quercetin that in SPR analysis has been shown to increase the lumacaftor affinity to F508del-CFTR, I also searched for already approved small molecules able to synergize together with lumacaftor in its putative binding pocket (DP1) for the protein rescue. According to calculations, DP1 has a large volume, and lumacaftor alone does not fill it completely, possibly allowing another drug to bind inside the pocket.

During the last years a great effort has already been undertaken by pharmaceutical companies and by the scientific community to identify alternative pharmacological targets with important results achieved in the development of efficacious combinations of CFTR-rescuing drugs. However, these combinations can be used only for patients older than age 12 [101] and they are not available worldwide due to their cost. In this scenario, the identification of agents endowed with a dual-target activity (i.e., antibacterial and antioxidant or antioxidant and CFTR-rescuing activities combined in one molecule) may lead to increased therapeutic benefits for cystic fibrosis patients [102]. Furthermore, the dietary implementation of nutraceuticals in cystic fibrosis treatment has already been considered [103-105]. Accordingly, CFF USA Registry reported in 2020 that about 40% of CF patients required oral nutrition supplements, such as minerals (sodium chloride, magnesium, zinc), fat-soluble vitamins (A, D, E, K), fatty acids, and probiotics.

Different molecules, among which rutin/quercetin [106-107] are also endowed with CFTR-rescuing activity, as discussed above. Finally, it is interesting to note how some natural compounds (including genistein, digitoxigenin, curcumin, resveratrol, and latonduine) show additive/synergic effects when administered in combination with already approved cystic fibrosis drugs [108-112].

Taken together, these observations call for the systematic search of nutraceuticals with CFTR-rescuing potential for treatment in cystic fibrosis, an aim that here was set out to achieve by a drug approach selecting NAM as a possible hit. NAM is the main precursor of NAD<sup>+</sup> that, in turn, is an essential co-enzyme of redox reactions for adenosine triphosphate production and ATP-dependent metabolic processes critical in maintaining cellular energy. NAD<sup>+</sup> and its precursor NAM are therefore essential for metabolically active tissues such as epithelia. Relevant to cystic fibrosis, NAM exerts a protective effect on acute lung damage caused by ischemia, endotoxin, or oxidative stress [113], and reduces the levels of pro-inflammatory cytokines, inducible nitric oxide synthase, and other cellular and biochemical inflammation markers [114]. Interestingly, the NAM intracellular pyridine nucleotides derivatives have been demonstrated to regulate CFTR-mediated cAMP-dependent Cl<sup>-</sup> conductance [115], strengthening the hypothesis that NAM may positively affect CFTR activity, pointing to NAM as a putative multitarget compound able to act on different pathological aspects of the cystic fibrosis disease (i.e., reducing oxidation, inflammation, and CFTR conductance). Here, NAM was predicted to bind to F508del-CFTR into the putative lumacaftor binding pocket DP1 located in a dynamic interface crucial for CFTR gating

[98]. Interestingly, during the molecular dynamic simulation, NAM created a hydrophobic interaction with W496 and L1065. W496 is a residue that is part of the hydrophobic pattern altered by the deletion of F508. In wild-type CFTR, W496 forms a hydrophobic cluster with F508 itself, F1068, and, partially, with F1074 [96]. Moreover, in agreement with a recently published paper [98], NAM also interacts with L1065, a revertant residue able to partially rescue F508 deletion. Furthermore, the binding of NAM to R1070 could have biological importance. In literature, it is reported how R1070W mutation acts by restoring interactions at the ICL4/NBD1 interface and thus reinstates the protein functionality compromised by F508 deletion [116]. In the same way, NAM bound to R1070 could partially mimic F508, helping the F508del-CFTR rescue. In conclusion, the binding of NAM in the NBD1:ICL4 interface and its interaction with L1065, R1070, and W496 allow to hypothesize a synergic effect between lumacaftor and NAM in F508del-CFTR rescue. Surface plasmon resonance binding analysis experimentally confirmed that NAM effectively binds to F508del-CFTR with an affinity that is even higher than that of lumacaftor, but cell-based assays demonstrated it does not rescue CFTR function when assayed alone in those same experimental conditions in which lumacaftor results effective. Instead, a preliminary evaluation of the effects of NAM/lumacaftor co-treatment on immortalized bronchial cells demonstrated a modest improvement in mutant CFTR rescue. The dissociation between the CFTR-binding and rescuing activity of NAM could be caused by the fact that, when alone, NAM binds to the ICL2:NBD2 interface which, to date, is not known to be implicated in the F508del-CFTR rescue. At variance, when NAM is administered in combination with lumacaftor, the binding of the latter could contribute to the subsequent binding of NAM to the DP1 sub-region at the NBD1:ICL4 interface, synergizing with lumacaftor for the F508del-CFTR rescue. Furthermore, the fact that NAM is a very small molecule (MW equal to 122 Da) could account for its capacity to be accommodated in the DP1 sub-region without competing with lumacaftor, but rather synergizing in correcting the F508del-CFTR folding defect. When administered in combination with lumacaftor, NAM increases the rescuing activity of lumacaftor in a bell-shaped way. Amazingly, a similar dose-response is also displayed by digitoxigenin when administered in combination with lumacaftor, a behavior that has been tentatively explained with a toxic effect exerted by the higher doses of digitoxigenin [112]. However, NAM has been demonstrated to decrease cell viability in culture only at 5 mM [117], a dose that is 50 times higher than those used in the experiments carried out in this work. On the other hand, genistein has also been



demonstrated to affect CFTR activity in a bell-shaped way, possibly due to its interaction at two binding sites of the mutated protein: a high-affinity site that decreases the closing rate and a low-affinity site that reduces the opening rate [118]. It is thus tentative to hypothesize that, at lower doses, NAM adopts the high-affinity binding mode here described that leads to the synergism with lumacaftor in rescuing CFTR activity while, at higher doses, low-affinity aspecific binding(s) could occur that counteract the rescuing effect. Further investigations will be performed as a prosecution of this research.

The combination of drugs is currently an almost mandatory approach to increasing the therapeutic benefits of the treatment of cystic fibrosis [119]. Among drug combinations already approved for clinical use, Orkambi and Symdeko contain correctors and potentiators that act by distinct mechanisms of action in turn mediated by direct binding of the drugs to different regions of the mutated protein [71,120]. Furthermore, the combination of three modulators (Trikafta) shows a synergic effect that is achieved, once again, by the binding of the drugs to different CFTR regions [121]. Finally, the CFTR-rescuing activity of lumacaftor is increased when it is used in combination with digitoxigenin [112] and with C407, another corrector that binds the mutated CFTR in the same region here identified for NAM [122]. Interestingly, two of the three residues interacting with NAM (W496 of NBD1 and R1070 of ICL4) were found to interact also with C407 [122]. In addition to these combinations, the data here presented suggest an alternative approach consisting of structural modifications of known CFTR-rescuing molecules (such as lumacaftor) including the moiety corresponding to a small molecule (such as NAM) that, even if devoid of intrinsic CFTR-rescuing activity, could stabilize the whole binding to the mutated protein, increasing the global CFTR-rescuing effect.

During the writing of this thesis, a Cryo-EM study was published, highlighting the F508del/E1371Q-CFTR structure in complex with lumacaftor and elezaftor, and with lumacaftor, elezaftor, and ivacaftor, respectively [123]. These data also confirm the findings of another recent paper focused on determining the lumacaftor binding site [124]. These Cryo-EM results thus represent a pioneer study in the field of CFTR research, able to pave the way to further achievements. Nevertheless, to fully elucidate the binding mode and the molecular mechanism of action of the modulator approved drugs, other investigations are needed. It is worth noting that the lumacaftor mechanism of action has not been completely elucidated yet. Literature data in fact show that: lumacaftor could bind and stabilize CFTR intermediates that are partially folded [125], while other works suggested how lumacaftor could bind

and stabilize near-native conformation F508del-CFTR trafficking [126,127]. Moreover, as reported in the present thesis, a lot of experimental studies were carried out in the last two decades, proposing a lumacaftor binding site that does not appear to be confirmed by the Cryo-EM. In fact, different biochemical studies evidenced how lumacaftor, as well as other type-I correctors, could bind to the NBD1 or at the NBD1:ICL4 interface and then allosterically stabilize the interface between NBD1 and TMD2 [126,128-130]. These data strongly suggest that lumacaftor could possibly have different binding sites.

## 3 Materials and Methods

### 3.1 Ligand dataset preparation

The ligand dataset was prepared as follows. Every molecule was saved as SMILE in a single file and then converted into the mae file format that is used by the Schrodinger software package. Successively, LigPrep was used to generate accurate, and energy minimized 3D structures of the molecules. The geometries of the generated structures were optimized using a restricted version of the MacroModel™ (Schrodinger package) which mixes high-quality force fields and GB/SA effective solvation model to obtain the optimized structures. LigPrep is also able to correct the ligand structure to reduce downstream computational errors. Epik was used to charge the obtained molecules at physiological pH 7.4. Epik is parametrized to return accurate pKa values of molecules in both water and DMSO. The calculation for this ligand dataset was done in water using the OPLS4 force field. Moreover, the option to run Epik together with LigPrep in a single command was used, so obtaining the minimized 3D molecule structure already protonated. Then, the charged molecules were converted in the mol2 file format and further in the pdbqt file format without modifying the molecules charges and then used for the docking experiments.

### 3.2 Computational and analytical set-up in the searching for approved drug to rescue the F508del-CFTR.

#### 3.2.1 Drug-repositioning pipeline for the searching of approved drugs to rescue the F508del-CFTR in the DP1

A dedicated computational infrastructure was developed to accommodate all computational needs of the repositioning study (i.e., preparation of the dataset for repositioning, docking virtual, screening and MDs). The pipeline was implemented on a High-Performance Computational Infrastructure base on OpenStack Hybrid Cloud Infrastructure including High-Performance Storage, Multi-Core Molecular Screening pipeline for docking studies, and GPU-based MDs.

The pipeline (see 2.2.1 Figure 2.3) is composed of two main parts and the Part B, the one relative to this thesis, is the drug repositioning pipeline in which the AIFA library (see 3.1.1) was first docked against

the Druggable Pocket 1 from the F508del-CFTR-lumacaftor complex [131], using Autodock Vina (see 3.2.2). The resulting docking poses are filtered by binding energy, taking into account those with an energy like lumacaftor [131], then by visual inspection, further considering the non-covalent interactions detected by PLIP [132], and by clinical and literature considerations. The remaining poses of relevant ligands are finally validated with MDs (Amber - see 3.2.3) by stability through RMSD and RMSF analysis. Moreover, a parallel path of the pipeline processes the F508del-CFTR apo form frames from the previous study [131], by means of Fpocket (see 3.2.5) to find all druggable pockets along the dynamics and to compare them with the key residues surrounding the ligand selected poses.

### **3.2.2 Preparation of the dataset for drug repositioning analysis**

The database used for the present study was obtained starting from the approved list of the AIFA drugs containing, at the time of the work, 1130 compounds [www.farmaci.agenziafarmaco.gov.it/bancadatifarmaci](http://www.farmaci.agenziafarmaco.gov.it/bancadatifarmaci). Not appropriate molecules (such as inorganic salts, peptides, proteins and hormones, and contrast agents) were discarded, thus obtaining a set of 846 small-molecule drugs. Then the molecules were prepared as described in 2.1.1.

### **3.2.3 Drug repositioning using docking experiment**

The F508del-CFTR protein model already reported [131] was used for the repositioning task. The choice has been done taking into account that it displays good stereo-chemical parameters, favorable agreement with experimental data [133] and with the available human cryo E.M. models and potential physiological relevance since it refers to a chloride channel protein in its close state.

A local docking was performed to reposition the already approved drugs retrieved from AIFA using Autodock Vina [134]. Semi-flexible docking was used to allow the ligands to sample various conformations, whereas the protein receptor was kept rigid. The drug repositioning was based on 17 representative conformations retrieved from the MDs of the apo F508del-CFTR and 1 representative conformation from the F508del-CFTR-lumacaftor complex, already selected in the same study cited above. The grid box was centered on the DP1 geometric center and the search was performed with the default parameters except for exhaustiveness set at 24.

### **3.2.4 Molecular dynamics simulations replica of the F508del-CFTR in complex with the repositioned drugs**

Three independent MDs analyses for each complex were carried out using ff12SB force field for the protein, phosaa10 for the modified residues of the protein, and lipid14 [135] for the DOPC lipid bilayer, using the AMBER-14 package [136]. The complexes were solvated with the TIP3P water model and neutralized by the addition of counter ions. The following parameterization for the simulations was used:

- Twelve minimizations of 5000 steps with decreasing restraints on the whole system, and a thirteenth minimization of 50000 steps without restraints, checking the reaching of the system energy plateau. The non-bonded cut-off was set at 8 Å.
- Heating at constant volume, increasing temperature (T) from 0 K to 100 K in 50 ps, then 50 ps at constant 100 K with 6 kcal/mol restraint for either protein and ligand; then 75 ps increasing T from 100 K to 300 K and 25 ps at constant 300 K with 4 kcal/mol and 6 kcal/mol restraint for protein and ligand, respectively, checking the reaching of the desired temperature.
- Equilibration at constant pressure of 4 ns at constant 300 K with a restraint of 2 kcal/mol and 6 kcal/mol for protein and ligand, respectively.

From equilibration, each independent simulation was carried out using different initial velocities.

- MDs production of 25 ns without any restraints.

The clusterization of the initial apo model has been performed with AmberTools (cpptraj).

### **3.2.5 MM-GBSA analyses of the F508del-CFTR-ligand complexes**

The Molecular Mechanics Generalized Born Surface Area (MMGBSA) analysis was performed for each simulation using Amber-Tools (MMPBSA.py) to obtain the binding free energy ( $\Delta G$ ) between the ligand (rutin, telmisartan or tadalafil) and the receptor (F508del-CFTR). To perform the analysis, snapshots were obtained from the last 400 ps of each simulation, setting igb = 5 and the ionic strength to 0.100.

### **3.2.6 Fpocket analysis for the identification of DP1 druggable transient sub-pockets**

Fpocket analysis has been performed outside the pipeline with the default parameters in order to evaluate the druggability of the selected pockets. The program Fpocket [137] allows the identification

of druggable cavities on protein complexes from multiple structures sampled in MDs frames and gives for each pocket a druggability score, which was used to analyze the obtained results. Fpocket uses the concept of alpha spheres. An alpha sphere is a sphere that contacts four atoms on its boundary and does not contain an internal atom. During the checking of the protein, very small spheres can be located inside the core of the protein, while large spheres at its surface and intermediate spheres can be placed in clefts and cavities. In this way, a pocket can be detected by filtering an ensemble of alpha spheres of proper radius, from which it is possible to obtain the properties of the pocket by studying the involved residues. So, the pocket detection can be divided into three parts: A) Determination of the entire ensemble of alpha spheres from the protein structure. B) Identification of clusters of spheres close together in order to identify pockets and remove clusters of poor interest. C) Calculation of properties from the atoms of the pocket, in order to score each pocket.

Here, Fpocket has been used to compute the best druggable pockets along the trajectory of the MDs (1250 frames spaced by 20 ps). For each frame, the number of residues of the pocket that belong to a predetermined set (in particular, imatinib and lumacaftor sub-pockets) has been computed and plotted to evaluate a sort of druggability score of these pocket residues along the dynamics.

### **3.3 Computational and analytical set-up in the searching for small molecules able to rescue the F508del-CFTR together with lumacaftor**

#### **3.3.1 Drug repositioning pipeline update**

A new drug repositioning procedure (see 2.3.1 Figure 2.17) was applied to scout small molecules able to induce DP1 fitting upon binding in combination with lumacaftor in the same pocket by further re-examining the above discussed drug repositioning results as follows: the best pose of each drug was selected and further filtered based on its molecular weight (<500 Da), binding in DP1 sub-regions, and pharmacological effects. Ultimately, nicotinamide (NAM) was found as a hit. Subsequently, the F508del-CFTR-lumacaftor-NAM complex stability was evaluated by molecular dynamic simulations using the Amber18 package (see Section 3.3.2). Two series of MDs were carried out: firstly, three MDs replicas were performed using the complex between the selected docking pose of NAM inside the DP1 sub-region of the apo F508del-CFTR conformation to evaluate the binding pose of NAM inside the DP1 sub-

region of the apo F508del-CFTR. Relevant to this point, data concerning the binding of lumacaftor to the apo F508del-CFTR has already been published [131]. Then, three MDs replicas of the selected NAM docking pose in complex with the representative conformation of the F508del-CFTR-lumacaftor complex [131] were performed to assess the ability of the two drugs to combine in the DP1. MDs were eventually followed by surface plasmon resonance analyses [138] to experimentally validate the computational findings.

### **3.3.2 Molecular Dynamic Simulations of the NAM in complex with the F508del-CFTR with or without lumacaftor**

MDs were performed using AMBER-18 [139] with the ff14SB force field for the protein [140] and lipid14 for the DOPC lipid bilayer [135]. The complexes were solvated with the TIP3P water model and neutralized by the addition of counter ions. Following the parameterization used for the simulations:

- Twelve minimizations of 5000 steps with decreasing restraints on the whole system, and a thirteenth minimization of 50000 steps without restraints, both setting the non-bonded cut-off at 8 Å. The reaching of the system energy plateau was checked.
- Heating at constant volume, increasing the temperature from 0 K to 300 K as follows: 0–100 K in 50 ps then further 50 ps at constant temperature 100 K, using 1 kcal/mol restraint for protein (and lumacaftor when present) and 4 kcal/mol for NAM in both steps; 100–300 K in 75 ps, then further 25 ps at constant temperature 300 K, decreasing the restraint on NAM to 2 kcal/mol. The reaching of the desired temperature was confirmed. From the initial step of the heating, each independent simulation was carried out using different initial velocities
- 4 ns of equilibration at constant pressure at 300 K. Protein and ligand(s) were either restrained with 1 kcal/mol
- 50 ns MDs production without any restraint.

All of the trajectory analyses were carried out using AmberTools (cpptraj), evaluating the binding pose of the ligand and the stability of the protein.

## PART 2

### Chapter 1 - LRIG2

#### 4 Introduction

##### 4.1 The Urofacial Syndrome

The Urofacial Syndrome (UFS), also called Ochoa Syndrome, is a rare genetic disorder characterized by congenital urinary bladder dysfunction, abnormal facial movements with a facial grimace when attempting to smile, as well as bowel dysfunction and nocturnal lagophthalmos (Figure 4.1) [141,142].



Figure 4.1. On the left, an enlarged organ as well as the retrograde passage of urine into one ureter is visible in black due to the introduction of a radio-opaque dye (black) into the urinary bladder, after the bladder outflow obstruction. On the right, the abnormal facial movement of the child's face while smiling (Picture from Stuart HM et al., *Am J Hum Genet.*, 2013, 7;92(2):259-64).

UFS Complications include constipation, urinary incontinence, ascending bacterial infections, vesicoureteral reflux, and renal failure [141,143].



The UFS prevalence is unknown and only about 150 cases have been reported in the literature [141], being one of the rarest human congenital urinary tract diseases (Figure 4.2).

Urinary tract disease	Prevalence
Bladder exstrophy	0.002% births
Megabladder	0.30–0.06% in first trimester
Posterior urethral valves	0.01% births
Primary non-syndromic vesicoureteric reflux	Estimated 1–10% in young children
Prune belly syndrome	0.004% live births
Ureteropelvic junction obstruction	Up to 0.4% of newborns
Urofacial syndrome	Prevalence unknown but around 150 postnatal cases reported

Figure 4.2. Table from *Woolf AS et al., Front Pediatr., 2019, 11;7:136* showing the prevalence of human congenital urinary tract diseases.

The UFS prevalence is likely to be higher in some populations as a result of associated consanguinity or founder variants (e.g., in Columbia) [141,143], however, the total UFS prevalence may be underestimated because the subtleness of the abnormal facial movement may not make it recognizable until late during the renal tract disease [144].

UFS is characterized by a biallelic loss of function mutations in the *HeParanaSE 2 (HPSE2)* or in the *Leucine-Rich-repeats and ImmunoGlobulin-like-domains 2 (LRIG2)* genes located in the chromosome 10 and 1, and resulting in UFS1 or UFS2, respectively [141]. Since UFS1 and UFS2 are reported to have the same phenotype, thus being clinically indistinguishable, an overlapping role of *HPSE2* and *LRIG2* in tissue differentiation and neural growth into the bladder has been suggested, as the two proteins were found to be co-expressed in the human fetal bladder and to be immuno-localized to nerve fascicles growing into the bladder wall, even if the biological functions of the two proteins are yet to be elucidated [145-147].

Also, UFS2 has been recently proposed to be part of a disease spectrum (Figure 4.3) including different syndromes characterized by bladder voiding dysfunctions with or without facial dysmorphisms, such as

the Hinman syndrome [145,148], for whom pathogenic compound heterozygous mutations in the LRIG2 gene were found [147].

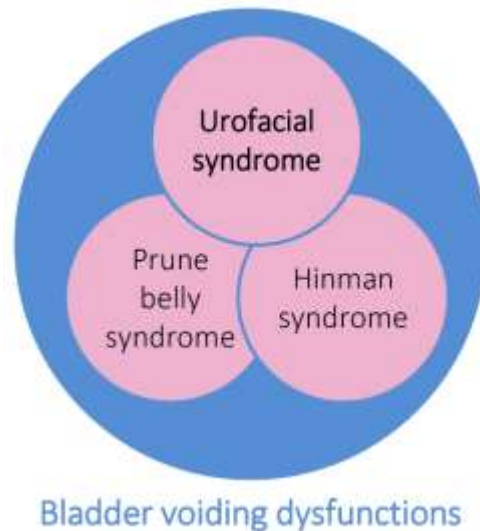


Figure 4.3. diseases characterized by bladder voiding dysfunctions with or without facial dysmorphisms.

#### 4.2 LRIG2 protein

LRIG2 belongs to the LRIG protein family, that in humans is composed of three members (LRIG1-3) which are key regulators of growth factor receptors, including tyrosine and serine/threonine kinase receptors. All LRIG proteins share a common protein structure characterized by an ectodomain, consisting of 15 leucine-rich repeats (LRR), 3 immunoglobulin-like domains (Ig-like), a transmembrane segment, and a large cytoplasmic tail (Figure 4.4) [149,150]. In particular, LRIG2 is composed of 1065 amino acids and in comparison to LRIG1 and LRIG3 lacks the LRR flanking cap.

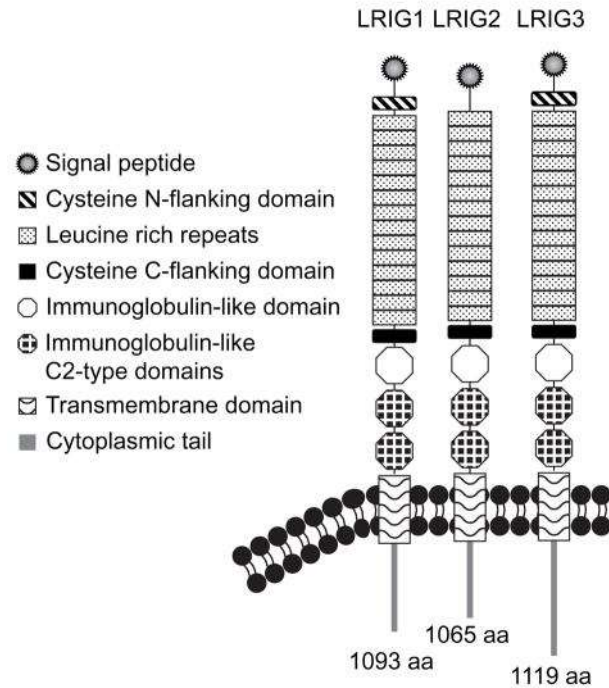


Figure 4.4. Representation of the protein structure of the human LRIG1, LRIG2, and LRIG3 (Picture from Simion C. et al., *Endocr Relat Cancer.*, 2014, 21(6):R431-43).

LRIGs show a high homology in their domains, except for the cytoplasmic portion, which differs significantly among LRIGs and may provide unique functions [151,150]. LRIG1 is the most studied protein of this family, and its structure and function are well defined [152], while the structure and the role of LRIG2 is still poorly understood, as well as its role in UFS development.

LRIG2 is reported to act as an oncogene in glioblastoma cells by physically interacting with PDGFR $\beta$  and by the engagement of the ectodomain with EGFR [153,154], thus promoting cell proliferation and angiogenesis and inhibiting apoptosis. Indeed, increased LRIG2 expression correlates with a poorer prognosis in patients with oligodendroglioma [155], cervical squamous cell carcinoma [156], non-small cell lung cancer [153], and glioblastoma [157,158]. Instead, the role of LRIG2 in bladder innervation and the effect of LRIG2 mutations in the pathogenesis of UFS is not yet fully elucidated. LRIG2 is, indeed, reported to play a role in the regulation of neuronal migration during embryonic development through a mechanism not well defined yet [159]. To date, only one nonsense and two frameshift pathogenic variants in homozygosity or compound he

terozygosity in the *LRIG2* gene were reported in UFS2 patients [147]. Recently, the identification of *LRIG2* pathogenic mutations in patients with Hinman syndrome, or “non-neurogenic neurogenic bladder”, characterized by severe bladder dysfunction without the typical grimace of UFS [147], supported the hypothesis that UFS may be a disease spectrum, including different conditions characterized by overlapping clinical, radiological, and urodynamic features, with or without facial defects. This evidence suggests that the molecular pathways in which *LRIG2* and *HPSE2* function, may also be implicated in other common disorders of bladder voiding, including non-syndromic vesicoureteral reflux [147].

## 5 Results

### 5.1 Identification of a novel LRIG2 pathogenic variant

The identification of the impact of pathogenic mutations on protein function is mandatory to highlight the LRIG2 mechanism of action and its role in UFS, but no studies have yet investigated the effect of the observed mutations on LRIG2 function, despite their well-established role in the UFS development. Probably, this is mainly because of the lack of the LRIG2 crystallographic structure, only available for the best characterized member of the protein family, LRIG1 [160].

A new pathogenic LRIG2 isoform, a homozygous splice site variant causing the deletion of the first Ig-like domain (Ig-1) of the LRIG2 protein (Ig1del-LRIG2), was observed and characterized for the first time in two siblings both affected by UFS by Dott.ssa Laura Fontana and her team from the Medical Genetics Lab of ASST Santi Paolo e Carlo (Milan).

The LRIG2 protein and its known pathogenic mutations in the Ig1 domains (H566Y, R550C, and S523R) were studied *in silico*.

The results, which will be presented below in the thesis, led to a manuscript in preparation: *Rondinone O<sup>S</sup>, Uggeri M<sup>S</sup>, Mauri A, Moresco G, Costanza J, Santaniello C, Colapietro P, Nizzardo G, Marfia M, Guarnaccia L, Navone S, Manzoni G, Minoli DG, Grilli F, Milani D, D'Ursi P, Miozzo M.R, Fontana L (\$ equally contributed to this work). Identification of a novel physiological LRIG2 splicing variant associated with motoneuron differentiation and Urofacial Syndrome development by loss of protein dimerization.*

### 5.2 Primary sequence analysis of LRIG1 and LRIG2 Ig-like domains

According to the LRIG2 splicing variant above mentioned, the LRIG2 alternative isoform leads to a protein lacking the first Ig-like domain (Ig1del-LRIG2). To investigate the effect of Ig-like domain loss and, thus, the function of the new LRIG2 alternative isoform, a *in silico* protein modelling and molecular dynamic simulations were performed.

Since the LRIG2 protein structure is not available in the literature, the LRIG2 structure of the 3 Ig-like domains was inferred by comparison with the homologs X-ray LRIG1-3Ig (PDB ID: 4U7M). The amino acid sequence alignment between the two protein domains showed in these domains a 63.9 % of sequence similarity (Figure 5.1). This high similarity allowed the creation of a reliable LRIG2-3Ig model

from the refined structure of LRIG1-3Ig.

Furthermore, LRIG1 dimerizes through its Ig1 and Ig2 domains [160]. In particular, the LRIG1 dimer interface involves the side chains of Val546, Val548, Met556, and Tyr558 of Ig1, which are inserted into the pocket formed by Trp632, Phe640, Met648, Val650, and Phe657 of Ig2, here called Interface-1 (Int-1). In addition, the Phe545 of Ig1 is in close contact with Ala631, Asn633, and Thr673 of Ig2, here called Interface-2 (Int-2) [160]. These residues are highly conserved in LRIG2 (Figure 5.1), thus suggesting that they may be involved in LRIG2 homodimerization.

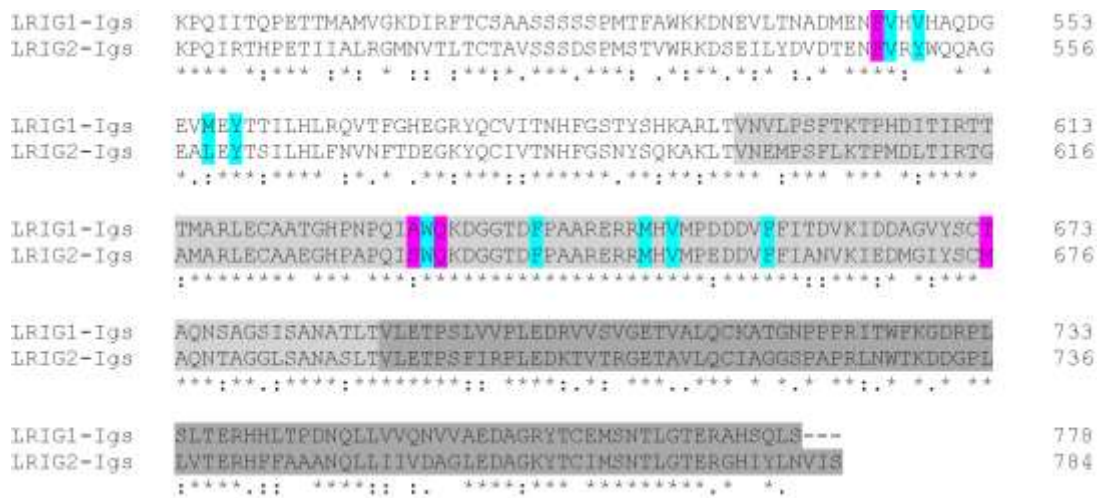


Figure 5.1. Sequence alignment of the LRIG1-3Ig and LRIG2-3Ig. Ig-1, Ig-2, and Ig-3 are highlighted in white, light grey, and heavy grey, respectively. Residues forming Int-1 and Int-2 are highlighted in blue and purple, respectively.

### 5.3 Modelling the dimeric forms of LRIG1-3Ig and LRIG2-3Ig

In order to correlate the del-Ig1-LRIG2 to UFS and investigate the physiological function of the new LRIG2 alternative isoform, the LRIG2-3Ig dimerization was explored by an integrated approach of computational methods such as homology modelling, docking, modelling of the dimeric forms of LRIG1-3Ig and LRIG2-3Ig, dimer stability evaluation by molecular dynamics (MD).

The derived LRIG1- and LRIG2-3Ig models were used to obtain their homodimers applying the procedure described in section 6.1 Protein-protein docking experiments were driven by the experimental data of the LRIG1 homodimer X-ray structure, and LRIG1 was re-docked, using it as a positive control for the validation of the computational pipeline. The 10 best docking poses of the

LRIG1- and LRIG2-3Ig dimers were analyzed by visual inspection. Pose 1 of LRIG1-3Ig homodimer corresponds to the interface described in the literature [160], while the pose n. 5 of LRIG2-3Ig homodimer was the model that best matched the LRIG1-3Ig dimer. These homodimers were superimposed on the LRIG1 homodimer X-ray structure (Figure 5.2) and used for further analyses.

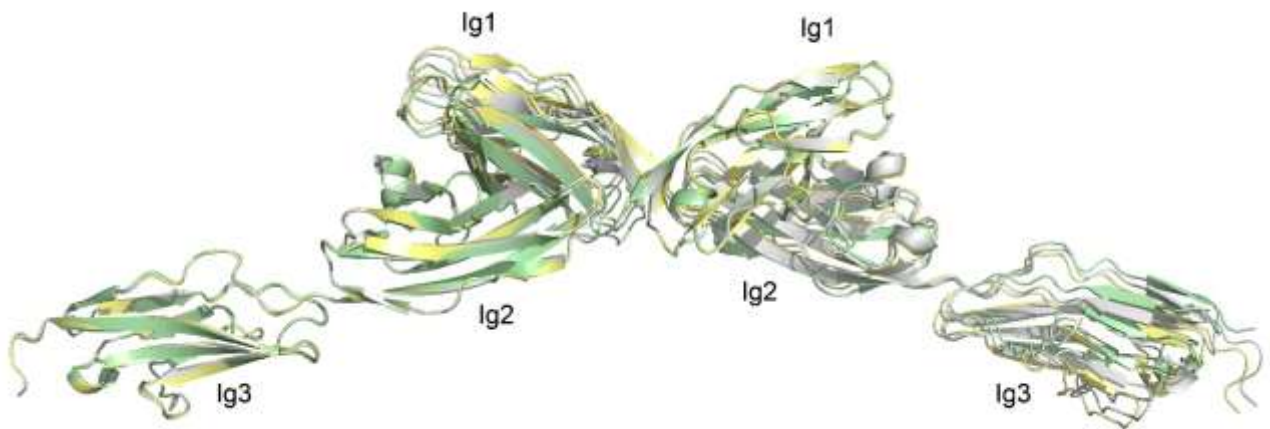


Figure 5.2. LRIG2-3Ig dimer docking pose (green) and LRIG1-3Ig dimer docking pose (yellow) superimposed to LRIG1 X-ray homodimer (grey).

#### 5.4 LRIG1-3Ig and LRIG2-3Ig dimer characterization by molecular dynamic simulations

The stability of the dimeric models obtained from docking analysis was evaluated by 75 ns of MD. Trajectories were analyzed and the Root-Mean-Square Fluctuation (RMSF) showed  $C\alpha$  values of LRIG1- and LRIG2-3Ig subunits in the range of 2.5-5 Å (Figure 5.3). As expected, residues of the Ig3 domain of both dimers had a higher average fluctuation value than Ig1 and Ig2.

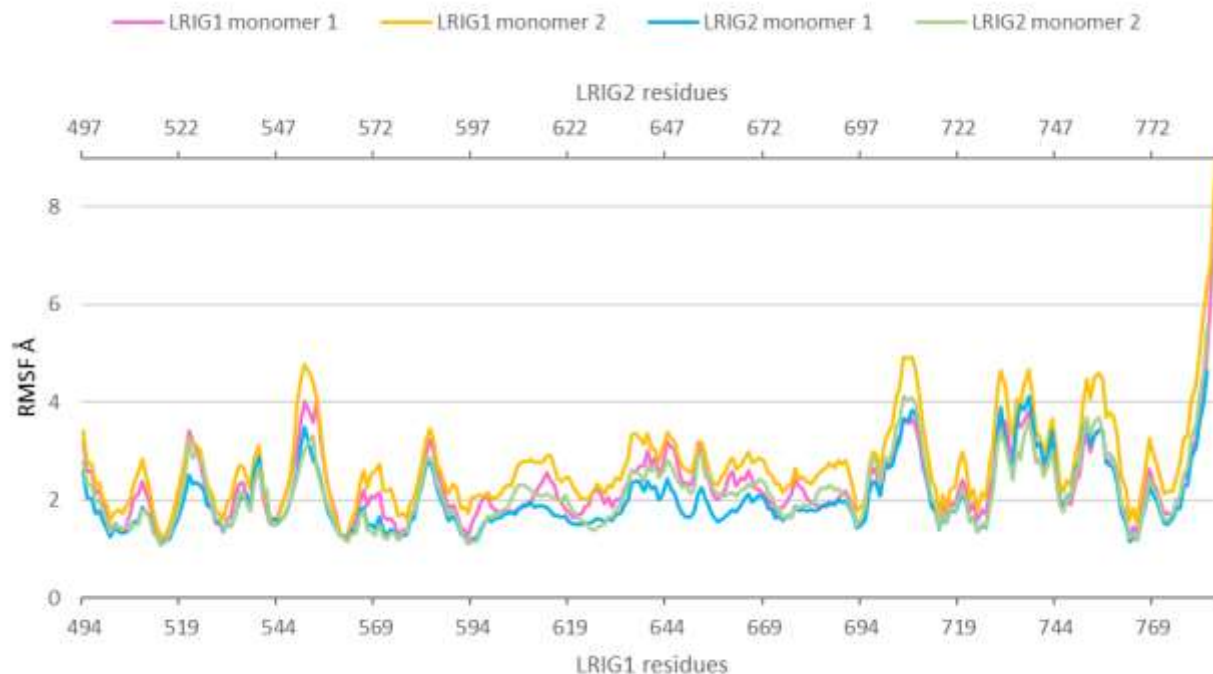


Figure 5.3. Molecular dynamics trajectory analysis of the LRIG1- and LRIG2-3Igs dimers. RMSF analysis of the Cα atoms of each monomer forming the dimer.

Also, the RMS Deviation (RMSD) of residues forming the dimer interfaces was calculated, finding an average value  $< 3.5 \text{ \AA}$  for both LRIG1 and LRIG2 (Figure 5.4). Overall, these results indicated stable dimers and interfaces.



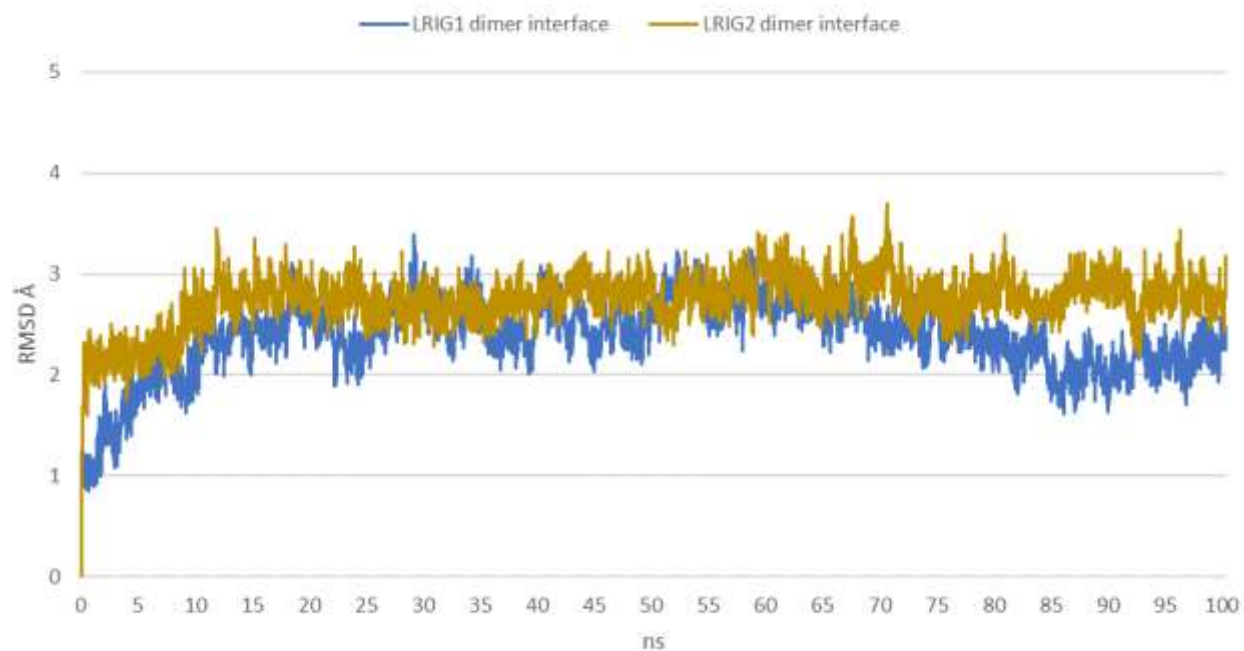


Figure 5.4. Molecular dynamics trajectory analysis of the LRIG1- and LRIG2-3Igs dimers. RMSD analysis of the Ig1-Ig2 dimer interfaces residues (Int-1 and Int-2).

Furthermore, the dimer interface of LRIG1-3Ig and LRIG2-3Ig was characterized by H-bond and hydrophobic interaction analysis. The most representative conformations of LRIG1-3Ig and LRIG2-3Ig dimers (three for LRIG1-3Ig representative of 59% and three for LRIG2-3Ig representative of 69.5%) were obtained by hierarchical cluster and analyzed. The LRIG1-3Ig dimer interface hydrophobic interactions were identified using the X-ray structure (Table 5.1) [160], and the MD results for both LRIG1 (Table 5.2) and LRIG2 (Table 5.3) were comparable.

Table 5.1. LRIG1-3Ig X-ray dimer interaction analysis.

H-bond Interactions									
Position	Residue	Int	Ig	Chain	Position	Residue	Int	Ig	Chain
546	VAL	Int-1	1	A	632	TRP	Int-1	2	B
632	TRP	Int-1	2	A	546	VAL	Int-1	1	B
Hydrophobic Interactions									
Position	Residue	Int	Ig	chain	Position	Residue	Int	Ig	chain
545	PHE	Int-2	Ig1	A	631	ALA	Int-2	Ig2	B
545	PHE	Int-2	Ig1	A	633	GLN	Int-2	Ig2	B
545	PHE	Int-2	Ig1	A	673	THR	Int-2	Ig2	B
546	VAL	Int-1	Ig1	A	632	TRP	Int-1	Ig2	B
546	VAL	Int-1	Ig1	A	640	PHE	Int-1	Ig2	B
548	VAL	Int-1	Ig1	A	632	TRP	Int-1	Ig2	B
548	VAL	Int-1	Ig1	A	648	MET	Int-1	Ig2	B
548	VAL	Int-1	Ig1	A	650	VAL	Int-1	Ig2	B
548	VAL	Int-1	Ig1	A	657	PHE	Int-1	Ig2	B
631	ALA	Int-2	Ig2	A	545	PHE	Int-2	Ig1	B
632	TRP	Int-1	Ig2	A	546	VAL	Int-1	Ig1	B
632	TRP	Int-1	Ig2	A	548	VAL	Int-1	Ig1	B
633	GLN	Int-2	Ig2	A	545	PHE	Int-2	Ig1	B
640	PHE	Int-1	Ig2	A	546	VAL	Int-1	Ig1	B
648	MET	Int-1	Ig2	A	548	VAL	Int-1	Ig1	B
650	VAL	Int-1	Ig2	A	548	VAL	Int-1	Ig1	B
657	PHE	Int-1	Ig2	A	549	VAL	Int-2	Ig2	B
673	THR	Int-2	Ig2	A	545	PHE	Int-2	Ig1	B

Table 5.2. LRIG1-3Ig dimer interaction analysis from MD.

H-bond Interactions										
Position	Residue	Int	Ig	Chain	Position	Residue	Int	Ig	Chain	Lifetime
546	VAL	Int-1	1	A	632	TRP	Int-1	2	B	99.90%
632	TRP	Int-1	2	A	546	VAL	Int-1	1	B	99.90%
Hydrophobic Interactions										
Position	Residue	Int	Ig	Chain	Position	Residue	Int	Ig	Chain	Cluster
545	PHE	Int-2	1	A	631	ALA	Int-2	2	B	1, 2, 3
545	PHE	Int-2	1	A	633	GLN	Int-2	2	B	1, 2, 3
545	PHE	Int-2	1	A	673	THR	Int-2	2	B	1, 2, 3
546	VAL	Int-1	1	A	632	TRP	Int-1	2	B	1, 2, 3
546	VAL	Int-1	1	A	640	PHE	Int-1	2	B	1, 2, 3
546	VAL	Int-1	1	A	648	MET	Int-1	2	B	1, 3
548	VAL	Int-1	1	A	632	TRP	Int-1	2	B	2, 3
548	VAL	Int-1	1	A	648	MET	Int-1	2	B	1, 2, 3
548	VAL	Int-1	1	A	650	VAL	Int-1	2	B	1, 2, 3
548	VAL	Int-1	1	A	657	PHE	Int-1	2	B	1, 2, 3
556	MET	Int-1	1	A	648	MET	Int-1	2	B	2, 3
631	ALA	Int-2	2	A	545	PHE	Int-2	1	B	1, 2, 3
632	TRP	Int-1	2	A	546	VAL	Int-1	1	B	1, 2, 3
632	TRP	Int-1	2	A	548	VAL	Int-1	1	B	3
633	GLN	Int-2	2	A	545	PHE	Int-2	1	B	1, 3
640	PHE	Int-1	2	A	546	VAL	Int-1	1	B	1
648	MET	Int-1	2	A	548	VAL	Int-1	1	B	1, 2
648	MET	Int-1	2	A	556	MET	Int-1	1	B	1, 2, 3
650	VAL	Int-1	2	A	548	VAL	Int-1	1	B	1, 2, 3
657	PHE	Int-1	2	A	548	VAL	Int-1	1	B	1, 2, 3
673	THR	Int-2	2	A	545	PHE	Int-2	1	B	1, 2, 3

Table 5.3. LRIG2-3Ig dimer interaction analysis from MD.

H-bond Interactions										
Position	Residue	Int	Ig	Chain	Position	Residue	Int	Ig	Chain	Lifetime
546	VAL	Int-1	1	A	632	TRP	Int-1	2	B	99.90%
632	TRP	Int-1	2	A	546	VAL	Int-1	1	B	99.90%
Hydrophobic Interactions										
Position	Residue	Int	Ig	Chain	Position	Residue	Int	Ig	Chain	Cluster
548	PHE	Int-2	1	A	634	SER	Int-2	2	B	1, 2, 3
548	PHE	Int-2	1	A	676	MET	Int-2	2	B	1, 2, 3
548	PHE	Int-2	1	A	636	GLN	Int-2	2	B	1, 2, 3
549	VAL	Int-1	1	A	651	MET	Int-1	2	B	1
549	VAL	Int-1	1	A	643	PHE	Int-1	2	B	1, 2
549	VAL	Int-1	1	A	635	TRP	Int-1	2	B	1, 2, 3
551	TYR	Int-1	1	A	651	MET	Int-1	2	B	1, 2, 3
551	TYR	Int-1	1	A	660	PHE	Int-1	2	B	1
551	TYR	Int-1	1	A	653	VAL	Int-1	2	B	3
634	SER	Int-2	2	A	548	PHE	Int-2	1	B	1, 2, 3
636	GLN	Int-2	2	A	548	PHE	Int-2	1	B	1, 2, 3
643	PHE	Int-1	2	A	549	VAL	Int-1	1	B	2
643	PHE	Int-1	2	A	551	TYR	Int-1	1	B	1
651	MET	Int-1	2	A	551	TYR	Int-1	1	B	1, 2, 3
653	VAL	Int-1	2	A	551	TYR	Int-1	1	B	1, 2, 3
660	PHE	Int-1	2	A	551	TYR	Int-1	1	B	1, 2, 3
676	MET	Int-2	2	A	548	PHE	Int-2	1	B	1, 2, 3

Altogether these results evidenced that the computational protocol could reproduce the LRIG1-3Ig homodimer interaction and supported the hypothesis that also LRIG2 may function as a homodimer.

### 5.5 Characterization of LRIG2 point mutations in the Ig1 domain

In order to deepen the role of Ig1 in LRIG2 dimerization and function, the effect of pathogenic mutations affecting the Ig-1 domain already reported in the literature was studied. In particular, the p.H566Y and the p.S523R are associated with UFS, in addition to the p.R550C that has been reported in a patient with Hinman syndrome [161].

S523 and R550 are nearby the residues forming Int-1 and Int-2, while H566 is less close. Residues forming the Ig1 dimer interface (Int-1 and Int-2) are structurally located on the Beta-Strand (BS) 5 and 6 that, together with BS 3, form the Beta-Sheet 1 of Ig1. Based on this evidence, it is conceivable that these mutations may affect the LRIG2 dimer interface, thus interfering with the dimerization. To

deepen this hypothesis, the monomeric LRIG2 p.H566Y, p.R550C, and p.S523R mutations were characterized and compared to the wt-LRIG2 by 500 ns of accelerated MD (aMD).

### 5.5.1 LRIG2 p.H566Y

H566 is on the BS6, relatively far away from Int-1 and, partially, from Int-2. The H566 is surrounded by the side chain of 6 residues of BS3, BS5, and BS6: N514, T516, D543, E546, I564, and F568, but it does not directly interact with anyone, except for the backbone H-bond interaction with D544 of the BS5. The substitution of histidine with tyrosine determines a higher steric hindrance and lipophilicity, which could affect the surrounding environment. Simulation analysis highlighted both a mild increase in the RMSF of the Y566 side chain over the H566, from 1.4 to 1.8 Å, indicating a less stable side chain, and a partial loss of the H-bond interaction between Y566 and D544 (93% to 62% of a lifetime). Interestingly, the comparison between the representative conformations of the wt-LRIG2 and the H566Y-LRIG2 highlighted a huge rearrangement in the quaternary structure of the Ig1-Ig2 connection region, in the mutated protein, which becomes closer (Figure 5.5).

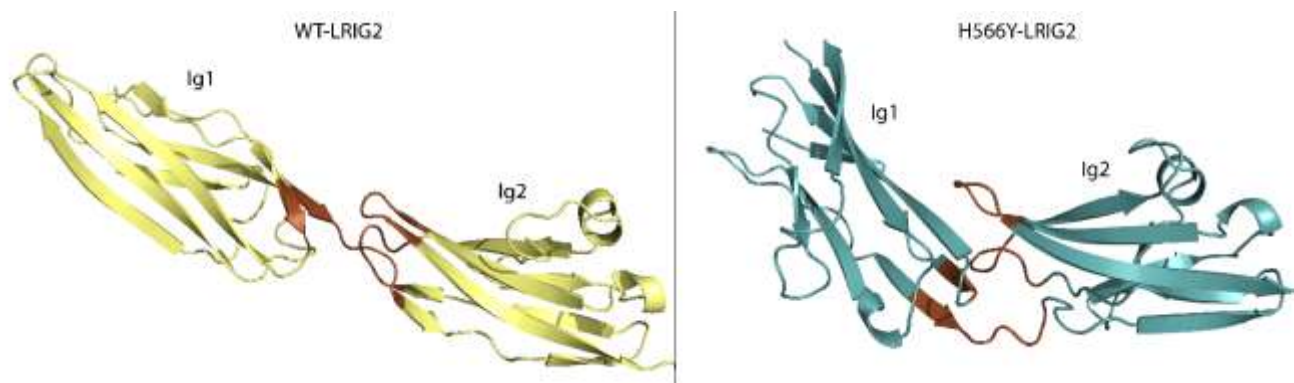


Figure 5.5. Rearrangement of the Ig1-Ig2 connection region (brown) in the quaternary structure. Wt-LRIG2 on the left, H566Y-LRIG2 on the right.

Moreover, Principal Component Analysis (PCA) performed on the H566Y-LRIG2 simulation (Figure 5.6) also confirmed that the Ig1 and Ig2 domains get close together.

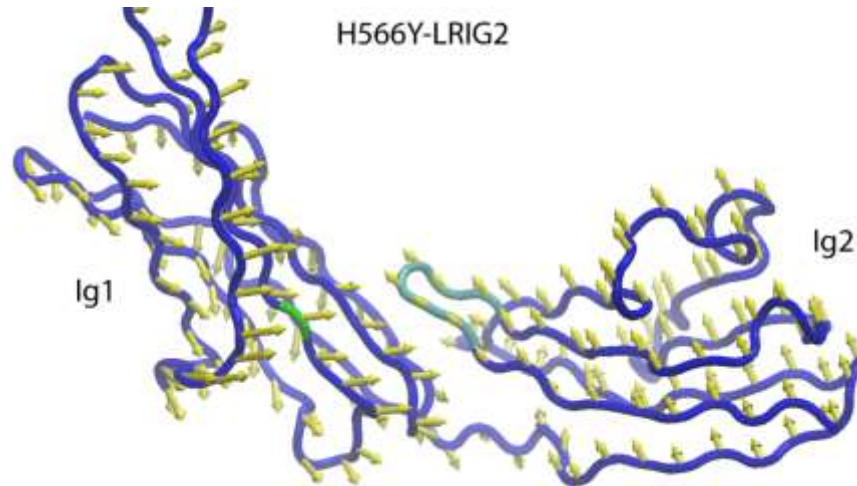


Figure 5.7. Porcupine plots represent the movements of the second eigenvector obtained from the PCA analysis of the H566Y-LRIG2 trajectory. In green the Y566, and in cyan the part of the Ig1-Ig2 connection region which gets close to the Ig1 domain.

Moreover, H-bond analysis showed that residues nearby the Ig1-Ig2 connection region form stable interactions in the wt-LRIG2, which are mainly lost in the H566Y mutated protein. The loss of H-bond interactions (Figure 5.8) made the Ig1-Ig2 hinge more flexible, contributing to bringing the two domains closer.

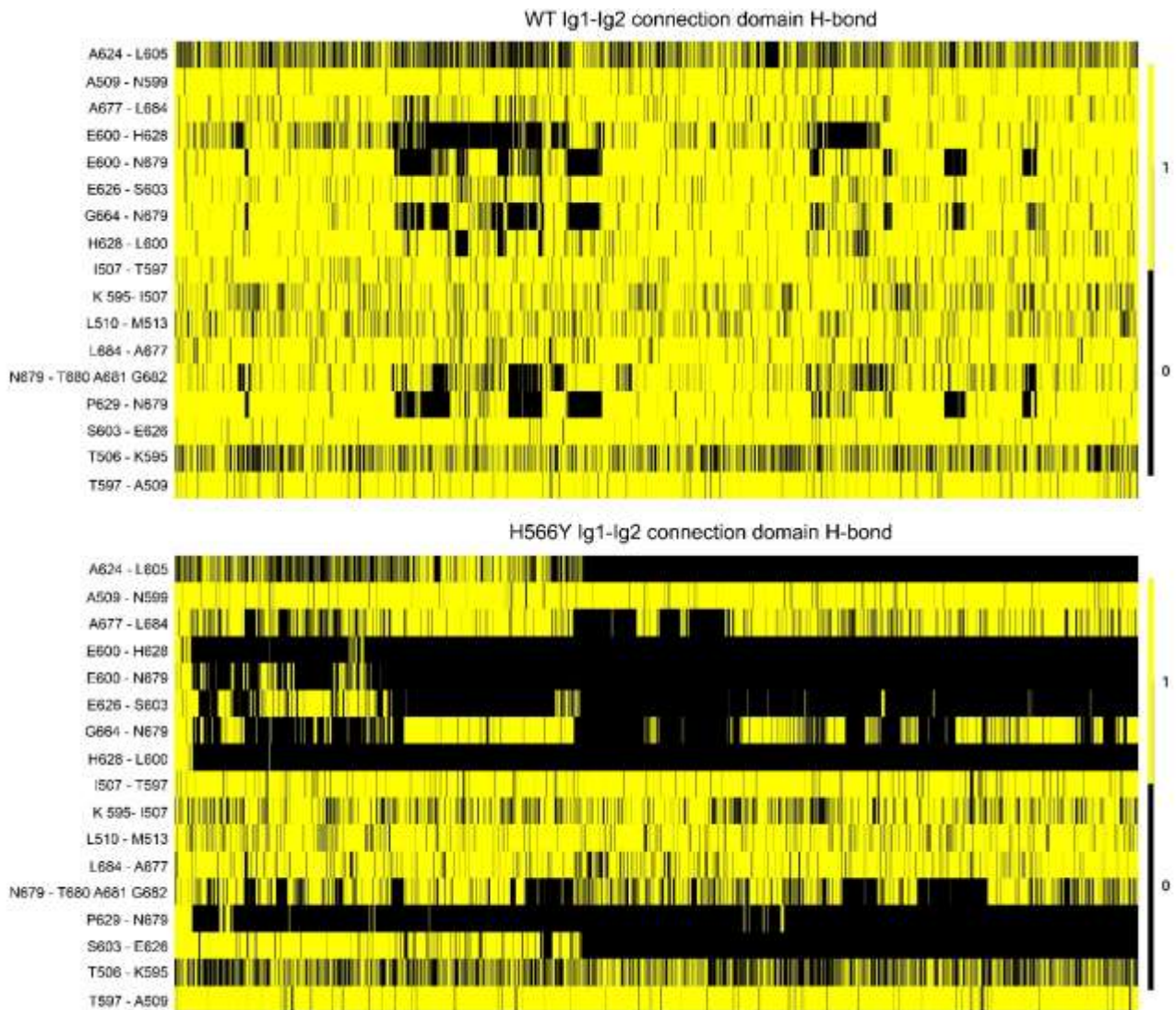


Figure 5.8. Comparison of the H-bond network of the residues nearby the Ig1-Ig2 connection region between wt- and H566Y-LRIG2 during the simulation. For each residues pair, 1 represents the presence of the H-bond, while 0 the absence.

The histidine to tyrosine mutation at position 566 leads to a protein structure rearrangement, bringing the Ig1 and Ig2 domains closer together, altering the binding region for the second monomer and impairing the dimerization.



### 5.5.2 LRIG2 p.R550C

R550 is located on the BS5 of Ig1, between the two Int-1 residues V549 and Y551, but its side chain points in the opposite direction to them. The study of the monomeric wt-LRIG2 aMD trajectory evidenced that the R550 side chain creates a stable H-bond interaction with the carboxyl group of E560 (located on the BS6), with a lifetime of 84%. In turn, E560 creates a stable H-bond with the side chain of R501 (located on the beta sheet 1) with a lifetime of 76%, forming a strong H-bond network and giving stability and rigidity to the beta sheet 1. The R550C substitution leads to the loss of an H-bond between the side chains of C550 and E560 (lifetime 0%), destabilizing the network. In the R550C-LRIG2 protein, E560 strengthens its interaction with R501 (lifetime from 76% to 92%) leading to the creation of a new network, through the H-bond interaction between E560 and T520 of BS3 (70% of lifetime) (Figure 5.9).

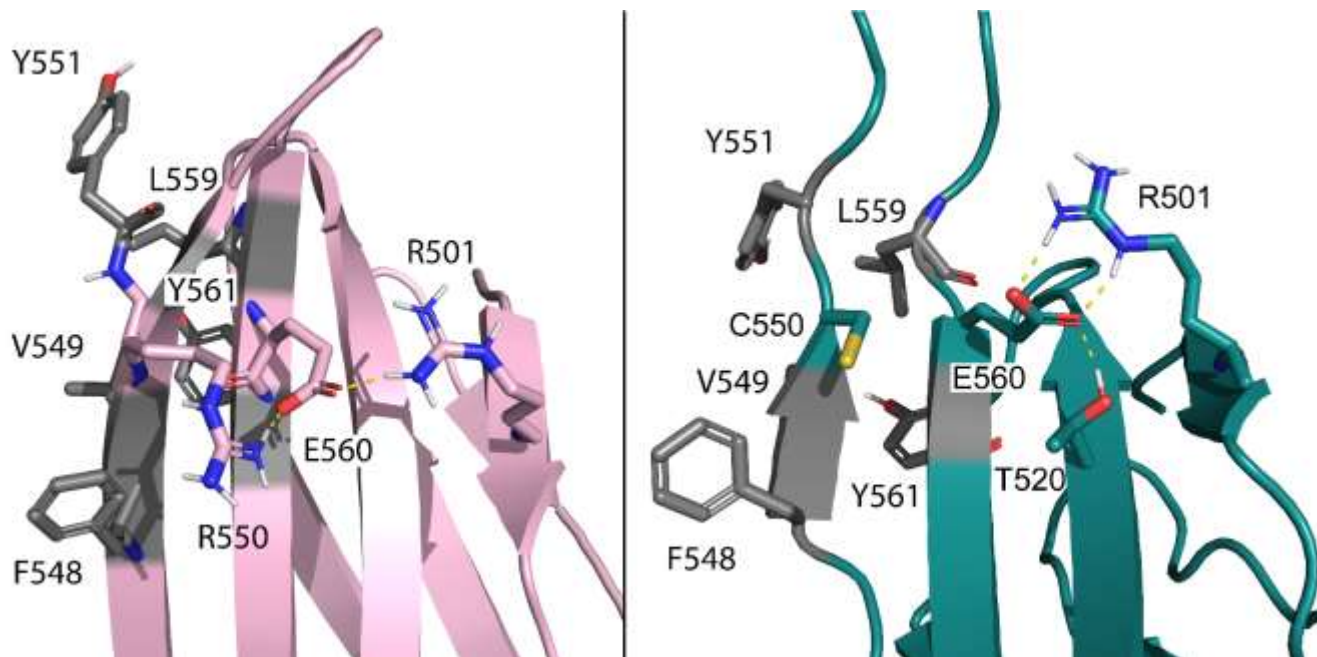


Figure 5.9. H-bond network (yellow dashes) formed by R550, E560, and R501 (pink stick) of the Ig1 domain in the wt-LRIG2 (left), which is lost in the R550C-LRIG2 (right). In grey are also reported residues forming the Ig1 dimer interface.



This new H-bond network between BS6, BS3, and BS1, may change the rigidity and the stability of the BS5, leading to a partial unfolding of the Beta-Sheet 1 (Figure 5.10 and 5.11) and subsequent instability of the Ig1 Int-1 and Int-2.

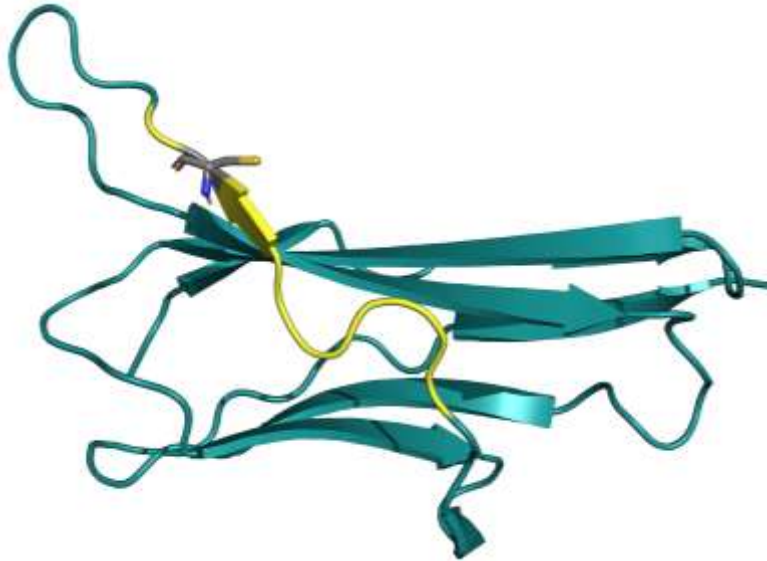


Figure 5.10. Representative structure of the R550C Ig-1 domain after 500 ns of aMD, where it is possible to note the unfolding of the BS5 due to the R550C mutation. In teal is shown the Ig-1 domain, in grey stick the C550 mutated residue, and in yellow the unfolded BS5.



The partial unfolding of the beta sheet 1 together with flexibility changes of BS5 and BS6, which include Int-1 and Int-2 residues, may thus affect protein dimerization in presence of the R550C variant.

### 5.5.3 LRIG2 p.S523R

The S523 is in a loop connecting BS3 and BS4 (523-529 spanning residues) near the Ig1 residues forming Int-1 and Int-2. The wt-LRIG2 simulation showed that the studied loop and the Int-1 region had negative and neutral charged electrostatic potential molecular surface patches. The substitution of the serine to a positively charged residue (arginine) strongly alters the electrostatic potential of the region (Figure 5.12).

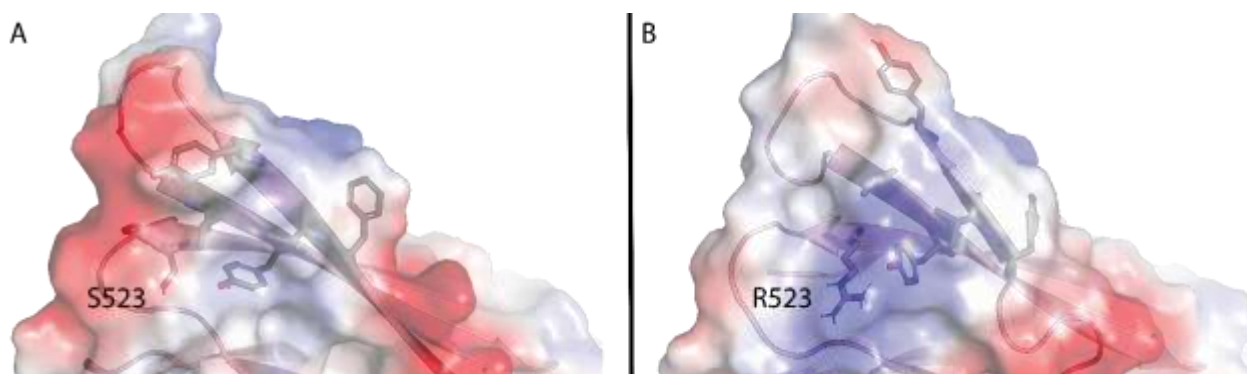


Figure 5.12. Electrostatic potential molecular surface of the wt-LRIG2 (left) and S523R-LRIG2 (right) after 500 ns of aMD.

Moreover, an instability of the R523 residue during the S523R-LRIG2 simulation was observed. In particular, the bulky side chain of the arginine and its charge did not allow it to correctly fit in this specific region. Therefore, the R523 residue cyclically changed the orientation of the side chain (Figure 5.13), pointing towards or outside Int-1 residues, and thus leading to both the loss of the H-bond between the backbone of R523 and L559 (from 94% to 32% of lifetime), one of the Int-1 residues, and a continued rearrangement of the surrounding residues.

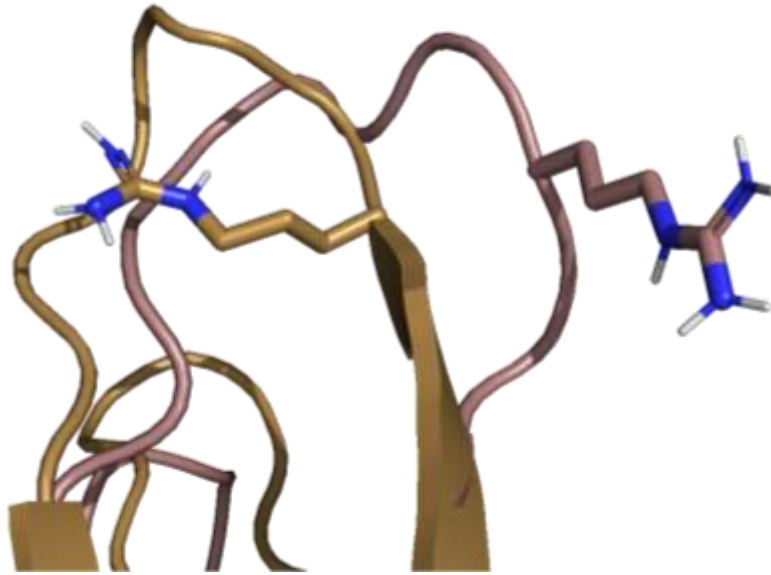


Figure 5.13. R523 side chain cyclically changes during the simulation. The two possible side chain orientations of R523 are reported and represented as stick.

As a consequence of the substitution of the serine to the arginine, the beta sheet 1 partially unfolds (Figure 5.14) as observed in the R550C-LRIG2 protein.

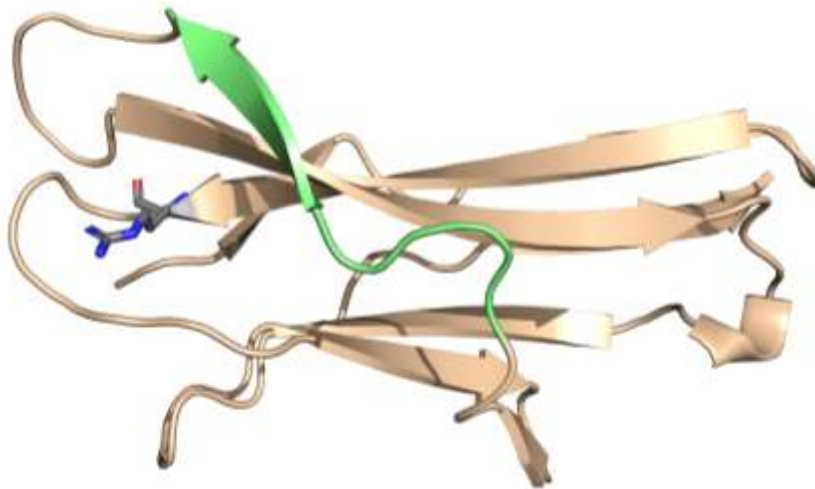


Figure 5.14. Representative structure of the S523R Ig-1 domain after 500 ns of aMD, where it is possible to note the unfolding of the BS5 due to the S523R mutation. In okra is shown the Ig-1 domain, in grey stick the R523 mutated residue, and in green the unfolded BS5.

Altogether, this evidence suggests a dimerization impairment of the S523R-LRIG2 protein.

## 5.6 Discussion and conclusions

The Ig1del is the first splicing variant associated with UFS development. Most LRIG2 mutations in UFS are, indeed, missense and nonsense mutations [161,162]. The identified variant affects the canonical splicing of the LRIG2 pre-mRNA, leading to the skipping of exon 13, and is thus predicted to trigger the expression of a protein isoform lacking the first Ig-like domain.

To deepen the role of this new LRIG2 isoform, the function of the Ig1 domain was computationally studied. Since no crystallographic structure of the LRIG2 protein is available, the structure of the LRIG2 Ig-like domains (3Ig) was inferred by comparison with the LRIG1-3Ig [160]. The comparison demonstrated that LRIG1 function also as a homodimer through the interaction between the Ig1 and Ig2 domains, so in this research it was hypothesized that LRIG2 may function, as well as LRIG1, as dimer with a central role of the Ig1 in protein dimerization. This hypothesis was supported by the high sequence similarity between the LRIG1-3Ig and the LRIG2-3Ig, which has never been proven by *in silico* or *in vitro* studies. By homology modeling and docking-docking approaches, the stability of the dimer and the involvement of the Ig1 and Ig2 domains in protein dimerization were demonstrated. Results showed that, despite some differences in the amino acid composition in the Int-1 and Int-2 regions, LRIG2-3Ig can dimerize using its Ig1 and Ig2 domains, in a similar way to LRIG1-3Ig. Val549 and Met556 of LRIG1-3Ig Int-1 correspond to Tyr551 and Leu559 of LRIG2-3Ig Int-1 respectively, while Ala630 and Thr673 of LRIG1-3Ig Int-2 correspond to Ser633 and Met676 of LRIG2-3Ig Int-2 respectively. Theoretically, methionine to leucine in Int-1 and alanine to serine in Int-2 are well-tolerated substitutions, with similar steric hindrance. Threonine to methionine in Int-2 is a less-tolerated substitution because of the different polarity of the two side chains, but this does not seem to affect the stability of Int-2. Finally, valine to tyrosine in Int-1 is the less-tolerated substitution, with a small and nonpolar residue substituted with a bulky and polar residue. However, also this substitution has no destabilizing effect on Int-1.

The obtained results support a fundamental role of Ig1 in LRIG2 dimerization and allow to hypothesize that the Ig1del-LRIG2 mutant may not be able to create a homodimer, thus possibly contributing to UFS development.

Few other LRIG2 mutations, already reported in the literature, cluster in exons 13-15 that encode for the Ig1 domain. The assumption was that these mutations may affect the Ig1 structure and, thus, prevent protein dimerization, with a comparable effect on protein function as the splicing mutation reported in this study. To deepen this hypothesis, accelerated MD of all these mutations was carried out. In particular, the effect of two variants associated with UFS, H566Y reported as a variant of uncertain significance (VUS), and the S523R reported in ClinVar as likely pathogenic were analyzed. In addition, the R550C variant reported in a patient with Hinman syndrome was also investigated [161].

Regarding the H566Y variant, the affected residue is relatively far away from the Int-1 or Int-2 regions and it does not play a primary role in the stability of the interaction region. However, aMD simulation of the H566Y mutated protein showed a clear rearrangement in the protein structure, mainly due to the loss of the H-bond interactions located in the Ig1-Ig2 hinge, which brought the Ig1 and Ig2 domains closer compared to the wt-LRIG2. Furthermore, principal component 2 from PCA highlighted a correlation of the motion between Ig1 and Ig2 domains, which also involves Y566. Such movement is not observed in the wt-LRIG2, allowing to hypothesize that Y566 plays a role in the protein rearrangement, preventing the correct protein folding and, in turn, inhibiting the LRIG2 homodimerization.

The role of the Ig1 domain in LRIG2 dimerization is further supported by the analysis of the R550C mutated protein. The affected residue is, indeed, localized on the Beta-Strand of the dimeric interface and participates, with the E560 and R501 residues, in a hydrogen-bonding network opposite to the dimeric interface. The R550C mutation breaks the H-bond interaction with E560, thus disrupting the above-mentioned H-bond network and finally leading to conformational changes in the Ig1 domain that prevent LRIG2 dimerization.

Eventually, also the S523R variant was characterized. The mutated residue is, indeed, located in a sequence conserved also in LRIG1 [SSSxSPM] that forms a loop adjacent to the Int-1. The S523 residue is the only serine in the loop with the side chain pointed towards Int-1. Moreover, it creates a stable H-bond interaction (backbone-backbone) with the L559 residue of the Int-1. The S523R substitution strongly affects the protein in this region as a consequence of the completely different properties of the substituted amino acid, which leads to a different dynamic behavior during the aMD. Moreover, both R550C and S523R mutations trigger a closed protein conformation in which the 3Ig folds up and

lose the linear structure which could be fundamental for protein dimerization, thus preventing the interactions of Igs of the two monomers.

Altogether, the results presented on mutations affecting the Ig1 domain support the importance of this domain for protein dimerization and confirmed the pathogenicity of variants reported to date with uncertain significance. Moreover, the identification that also the *LRIG2* mutation associated with Hinman syndrome (HS) leads to the same effect on protein function, further corroborates a common pathogenetic pathway shared by both UFS and HS and the central role of the Ig1 in protein function.

Despite these interesting results, further experiments are needed to confirm these hypotheses.

## 6 Materials and Methods

### 6.1 LRIG1-3Ig and LRIG2-3Ig models, docking, and molecular dynamic simulation

The LRIG1-3Ig and LRIG2-3Ig sequences were retrieved from UniProt: Q96JA1 and O94898 (The UniProt Consortium, 2021), and aligned using Clustal Omega [162]. Swissmodel [163] was used to model the monomeric LRIG2-3Ig, using the LRIG1-3Ig X-ray structure, available in the pdb database (ID: 4U7M), as a template. The monomeric LRIG1-3Ig X-ray structure was firstly refined by adding missing residues, then it was used as input in protein-protein docking experiment to model its homodimer using the LZerD webserver with default settings [164]. The LRIG2-Igs homodimer was also modeled using the same software. LZerD docking program generates tens of thousands of docking models. After clustering (by default at an RMSD cutoff of 4.0 Å), which generally culls models pool from a few tens of thousands to a few thousand, the complex models are then scored by a ranksum method. Among the 10 best models, the homodimer resembling the X-ray LRIG1 dimeric structure, with an interface as described in the literature, for LRIG1-3Ig and LRIG2-3Ig were selected by visual inspection (pose n. 1 and n. 5, respectively) and underwent 100 ns of MD.

The two homodimer systems were solvated in a TIP3PBOX, and minimized in five steps, for a total of 12500 steps of steepest descent and 12500 steps of conjugate gradient with decreasing restraint, from 5 Kcal/mol to 0 Kcal/mol. Then, the systems were heated from 0 to 300 K in 400 ps, applying 1 Kcal/mol restraint on both dimers in an environment with constant volume. For the equilibration step, the systems were switched to an environment with constant pressure and simulated for 4 ns at 300 K with 1 Kcal/mol restraint on both dimers. Subsequently, a 100 ns MD at 300 K was carried out on both systems, without any restraint. The ff14SB forcefield and the Langevin thermostat were used for the simulation. The MD trajectories and H-bond interactions were analyzed using the Ambertools package, while the hydrophobic interactions (max 5 Å) of the LRIGs-3Ig dimer interface were studied using the LigPlot+ software [165].

### 6.2 Modelling of LRIG2 variants and accelerated molecular dynamic simulation

The three mutated monomeric LRIG2-3Ig structures: H566Y, R550C, and S523R were modelled using SwissModel [163] starting from previously obtained LRIG2-3Ig model and characterized by 500 ns of



accelerated molecular dynamic simulation (aMD) using Amber18 [166]. aMD was used to allow the protein to overcome potential energy barriers and better exchange the rate between low energy conformation states. The system was minimized and heated as before, then 10 ns of classic MD was carried out to obtain the value to be applied for the aMD. Trajectories were analyzed using the Ambertools package.

## Chapter 2 - NUP98

### 7 Introduction

#### 7.1 Intrinsically Disordered Proteins and Intrinsically Disordered Regions

Despite in biology the unique biological function of a protein is typically defined by its specific and unique 3D structure, many of them do not need a unique structure to exert their function. These proteins can be completely unstructured or hybrid proteins, containing both structured and unstructured domains, and they are called Intrinsically Disordered Proteins (IDPs) or Intrinsically Disordered Protein Regions (IDPRs), respectively (Figure 7.1).



Figure 7.1. From left to right, an example of a possible intrinsically disordered protein (red loop), intrinsically disordered protein regions (red loop), and a structured protein (white cartoon) (Picture from Raices M. et al., *Nat Rev Mol Cell Biol.*, 2012, 13, 687-99).

IDRs can often be found in nature as they are ubiquitous in all living beings and not, as they were also found in viruses [167,168].

Interestingly, proteins for which 3D structure is available and deposited in the Protein Data Bank (PDB) showed a very high presence of IDPRs with relatively long unstructured domains, with a higher abundance in eukaryotes (more than 30% of proteins are predicted to have IDPRs) in comparison to bacteria and archaea [169-171]. Furthermore, both IDPs and IDPRs can be present in the PDB also as structured proteins since their folding can be induced by a specific binder able to order them [172].

Moreover, IDPs and IDPRs are also capable of simultaneously binding to different receptors, further increasing the complexity [173].

The disordered regions of IDPs and IDPRs are often involved in the signaling, regulation, or control. In fact, in comparison to a region with a rigid and well-defined 3D structure, that has reached its energetic minimum, a disordered region with the lack of secondary or tertiary structure has a conformational dynamism that ensures extraordinary flexibility in binding. IDRs can easily and faster change their conformation due to an environmental perturbation. Consequently, as a direct result of the increasing complexity of living beings, IDRs are more abundant in eukaryotes as increasing signaling, regulation, and control are needed (Figure 7.2) [174].

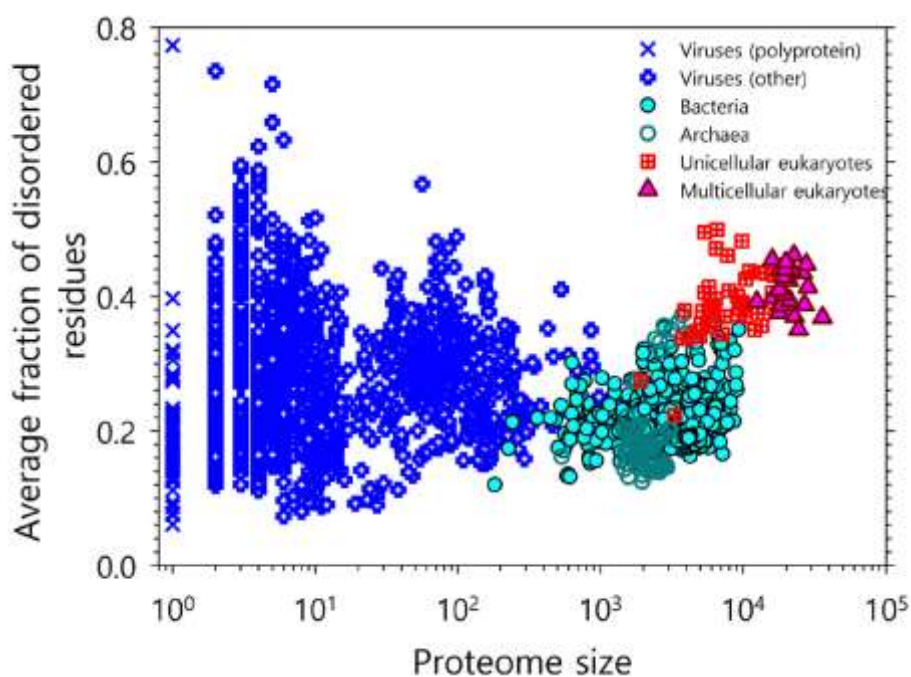
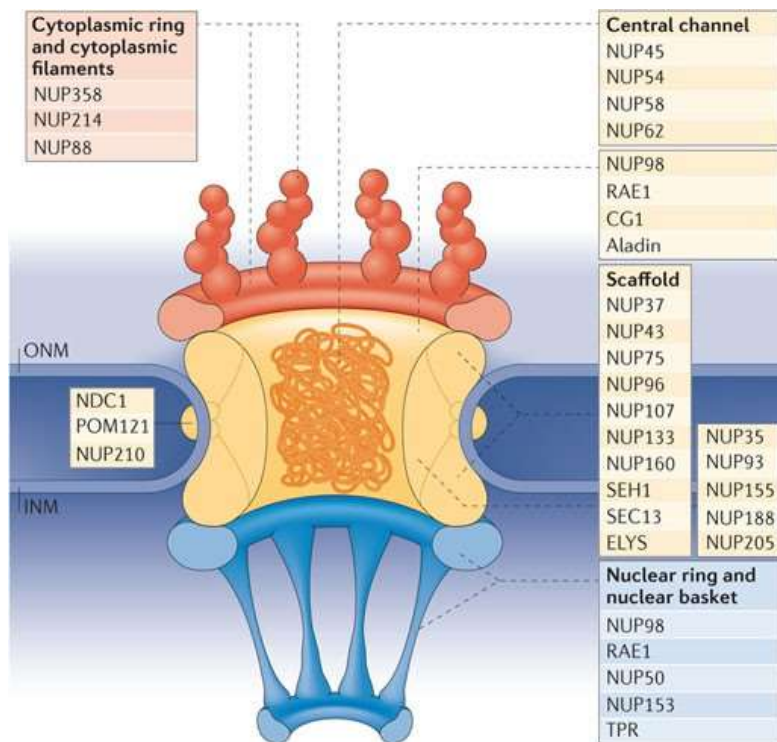


Figure 7.2. Correlation between the intrinsic disorder content and proteome size. Viruses, archaea, bacteria, and eukaryotes were studied, for a total of 3,484 species (Picture from Uversky VN, *Frontiers in Physics*, 2019, 7:10).

The study of IDPs or IDPRs can be hard and there is a common agreement with the fact that to obtain an IDRs structural description, experimental assays or computational ensemble approaches are needed [172].

## 7.2 Nucleoporin 98 and FG-repeats motif

The nuclear pore complex (NPC) is a very large regulatory channel formed by different proteins and embedded in the nuclear envelope of the eukaryotes. It primarily acts as a gatekeeper for the nucleus-cytoplasmic transport of macromolecules, but it is also involved in chromatin organization and regulation of gene transcription [175]. It is almost entirely made up of Nucleoporins (Nups), each is present in multiple copies. They are evolutionary conserved and 34 different Nups form the NPC, in which Nups are organized in sub-complexes: the outer transmembrane Nups, the intermediately positioned scaffold Nups and the inner Nups, forming the NPC cytoplasmic ring and filament, the central channel and the scaffold, and the nuclear ring and basket, respectively (Figure 7.3).



Nature Reviews | Molecular Cell Biology

Figure 7.3. Schematic illustration of the NPC structure and composition. ONM = Outer Nuclear Membrane and INM = Inner Nuclear Membrane (Picture from Raices M et al., *Nat Rev Mol Cell Biol*, 2012, Nov;13(11):687-99).

Several Nups, among which NUP98, contain phenylalanine-glycine (FG) repeats (FG-Nups) and they shuttle off the NPC to the nucleoplasm where, based on their interaction with chromatin modifiers and actively transcribed genes, participate in epigenetic and transcription regulation [176-178] and are involved in mitotic progression [179,180]. These functions are beyond the canonical function of nucleocytoplasmic transport. Moreover, in particular, NUP98 was found to be part of both the cytoplasmic and nuclear NPC sides. It represents an example of a peripheral NPC ambivalence component, able to come off the nuclear pore [181].

Structurally, NUP98, as well as other NUPs, is known to contain Phenylalanine-Glycine (FG) repeats, which are typically spaced by about 20 amino acids and represent an example of IDRs. FG motifs allow binding to different Nuclear Transporter Receptors (NTRs), regulating the bi-directional trafficking of proteins and RNA between cytoplasm and nucleus [182]. Overall, NUP98 is an extremely disordered protein, containing about 80% of IDRs by which two unstructured FG-repeats domains [183]. This feature confers to Nup 98 a high flexibility and a potential involvement in various protein-protein interaction networks (PPI) [184]. The first FG domain is located at the N-terminal of the protein (spanning residues 1-156) before a structured domain, which is followed by the second FG domain (spanning residues 214-480). As for any other FG-Nups, the NUP98 full-length structure is not yet available, nor is the FG-repeat domain structure.

Nevertheless, in the last decade, the number of disease-associated Nup genes has been rapidly expanding in a variety of human hereditary disorders, including developmental, neurodevelopmental, neurodegenerative, cardiac, and nephrotic syndromes [184], making it a protein of therapeutical interest.

## 8 Results

### 8.1 Identification of novel NUP98 pathogenic variants

A recent study from the Laboratory of Medical Cytogenetics and Human Molecular Genetics, Istituto Auxologico Italiano, has identified two mutations on the NUP98 gene: G28D and E875A, related to the Rothmund-Thomson syndrome, a very rare genetical disorder, implying fragile hair, absent eyelashes/eyebrows and bilateral cataract, mottled pigmentation, dental decay, hypogonadism, and osteoporosis.

Though both missense variants affect highly conserved amino acids, the higher pathogenicity score of the G28D variant and its location between Phenylalanine-Glycine (FG) repeats within the first NUP98 Intrinsically Disordered Region (IDR), made it suitable for being investigated by computational studies. The results, which will be presented below in the thesis, led to a manuscript that has been published: *Colombo EA, Valiante M, Uggeri M, Orro A, Majore S, Grammatico P, Gentilini D, Gervasini C, Finelli P, D'Ursi P, Larizza L. Germline biallelic Nucleoporin 98 variants in two siblings presenting a Rothmund-Thomson like spectrum: functional changes borne out by protein molecular modelling studies. Int. J. Mol. Sci. 2023, 24, 4028. <https://doi.org/10.3390/ijms24044028>.*

### 8.2 Molecular modeling study of NUP98 variant

A first analysis of the amino acid sequence of NUP98 showed that the G28D mutation was located near one of the FG repeats of the first FG domain (Figure 8.1).

```
WT      MFNKSFGTPEGGTGGFGTTSTFGQNTGFGTSSGGAFGTSAFGSSNNTGGLFGNSQTKPG 60
G28D    MFNKSFGTPEGGTGGFGTTSTFGQNTDFGTTSSGGAFGTSAFGSSNNTGGLFGNSQTKPG 60
*****

WT      GLFGTSSFSQPATSTSTGFGFGTSTGTANTLFGTASTGTSLFSSQNNAFAQNKPTGFGNF 120
G28D    GLFGTSSFSQPATSTSTGFGFGTSTGTANTLFGTASTGTSLFSSQNNAFAQNKPTGFGNF 120
*****

WT      GTSTSSGGLFGTTNTTNSNPFGSTSGSLFGPSSFTAA 156
G28D    GTSTSSGGLFGTTNTTNSNPFGSTSGSLFGPSSFTAA 156
*****
```

Figure 8.1. Sequence alignment between the WT and G28D NUP98 highlights the G28D mutation to be located before one of the FG repeats.

This domain is the N-terminal of the NUP98 protein, consisting of 156 amino acids (aa). The FG domain is structurally disordered, and it is characterized by several FG sequence motifs. The high dynamic structure of the domain excludes it from classical structural biology analyses such as X-ray crystallography; thus, bioinformatics approaches could be used to elucidate differences in the dynamic behaviour of the wild type and variant proteins by studying the ensemble structures. MD simulations were used to characterize the dynamic ensemble of structures of the FG domain proteins. Starting from a full-extended conformation of the FG domain (spanning residues 1-156), 20 replicas of MD simulations were performed for the wild type (WT) and the G28D variant, respectively. The structural diversity of each ensemble of biomolecular structures obtained along molecular dynamics simulations was analyzed by Root-Mean Square Deviation (RMSD) values. Starting from a fully extended conformation model, the system passed over to a new spatial arrangement which was stably maintained for 7 replicas in the wild type (replicas 2, 4, 7, 11, 16, 17, and 18) and 15 replicas in the G28D variant (replicas: 1, 2, 3, 5, 7, 8, 9, 10, 11, 13, 14, 15,16, 17, and 18) (Figure 8.2).

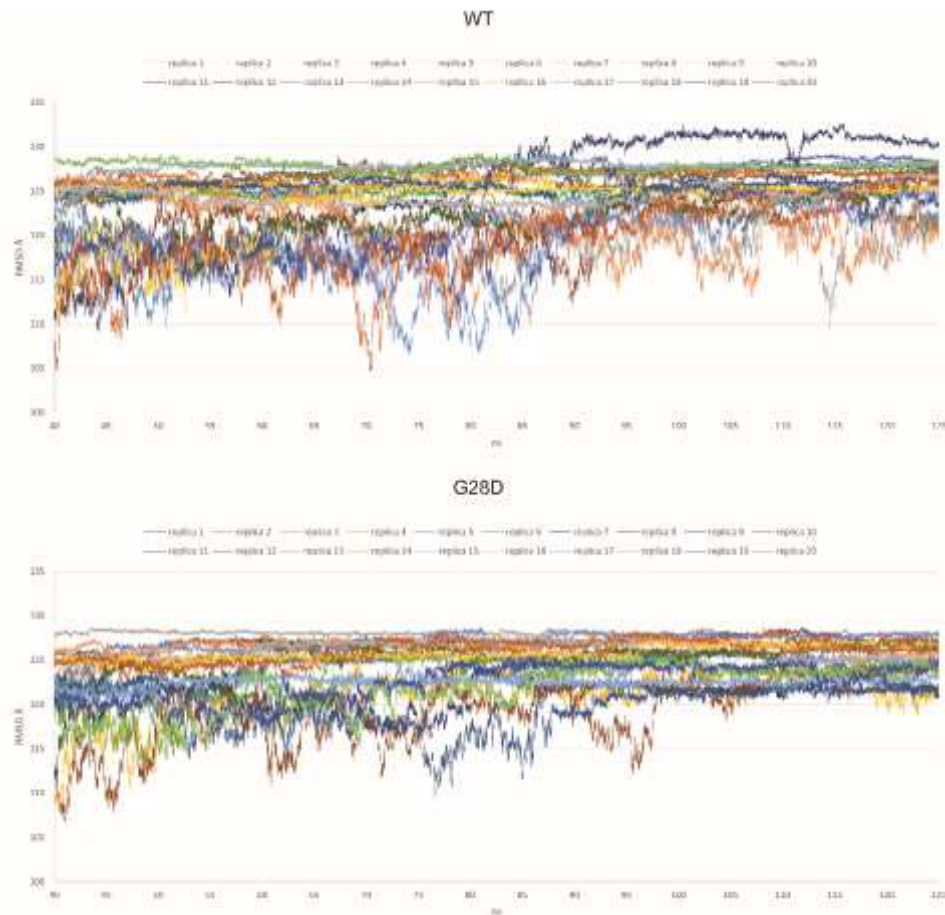


Figure 8.2. RMSD values from 40 ns of simulation to make easier the observation of the stable replicas.

Based on these first results, to further evaluate differences between the wild type and the pathogenic NUP98 variant, an evaluation of the representative conformations of WT and G28D, obtained from cluster analysis of the stable replicas, was carried out.

### 8.3 Compactness differences evaluation between the WT and the G28D

To verify differences between the WT and the G28D FG domains, the compaction level and the shape of WT and G28D were assessed by calculating the hydrodynamic radius, which can be determined by several biophysical experiments, such as dynamic light scattering experiments or it can be predicted from a computationally generated conformational ensemble of IDP, as in the present thesis. In fact, it is known that, depending on the overall composition of the amino acids of the IDP, their compaction may differ influencing the functional and the bio-physical properties of the protein [185].



Results showed an average hydrodynamics radius of 20,12 Å for the wild-type and 20,69 Å for the mutated FG domain (Figure 8.3), suggesting a different compaction for the two proteins. In particular, during the simulations, the initially fully extended conformations were shortened to globular conformation, in the case of the wild type, and were more elongated in the case of the G28D (Figure 8.4).

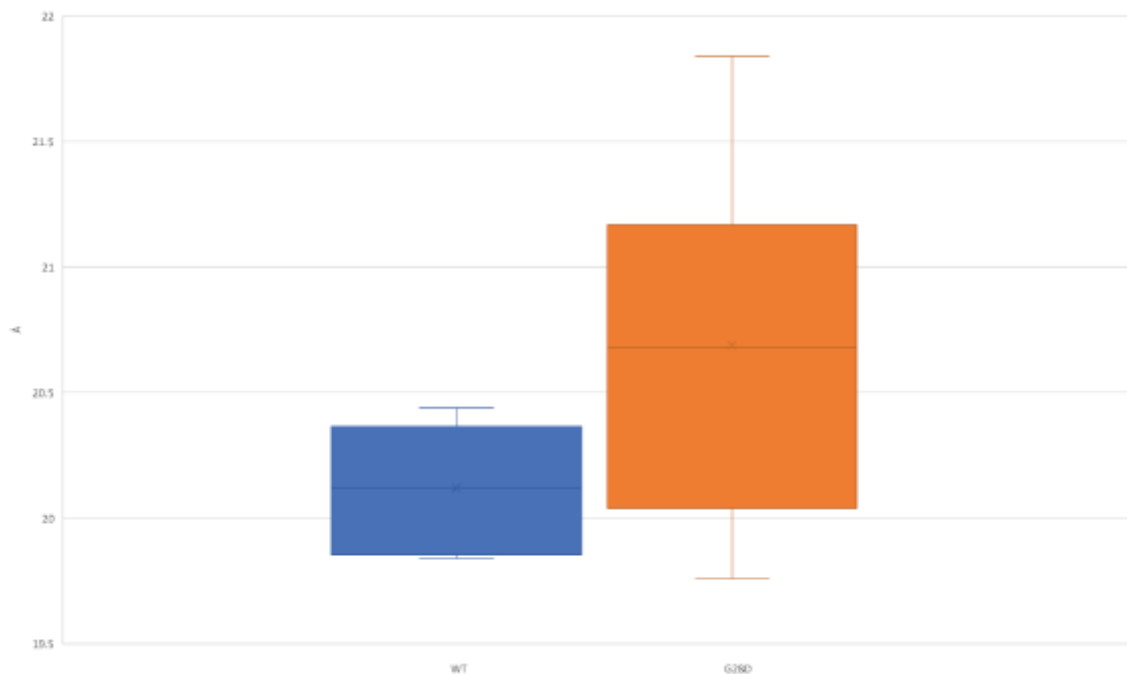


Figure 8.3. Boxplot of the average hydrodynamic radius of the WT and G28D FG domains calculated from the representative structures of stable replicas.

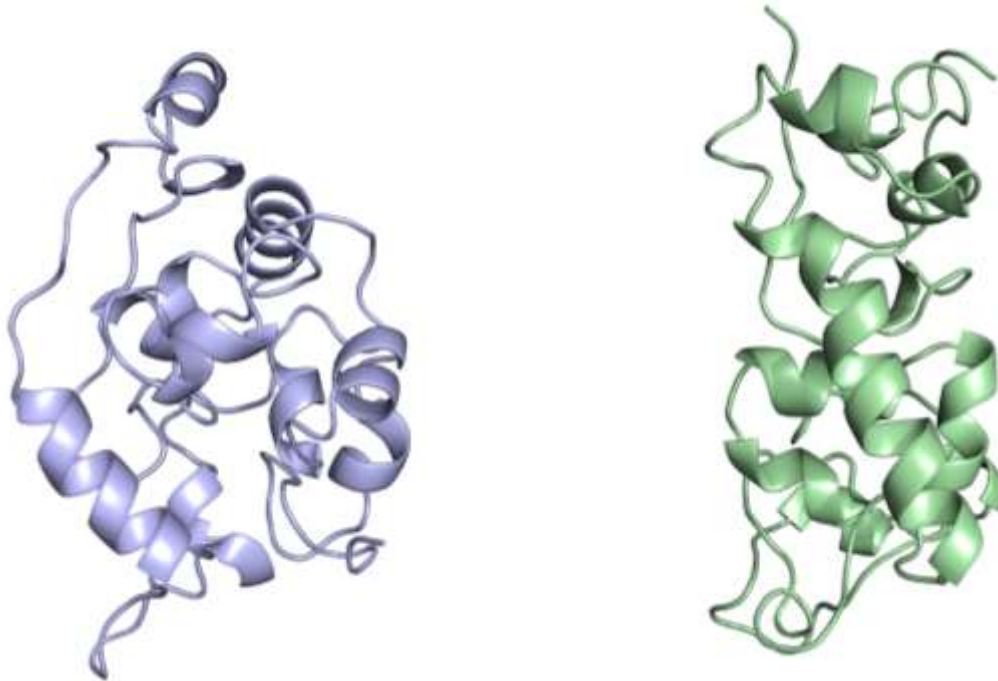


Figure 8.4. Differences in the shape of the WT- and G28D-NUP98 protein conformation. The WT-NUP98 (purple – left) showed a more globular conformation, while the G28D-NUP98 (green – right) a more elongated.

#### 8.4 G28D is less intramolecularly cohesive than WT NUP98

Since the G28D was found to be more expansive than the wild type, this provides a first indication that the G28D variant could be not as intramolecularly cohesive as the wild type, as shown by literature data from Krishnan *et al.*, according which phenylalanines in these FG motifs function as intramolecular cohesion elements imparting order to the FG domain and compacting its ensemble of structures in globular configurations [186]. To evaluate the intramolecular cohesion difference, the distances between the sites corresponding to the phenylalanine of the FG repeats (F-F pairs) were calculated and analyzed in the last 25 ns of each stable replica of the WT and the G28D. The average distance between the center of mass of phenylalanine residues belonging to the FG repeats is around 19 Å for the WT and 20.5 Å for the G28D (Figure 8.5A). Moreover, 23% of the G28D F-F pairs showed a significant variation in the distance (> 25%) in comparison to the WT (Figure 8.5B).



Figure 8.5. A) Distances between the centers of mass of the F-F pairs belonging to FG motifs. B) The G28D F-F pairs with a variation in distance > 25% compared to WT

To better describe the relationship between the FG motifs in the FG domain of NUP98, the distances of the F-F pairs illustrated in Figure 8.4 were used to generate a distance matrix which was subsequently used to construct the following network graph. The network graphic showed edges F-F pairs and edge weights are proportional to the distance value (Figure 8.6).

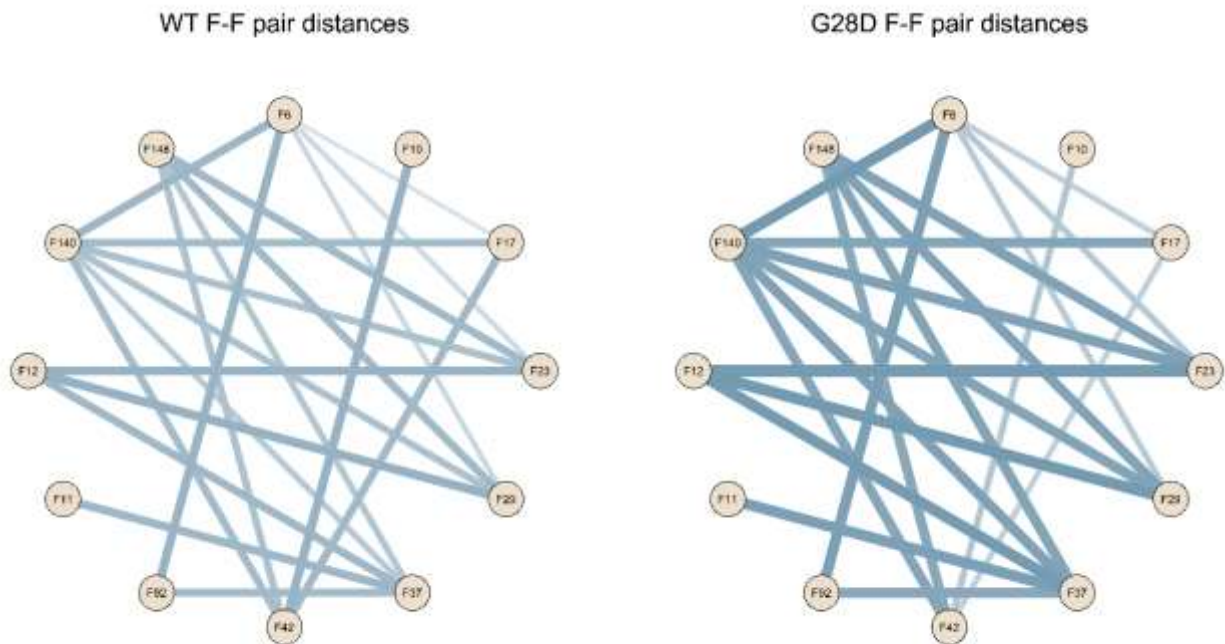


Figure 8.6. Schematic representation of F-F pair distances significant differences between the WT and the G28D variant

These findings highlighted that the intramolecular distances between phenylalanines of the FG repeats of the wild type and G28D differ quantitatively in the two proteins. In the wild-type, phenylalanines are generally closer to each other, while in G28D phenylalanine residues are more scattered.

To assess the effect of D28 on intramolecular cohesions, the solvent accessibility of this aspartic residue was calculated on the stable replicas of G28D by means of Solvent Accessible Surface Area (SASA) analysis (Figure 8.7).

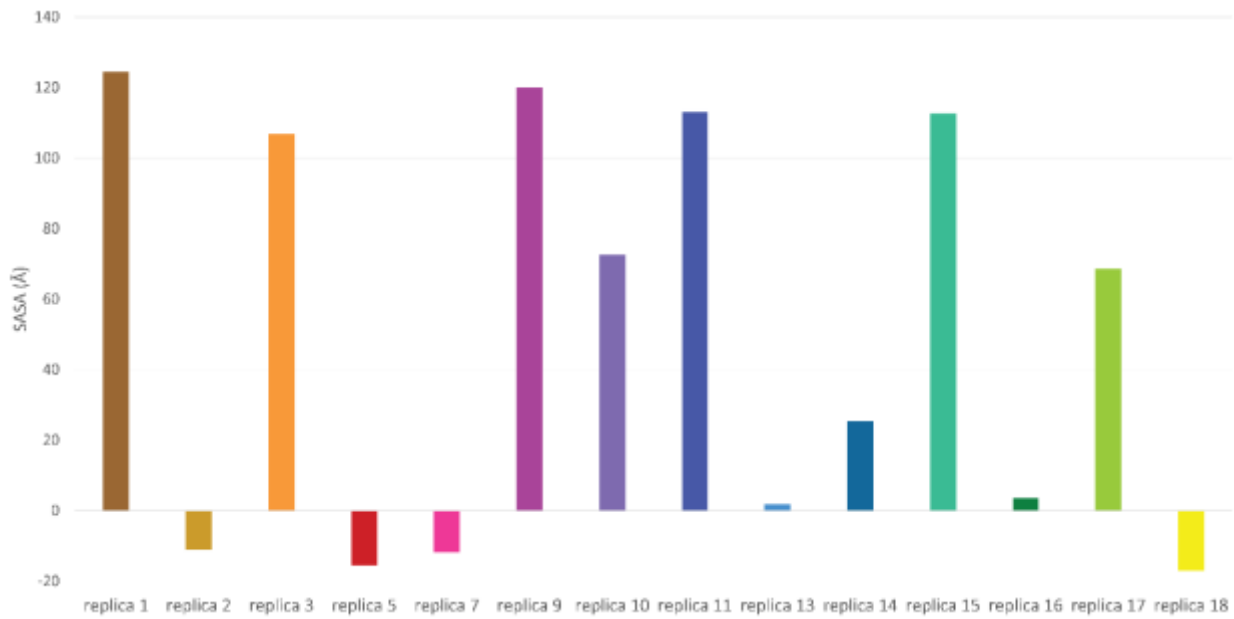


Figure 8.7. SASA analysis of D28 in the stable replicas showed two possible orientations: buried (SASA value near 0 below 0) or exposed (SASA positive value)

As shown in figure 8.8 the aspartic residue can be exposed to the solvent or can be buried within the FG domain.

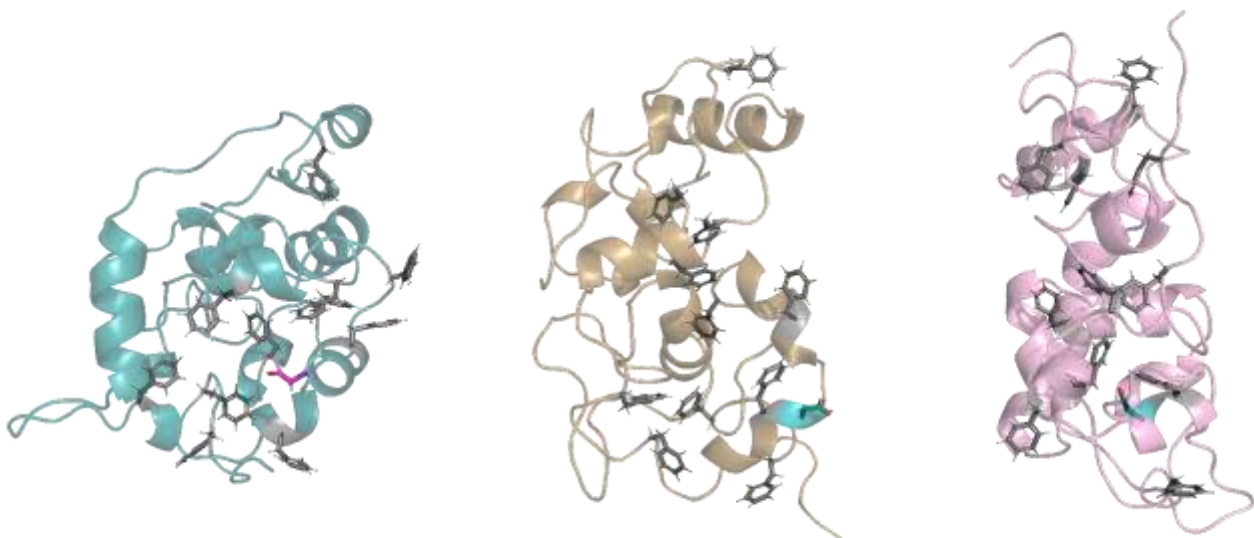


Figure 8.8. Exposed and buried conformations of G28D in comparison with the wild type. On the left, a conformation of the wild type (green ribbon), in the center, a conformation of the G28D with the D28 in the exposed conformation (orange ribbon), and on the right, a conformation of the G28D with the D28 in the buried conformation (pink ribbon). Phenylalanines of the FG motifs are shown in grey sticks, while the G28 and the D28 in blue and in cyan sticks, respectively.

Interestingly, in the buried conformation the D28 was found to make some common hydrogen bonds interaction between the replicas with specific residues of the protein: Gly30 residue in replicas 14 and 15, Thr31 in replicas 5, 14, and 15, Thr32 in replicas 7, 14, and 15, and Ser66 in replicas 2, 5, and 7.

### 8.5 RNA interaction impairment in the G28D variant

Milles *et al.* [187] showed that the N-terminal of NUP98 has the important functional role of multi-docking station for RNA, proteins, and self-assembly. In order to predict possible differences in the interaction with the RNA between the WT and the G28D protein, the Pprint webserver was used to evaluate the residues involved in the RNA interaction. As shown in Table 8.1, results highlighted that, in comparison to the WT, the G28D variant lost two residues predicted to interact with the RNA near the mutation: N26 and F37. Although this result suggests a minor ability of G28D to bind the RNA, further analyses are needed.

Table 8.1. Pprint analysis of the WT and G28D amino acid sequence.

WT				G28D			
Residue		SVM value	Prediction	Residue		SVM value	Prediction
1	M	-0.55378945	Non-Interacting	1	M	-0.55803765	Non-Interacting
2	F	-0.82641875	Non-Interacting	2	F	-0.87349547	Non-Interacting
3	N	-0.90220625	Non-Interacting	3	N	-0.959148	Non-Interacting
4	K	-0.61748008	Non-Interacting	4	K	-0.6271972	Non-Interacting
5	S	-0.60904646	Non-Interacting	5	S	-0.61693582	Non-Interacting
6	F	-0.3663883	Non-Interacting	6	F	-0.33438624	Non-Interacting
7	G	-0.86719585	Non-Interacting	7	G	-0.91670753	Non-Interacting
8	T	-0.15408726	Interacting	8	T	-0.08689689	Interacting

9	P	-0.38805357	Non-Interacting	9	P	-0.35942944	Non-Interacting
10	F	0.052464119	Interacting	10	F	0.15329547	Interacting
11	G	0.061878645	Interacting	11	G	0.16367349	Interacting
12	G	0.44208841	Interacting	12	G	0.6041482	Interacting
13	G	0.092849556	Interacting	13	G	0.19648674	Interacting
14	T	0.31123103	Interacting	14	T	0.45172504	Interacting
15	G	0.15037959	Interacting	15	G	0.26181989	Interacting
16	G	0.3290125	Interacting	16	G	0.47144538	Interacting
17	F	-0.16312306	Interacting	17	F	-0.07310067	Interacting
18	G	-0.52772285	Non-Interacting	18	G	-0.50382202	Non-Interacting
19	T	-0.54276185	Non-Interacting	19	T	-0.54183776	Non-Interacting
20	T	-0.2921471	Non-Interacting	20	T	-0.25404084	Non-Interacting
21	S	-0.67593245	Non-Interacting	21	S	-0.63787914	Non-Interacting
22	T	-0.53121105	Non-Interacting	22	T	-0.49591728	Non-Interacting
23	F	-0.52116954	Non-Interacting	23	F	-0.51303254	Non-Interacting
24	G	-0.96787691	Non-Interacting	24	G	-1.0327101	Non-Interacting
25	Q	-0.33007003	Non-Interacting	25	Q	-0.35700793	Non-Interacting
26	N	-0.09025613	Interacting	26	N	-0.22664901	Non-Interacting
27	T	0.20832611	Interacting	27	T	-0.09369108	Interacting
28	G	-0.30593648	Non-Interacting	28	D	-0.56705844	Non-Interacting
29	F	0.052686027	Interacting	29	F	-0.09500207	Interacting
30	G	-0.59153605	Non-Interacting	30	G	-0.60180787	Non-Interacting
31	T	0.19650102	Interacting	31	T	0.052826422	Interacting
32	T	-0.0532804	Interacting	32	T	-0.16177192	Interacting
33	S	-0.17722904	Interacting	33	S	-0.17024124	Interacting
34	G	-0.00635643	Interacting	34	G	-0.00468534	Interacting
35	G	-0.19332631	Interacting	35	G	-0.19660448	Interacting
36	A	-0.18253653	Interacting	36	A	-0.17248359	Interacting
37	F	-0.19816187	Interacting	37	F	-0.20855071	Non-Interacting
38	G	-0.84822047	Non-Interacting	38	G	-0.83480983	Non-Interacting
39	T	-0.41467499	Non-Interacting	39	T	-0.41739822	Non-Interacting
40	S	-0.28552534	Non-Interacting	40	S	-0.2767744	Non-Interacting

41	A	-0.32145082	Non-Interacting	41	A	-0.34539552	Non-Interacting
42	F	-0.35849272	Non-Interacting	42	F	-0.34433301	Non-Interacting
43	G	-0.24275524	Non-Interacting	43	G	-0.24665706	Non-Interacting
44	S	-0.38951034	Non-Interacting	44	S	-0.40358287	Non-Interacting
45	S	-0.70910621	Non-Interacting	45	S	-0.69937712	Non-Interacting
46	N	-0.41637418	Non-Interacting	46	N	-0.40777296	Non-Interacting
47	N	0.090789887	Interacting	47	N	0.092923449	Interacting
48	T	0.27514296	Interacting	48	T	0.28303749	Interacting
49	G	0.11327117	Interacting	49	G	0.11327132	Interacting
50	G	0.65111053	Interacting	50	G	0.6350224	Interacting
51	L	-0.02866753	Interacting	51	L	-0.04015428	Interacting
52	F	-0.25792442	Non-Interacting	52	F	-0.25522212	Non-Interacting
53	G	-1.0131399	Non-Interacting	53	G	-0.99196158	Non-Interacting
54	N	0.22322956	Interacting	54	N	0.21665862	Interacting
55	S	-0.21375527	Non-Interacting	55	S	-0.22446886	Non-Interacting
56	Q	-0.09758618	Interacting	56	Q	-0.11153087	Interacting
57	T	0.067543326	Interacting	57	T	0.047492371	Interacting
58	K	0.20446247	Interacting	58	K	0.18253409	Interacting
59	P	0.22663635	Interacting	59	P	0.20503075	Interacting
60	G	-0.11549157	Interacting	60	G	-0.14294754	Interacting
61	G	-0.28683165	Non-Interacting	61	G	-0.31742948	Non-Interacting
62	L	-0.33953353	Non-Interacting	62	L	-0.35158283	Non-Interacting
63	F	-0.73593201	Non-Interacting	63	F	-0.7412335	Non-Interacting
64	G	-0.74425387	Non-Interacting	64	G	-0.73547431	Non-Interacting
65	T	-0.78060069	Non-Interacting	65	T	-0.76434944	Non-Interacting
66	S	-1.1776457	Non-Interacting	66	S	-1.1687718	Non-Interacting
67	S	-0.96767276	Non-Interacting	67	S	-0.9390548	Non-Interacting
68	F	-1.0349774	Non-Interacting	68	F	-0.98125493	Non-Interacting
69	S	-1.0519283	Non-Interacting	69	S	-1.0413242	Non-Interacting
70	Q	-1.0165043	Non-Interacting	70	Q	-1.0056189	Non-Interacting
71	P	-1.0301084	Non-Interacting	71	P	-1.0140502	Non-Interacting
72	A	-0.95427477	Non-Interacting	72	A	-0.9495638	Non-Interacting



73	T	-1.0784161	Non-Interacting	73	T	-1.0786419	Non-Interacting
74	S	-0.95254672	Non-Interacting	74	S	-0.92545033	Non-Interacting
75	T	-0.9077265	Non-Interacting	75	T	-0.9022996	Non-Interacting
76	S	-0.56774113	Non-Interacting	76	S	-0.58594226	Non-Interacting
77	T	-0.62858207	Non-Interacting	77	T	-0.63028449	Non-Interacting
78	G	-0.47670062	Non-Interacting	78	G	-0.48366087	Non-Interacting
79	F	-0.08589069	Interacting	79	F	-0.10186104	Interacting
80	G	-0.50735971	Non-Interacting	80	G	-0.50449518	Non-Interacting
81	F	0.12438119	Interacting	81	F	0.1050222	Interacting
82	G	-0.38888995	Non-Interacting	82	G	-0.39747944	Non-Interacting
83	T	0.075502316	Interacting	83	T	0.041347035	Interacting
84	S	-0.17303079	Interacting	84	S	-0.19705206	Interacting
85	T	0.26220522	Interacting	85	T	0.21738126	Interacting
86	G	-0.03439857	Interacting	86	G	-0.06533779	Interacting
87	T	0.041335968	Interacting	87	T	-0.00559414	Interacting
88	A	-0.23688733	Non-Interacting	88	A	-0.29386109	Non-Interacting
89	N	0.05820687	Interacting	89	N	-0.01796369	Interacting
90	T	-0.24636999	Non-Interacting	90	T	-0.29849051	Non-Interacting
91	L	-0.41515678	Non-Interacting	91	L	-0.44493595	Non-Interacting
92	F	-0.75630273	Non-Interacting	92	F	-0.78610236	Non-Interacting
93	G	-0.51913145	Non-Interacting	93	G	-0.57522225	Non-Interacting
94	T	-0.57324359	Non-Interacting	94	T	-0.58675231	Non-Interacting
95	A	-0.90791421	Non-Interacting	95	A	-0.92717602	Non-Interacting
96	S	-0.56579222	Non-Interacting	96	S	-0.585956	Non-Interacting
97	T	-0.35803433	Non-Interacting	97	T	-0.38600291	Non-Interacting
98	G	-0.14025052	Interacting	98	G	-0.18097175	Interacting
99	T	-0.22947986	Non-Interacting	99	T	-0.24555298	Non-Interacting
100	S	-0.25903639	Non-Interacting	100	S	-0.26939633	Non-Interacting
101	L	-0.72867269	Non-Interacting	101	L	-0.7337462	Non-Interacting
102	F	-1.5055101	Non-Interacting	102	F	-1.4852729	Non-Interacting
103	S	-0.95047824	Non-Interacting	103	S	-0.93565842	Non-Interacting
104	S	-1.1651547	Non-Interacting	104	S	-1.1503788	Non-Interacting

105	Q	-0.52539393	Non-Interacting	105	Q	-0.52596819	Non-Interacting
106	N	-0.99015616	Non-Interacting	106	N	-0.98089022	Non-Interacting
107	N	-0.95779524	Non-Interacting	107	N	-0.94850312	Non-Interacting
108	A	-0.37984574	Non-Interacting	108	A	-0.38414445	Non-Interacting
109	F	-0.83849862	Non-Interacting	109	F	-0.83273493	Non-Interacting
110	A	-1.5557231	Non-Interacting	110	A	-1.5318974	Non-Interacting
111	Q	-0.8784608	Non-Interacting	111	Q	-0.87109706	Non-Interacting
112	N	-0.46131501	Non-Interacting	112	N	-0.462866	Non-Interacting
113	K	0.20884518	Interacting	113	K	0.18904333	Interacting
114	P	-0.23128636	Non-Interacting	114	P	-0.24036783	Non-Interacting
115	T	-0.2096631	Non-Interacting	115	T	-0.21949332	Non-Interacting
116	G	-0.15327384	Interacting	116	G	-0.16578721	Interacting
117	F	0.31626244	Interacting	117	F	0.29118409	Interacting
118	G	0.32724456	Interacting	118	G	0.30207997	Interacting
119	N	0.61000023	Interacting	119	N	0.57703863	Interacting
120	F	0.21606952	Interacting	120	F	0.19342053	Interacting
121	G	0.49218171	Interacting	121	G	0.46270835	Interacting
122	T	-0.07417111	Interacting	122	T	-0.08789188	Interacting
123	S	-0.28993514	Non-Interacting	123	S	-0.29757753	Non-Interacting
124	T	-0.64924421	Non-Interacting	124	T	-0.64966924	Non-Interacting
125	S	0.015272848	Interacting	125	S	-0.00252083	Interacting
126	S	-0.1431344	Interacting	126	S	-0.15467015	Interacting
127	G	0.012886299	Interacting	127	G	-0.00571911	Interacting
128	G	0.14138924	Interacting	128	G	0.11896952	Interacting
129	L	-0.49934315	Non-Interacting	129	L	-0.50369941	Non-Interacting
130	F	-0.23959612	Non-Interacting	130	F	-0.24987891	Non-Interacting
131	G	-0.35825414	Non-Interacting	131	G	-0.36702305	Non-Interacting
132	T	-0.72797192	Non-Interacting	132	T	-0.72785549	Non-Interacting
133	T	-0.6959551	Non-Interacting	133	T	-0.69549842	Non-Interacting
134	N	-0.44710493	Non-Interacting	134	N	-0.476782	Non-Interacting
135	T	-0.20438086	Non-Interacting	135	T	-0.24391674	Non-Interacting
136	T	-0.25912753	Non-Interacting	136	T	-0.28821713	Non-Interacting

137	S	-0.34135313	Non-Interacting	137	S	-0.33713262	Non-Interacting
138	N	0.14155492	Interacting	138	N	0.091506915	Interacting
139	P	-0.14956362	Interacting	139	P	-0.18465374	Interacting
140	F	-0.29269783	Non-Interacting	140	F	-0.28638717	Non-Interacting
141	G	-0.82672954	Non-Interacting	141	G	-0.81942059	Non-Interacting
142	S	-0.38685073	Non-Interacting	142	S	-0.42436374	Non-Interacting
143	T	-0.79021587	Non-Interacting	143	T	-0.80983266	Non-Interacting
144	S	-0.55153777	Non-Interacting	144	S	-0.58193886	Non-Interacting
145	G	-0.39222561	Non-Interacting	145	G	-0.43481243	Non-Interacting
146	S	-0.62833599	Non-Interacting	146	S	-0.66485185	Non-Interacting
147	L	-0.47007565	Non-Interacting	147	L	-0.48549155	Non-Interacting
148	F	-0.85006116	Non-Interacting	148	F	-0.85578051	Non-Interacting
149	G	-0.82726123	Non-Interacting	149	G	-0.82407064	Non-Interacting
150	P	-0.65474048	Non-Interacting	150	P	-0.64809849	Non-Interacting
151	S	-0.85642496	Non-Interacting	151	S	-0.84881163	Non-Interacting
152	S	-0.50571339	Non-Interacting	152	S	-0.50765667	Non-Interacting
153	F	-0.45379119	Non-Interacting	153	F	-0.45673466	Non-Interacting
154	T	-0.40774018	Non-Interacting	154	T	-0.41028561	Non-Interacting
155	A	-0.6588126	Non-Interacting	155	A	-0.65610309	Non-Interacting
156	A	-0.6588126	Non-Interacting	156	A	-0.6588126	Non-Interacting

## 8.6 Discussion and conclusions

Mutations of NUP98 have been so far investigated in cancer, mainly hematological malignancies, with the major focus on the expression of the NUP98 chimeric allele, while NUP98 depletion and the presence of only one copy of the WT allele has been overlooked [188]. As regards germline pathogenic variants, only one case of heterozygous mutation has been observed in a woman in which chromosome 11 breakpoint disrupts the NUP98 gene [189] without causing any relevant phenotype as the carrier had a normal phenotype and came to clinical observation due to bilateral renal angiomyolipoma. the G28D NUP98 variation above presented, causes a likely pathogenic variant in two siblings with a clinical presentation reminiscent of Rothmund-Thomson syndrome. The variant has been computationally studied in comparison to the WT NUP98. Differences in the dynamic behavior between

the two proteins were first highlighted by the RMSD analysis. Results evidenced how WT and G28D could explore different conformational spaces during their simulations. In particular, the mutated FG domain reached a minimum energy conformation easier than the wild type (15 vs 7 replicas), giving a first indication that the G28D substitution could facilitate this process. Moreover, the study of protein compactness showed interesting differences between variant and WT proteins. G28D was found to be less globular than the WT and this difference is in agreement with the different behaviour highlighted by the RMSD analysis. Furthermore, it is known that, depending on the overall composition of the amino acids of the IDPs, their compaction may differ, influencing the functional and the bio-physical properties of the protein [185]. So, based on these data, it was possible to hypothesize different functional properties for WT and G28D NUP98. With the aim to study at a deeper level these differences, the FG repeats, an important feature of the NUP98, were also investigated. Krishnan *et al.* [186] showed that phenylalanines in these FG motifs function as intramolecular cohesion elements imparting order to the FG domain and compacting its ensemble of structures in globular configurations. In agreement, the results presented above showed that the F-F pairs of the G28D are less close in comparison to the F-F pairs of the WT, once again highlighting the differences between the two NUP98 proteins.

Then, to further understand how the G28D substitution could affect this different dynamic behaviour, a SASA analysis was carried out finding two possible orientations of the D28 side chain. The mutated residue can be buried inside the protein or exposed. When D28 is buried, it induces an arrangement around itself, driving some protein residues and FG repeats close to it. This arrangement prevented the appropriate interactions between the FG repeats, thus influencing the folding of the FG domain. on the contrary, when D28 is exposed, it does not allow a proper arrangement of the FG motifs around itself, because of the chemical-physical properties of its side chain (steric hindrance and negative charge of the carboxyl group). Results obtained, highlighted that, in both possible conformations, D28 led to a moving away of the intramolecular cohesion elements and a more elongated conformation of the mutant in comparison to the wild type.

Finally, also an impairment regarding the RNA interaction was predicted.

In conclusion, it is possible to hypothesize that the G28D mutation may not allow the N-terminal of NUP98 to be as dynamic as it should be, leading to protein misfunction. The diminished structural

coherence of the NUP98 mutant FG domain could alter its interactivity network impacting both the canonical and the off-pore functions of this multitasking nucleoporin.

## 9 Materials and Methods

### 9.1. Molecular modelling simulation and analysis

The NUP98 wild type amino acid sequence spanning residues 1-156 was retrieved from UniProt: Q9HDC8 (The UniProt Consortium, 2021), and tleap software was used to create a fully extended 3D structure of the wild type and G28D mutation [165]. Then, 20 replicas of 125 ns each of molecular dynamic simulation in gas phase were carried out for both the WT and the G28D using Amber18. The systems were minimized in four steps, for a total of 5000 steps of steepest descent and 5000 steps of conjugate gradient with decreasing restraint, from 5 Kcal/mol to 0 Kcal/mol. Subsequently, starting from the minimized structures, 20 independent replicas with different seeds of 125 ns each were carried out on both systems, without any restraint. The ff14SB forcefield was used for the simulation, with a salt concentration of 0.15 mM. The MD trajectories analyses: RMSD, F-F pair distances, and SASA were carried out using the Ambertools package. The F-F pairs network graph was built using the RStudio software [190]. The hydrodynamic radiuses were calculated using the HullRad webserver [191] and the residues interacting with the RNA have been predicted using Pprint webserver with default settings [192].

## Summary

### 10. Methodologies Details

#### 10.1 Homology Modeling

Homology Modeling (HM) is a computational method used to predict the 3D structure of a target protein based on a homologous template protein with a known structure (X-ray or Cryo-EM). HM works by taking advantage of two observations. The first is that the amino acid sequence of a protein determines its 3D structure and second the structures of the proteins are conserved. So, proteins with similar sequences often have similar structures [193]. By using a protein with a known 3D structure as a template, it is possible to predict the structure of a related protein that has not been yet experimentally determined. It is possible to summarize the generation of a homology modeling in the following steps:

- Identification and selection of a related homologous template protein with a known 3D structure
- Amino acid sequence alignment between the target and the template protein to determine the conserved regions
- Building of the model
- Refinement and validation of the model

The identification of the template protein is usually done after a first alignment to determine the sequence similarity with the target protein. The 3D structure of the template is retrieved from the PDB database [194] selecting the best structure between the ones available. Then, a second and more accurate amino acid alignment between the template and the target protein can both be directly performed by the homology modeling program or carried out and manually checked by the user. The latter choice is always recommended to extensively check the alignment based on the knowledge of the template 3D protein structure. Successively, the model is built. There are different approaches to modeling a protein: rigid body assembly, segment matching, spatial restraint, and artificial evolution [193]. The rigid method is often used when the target protein has a high similarity to the template. It uses the template structures as is to do the homology modeling, with only small adjustments to the position and orientations of residues to avoid huge clashes. SwissModel is an example of a rigid body approach and one the most used software [195]. In segment matching, multiple templates are used to

generate a cluster of atomic positions which are used as a starting point. The target protein is then modeled, segment by segment based on the previously obtained atomic positions. The spatial restraint method is a more flexible and robust approach adopted by MODELLER [196]. It builds the model by using the stereochemical restraints on bond lengths, bond angles, dihedral angles, and van der Waals contact distances which came from the template structure/s. Finally, the artificial evolution method combines the rigid-body assembly method and stepwise template evolutionary mutations to fit the target protein on the template. In general, with a low amino acid similarity between the template and the target protein the use of a more flexible modeling approach with multiple templates is common to obtain a robust model. There are also other methods to predict the protein structures, such as the ab initio modeling, but they are more computationally demanding and used when the study is more focused on unstructured proteins or proteins without a homologous template with a known 3D structure. Finally, the model can be refined by energy minimization or short molecular dynamic simulations and its quality is then assessed. It can be done in different ways, like checking parameters such as torsion and rotational angles, and bond length or checking the Ramachandran plot, or, again, focusing on the determination of the spatial features of the model [193].

## **10.2 Molecular Docking**

Molecular docking is a computational method mainly used to predict the binding mode and the affinity of a small molecule to a target protein at the atomic level, but it can be also used to predict the binding between two proteins (protein-protein docking) or a small molecule/protein to the DNA/RNA. Based on the knowledge of the protein the docking experiment can be very specific, targeting a precise region of the protein (enzyme active site) or can be less defined, targeting the whole protein when the binding site is not known (blind docking). Generally, knowing the location of the binding site significantly increases the docking efficiency [197]. However, in the absence of a known binding site, there are software able to identify the most druggable pockets, such as Fpocket [198]. The docking experiment is basically performed in two steps:

- Sampling possible conformations of the ligand inside the grid which encloses the active site or the part of the interest of the protein
- Ranking the obtained conformations using a scoring function



The sampling of the possible conformations of the ligand is explored by different algorithms, among which, the currently more used are matching algorithms, incremental construction, the Monte Carlo, and genetic algorithms [199]. The matching algorithms are very fast and map the ligand into the pocket of the protein based on the molecular shape features and chemical information, representing the protein and the ligands as pharmacophores. So, each pharmacophore distance is calculated for a match and new ligand conformations are obtained by the distance matrix between the pharmacophore and the corresponding ligand atoms. The incremental construction divides the ligand into several fragments, breaking its rotatable bonds and then docking to the protein the largest fragment. Successively, the remaining fragments are incrementally added to the ligand in different orientations. Instead, the Monte Carlo generates ligand poses by rotating the ligand bond and by rigid-body translation or rotation and then testing the obtained conformation by energy-based criteria. If criteria are passed the pose is saved and a new conformation is generated and checked, until the reach of the defined quantity of conformations. Finally, genetic algorithms treat the degree of freedom of the ligand as “genes” which form a “chromosome”, that can be visualized as the pose of the ligand. Then, modifications to genes and chromosomes are applied. Mutations insert random changes into the genes, while crossover exchanges genes between two chromosomes. In this way, new ligand structures can be generated and used for the next ligand pose generations after a positive check is assessed by the scoring function [200]. After the generation of the ligand poses inside the binding pocket of the protein, the second step of the docking experiment is the calculation of their binding affinity, and it is evaluated by the scoring function. As for the sampling, each docking software uses a different scoring function, but they can be divided into four types: the physics-based, the empirical, the knowledge-based, and the one which uses machine learning [200]. The first type of scoring function uses the force field to assess the binding energy, calculating the sum of the non-bonded interactions, both electrostatic and van der Waals, and desolvation energy. Due to the computational cost, entropy and solvation terms are usually not calculated. These scoring functions are based on experimental data or *ab initio* quantum mechanical calculations. On the other side, the empirical scoring functions decompose the binding energy into several components, then each component is multiplied by a coefficient obtained from regression analysis and fitted into a high-quality test set of ligand-protein complexes, for which it is noted the binding affinity. Finally, the results are summed up obtaining the docking score. Another type

of scoring function is the knowledge-based. They evaluate the docking poses extracting information from the known crystal structures of protein-ligand complexes. The features of the binding between the ligand and the protein are retrieved from the structures and statistical analysis via the inverse Boltzmann distribution is performed. Finally, in the last few years machine learning is often time used to predict docking affinity. These types of functions do not assume a predetermined functional form, so they can implicitly capture intermolecular interactions that are hard to model explicitly, and results showed that they are able to overcome the older scoring functions methods [200].

Besides how docking software works, should also be mentioned the different types of docking experiments that can be performed. In fact, a docking experiment can be rigid, semi-flexible, or flexible. Rigid docking is less computationally demanding, but it is also less accurate. In this type of docking both the ligand and the protein are kept rigid, without any degree of freedom. Instead, in semi-flexible docking, the ligand is able to explore its possible spatial conformations, while the protein is kept rigid. Then, the last docking experiment is flexible docking. This is the most computationally demanding in which both the ligand and the protein are able to move, although the protein flexibility is mainly limited to the residues side chains. However, molecular dynamic simulation is one of the most reliable approaches to evaluate the intrinsic protein flexibility and to analyze the fitting induced by the binding of the ligand to the protein [201].

### **10.3 Molecular Dynamic Simulation**

Molecular Dynamic (MD) simulation is an advanced computational method able to explore the conformational spaces of the protein and other biomolecules, capturing their behavior in full atomic detail and simulating the cellular environment. With the MD simulation is possible to study and evaluate the position and the motion of each atom in the system at every point in time. MD is indeed a very powerful tool. However, since the computational demand for a quantum mechanic simulation at protein full scale is not yet available, the MD simulation simplifies the system using the Newtonian laws of motion treating the atoms as spheres connected by springs. Nevertheless, this simplification was extensively demonstrated during the years to be accurate enough to simulate the behavior of proteins or other biomolecules [202]. To simulate the system different force fields, which are derived from the results of quantum mechanical calculations and certain experimental measurements, are used. The

force field includes all the parameters to describe the system, the bonded and non-bonded interactions, and the inter- and intra-molecular forces. It is a collection of equations and associated constants used to reproduce molecular geometry and properties. Every biomolecule (i.e. protein, lipid membrane, DNA, RNA...etc) requires a different force field to be simulated at its best [203]. The MD algorithm uses the force field to obtain the net force and the acceleration of each atom at a certain position, which is treated as a point with a specific mass and fixed charge. This calculation allows to find how each atom will react and interact in a specific time-step (usually 1 fs) then allowing the creation of novel atom coordinates that, time-step after time-step, allows the simulation of the whole system.

The MD simulation can be divided into four different steps:

- Minimization
- Heating
- Equilibration
- Production

In the first step, the atoms of the system are minimized allowing them to reach the minimum energy conformation without giving any external energy (atoms velocity) to the system, setting the temperature at absolute zero. Then the system is gradually heated from 0 to the selected temperature (usually 300 K). This step gives gradually more energy to the atoms, allowing them to start moving and interacting with each other. Heating is usually done at constant volume, setting the pressure a 1 atm. Then, during the equilibration, the system is maintained at the reached temperature and switched to a constant pressure environment leaving the volume of the system free to expand. During the heating and the equilibration restraints/constraints can be added to the system to better stabilize and equilibrate it. Finally, during the production the atoms can fully explore their conformational spaces.

Different MD simulations can be used for different systems concerning what the goal is. The first distinction is the use of a gas-phase MD or a solvent-phase MD, in which the system is simulated in the absence or presence (implicit or explicit) of water or other solvents. The gas-phase MD is less computationally demanding, but unless specific cases such as protein folding, the solvent-phase MD is always recommended. It is also possible to carry out an all-atom simulation or simplify the system, mainly reducing the complexity of the side chain of the protein residues, performing a coarse-grained simulation. The coarse-grained is usually used when the system is too big, when a further increase of

the simulation time is needed, or when the aim is the observation of conformational changes of the protein. There are also other simulations, defined as “enhanced sampling methods”, which are considered advanced MD. They have been developed to facilitate the understanding of complex mechanisms regulating biological processes overcoming energy barriers [204]. The more used methods are the Replica-Exchange MD (REMD), the Accelerated MD (aMD), and the Metadynamic (MTD). In REMD, multiple simulations of the system are run in parallel at different temperatures and exchanged at a fixed interval to enhance the sampling of the conformational spaces of the protein. In aMD and MTD, the potential energy function is directly modified to speed up the transition of the biomolecule from local minimum, flattening the energy barrier [205]. aMD was used in the present thesis and it allows to reduce energy barriers which separate the different states of a system. aMD modifies the potential energy landscape raising energy wells so allowing an enhanced sampling of the conformational space, otherwise not accessible in cMD, still converging to the correct canonical distribution [205]. Finally, it is worth mentioning that the classical or advanced MDs are not able to simulate the formation or breakage of covalent bonds, and for this specific case, the use of Quantum mechanics/molecular mechanics (QM/MM) simulation is required [202].

## 11. Conclusions and Future Prospectives

In the present thesis, different computational methods have been successfully applied, examining a spectrum of proteins of therapeutic interest with a bunch of different features, functions, and interactions. *In-silico* methodologies have been applied in all their aspects, starting from the building of an in-house manually curated ligand dataset ready to use for the repositioning of already approved drugs by docking experiments, moving to the use of classical and advanced molecular dynamic simulations, and the development of specific computational pipelines. Membrane protein, protein-ligand and protein-protein interactions, and intrinsically disordered protein are the case studies that have been evaluated throughout my three-year PhD. The flexibility of computational methods allowed to virtually study all the proteins and their interactions here presented, highlighted, and discussed.

The study of the F508del-CFTR led to the development of two computational pipelines that were validated using a multidisciplinary approach by both SPR and cell-based assay. The two pipelines have extensively studied the binding pocket 1 of the F508del-CFTR to respectively repurpose already approved drugs for protein rescue and small molecules able to synergize together with lumacaftor. The first pipeline allowed the identification of rutin, a natural compound with a binding affinity for the F508del-CFTR similar to that of lumacaftor. Rutin is able to compete with lumacaftor for the binding to F508del-CFTR suggesting the same binding site, while quercetin, the molecule derived from the metabolism of rutin in the human gut, showed an even higher binding affinity for the F508del-CFTR, without competing with lumacaftor. These findings highlighted the possibility that a small molecule could be able to synergize with lumacaftor in the same binding pocket, enhancing the rescue ability of the already approved drug. The second pipeline was indeed focalized on that task, leading to the finding of nicotinamide as a small molecule able to bind the F508del-CFTR and increase its rescue. These pipelines could be further implemented with innovative features in both the screening of the ligand and the speed-up of the molecular dynamics simulations to better evaluate more efficient and safe drugs able to help the rescue of the F508del-CFTR, and its others known mutations.

Moreover, the thesis has also focused on the evaluation of two other proteins of therapeutic interest applying different computational protocols which highlighted how pathogenic mutations can affect protein functionality. LRIG2 was predicted to be able to homodimerize and the role of the Ig-1 domain in the dimerization was deeply studied, as a novel splicing variant deleted of the Ig-1 domain was found

to cause the urofacial syndrome. Results highlighted how three known mutations in the Ig-1 domain could all impair the LRIG2 dimerization, through the modification of the Ig-1 domain quaternary structure. These findings could give a better understanding of the LRIG2 protein function and, possibly, help the discovery of drugs able to restore the impairment of the LRIG2 dimerization due to LRIG2 point mutations, known to be pathogenic for such a rare disease as urofacial syndrome.

Eventually, also the study of the mutated disordered region of NUP98, found causative of a phenotype resembling the Rothmund-Thomson syndrome, led to the evaluation of the importance of the FG-motifs in the protein folding. A multi-replica computational approach showed how the G28D mutation might affect the folding of the NUP98 disordered domain. The understanding of how the dynamicity of a disorder domain is influenced by the intramolecular cohesions of the FG-motifs and can be altered by a single mutation is indeed helpful for the future comprehension of the multiple functionalities of this protein.

Finally, it is worth mentioning that the presented computational methodologies and the ones already existing, but not used and discussed in this thesis, will be of extreme importance in the next few years to help the search for new drugs and/or the identification of new mechanisms of actions of the proteins. The virtual screening will be incredible beneficial from the use of the machine and deep learning or artificial intelligence both to identify binding pocket, identify pose, de-novo, and rescore the docking results [206-211].

The continuous upgrading of the infrastructures has unlocked incredible time-scale simulations able to uncover allosteric and conformational changes. The development of new advanced molecular dynamic simulations allows the study of the mechanism of action, ligand binding, protein-protein interaction, protein folding, and so on, of a lot of different proteins of therapeutic interest [212]. However, it is also critical to stress that it is the deep knowledge of the biological question the real reason why all these technologies are so helpful. These new tools would be completely unuseful if not applied with the right criteria because every protein is different from the others and needs to be studied with an *ad hoc* procedure.

## 12 Bibliography

1. Elborn JS. Cystic fibrosis. *Lancet*. 2016 Nov 19;388(10059):2519-2531. doi: 10.1016/S0140-6736(16)00576-6. Epub 2016 Apr 29. PMID: 27140670.
2. Scotet V, L'Hostis C, Férec C. The Changing Epidemiology of Cystic Fibrosis: Incidence, Survival and Impact of the CFTR Gene Discovery. *Genes (Basel)*. 2020 May 26;11(6):589. doi: 10.3390/genes11060589. PMID: 32466381; PMCID: PMC7348877.
3. Bar-Or O, Blimkie CJ, Hay JA, MacDougall JD, Ward DS, Wilson WM. Voluntary dehydration and heat intolerance in cystic fibrosis. *Lancet*. 1992 Mar 21;339(8795):696-9. doi: 10.1016/0140-6736(92)90597-v. PMID: 1347582.
4. Morkeberg JC, Edmund C, Prause JU, Lanng S, Koch C, Michaelsen KF. Ocular findings in cystic fibrosis patients receiving vitamin A supplementation. *Graefes Arch Clin Exp Ophthalmol*. 1995 Nov;233(11):709-13. doi: 10.1007/BF00164674. PMID: 8566828.
5. Farinha CM, Callebaut I. Molecular mechanisms of cystic fibrosis - how mutations lead to misfunction and guide therapy. *Biosci Rep*. 2022 Jul 29;42(7):BSR20212006. doi: 10.1042/BSR20212006. PMID: 35707985; PMCID: PMC9251585.
6. Scott P, Anderson K, Singhanian M, Cormier R. Cystic Fibrosis, CFTR, and Colorectal Cancer. *Int J Mol Sci*. 2020 Apr 21;21(8):2891. doi: 10.3390/ijms21082891. PMID: 32326161; PMCID: PMC7215855.
7. Rivas Caldas R, Boisramé S. Upper aero-digestive contamination by *Pseudomonas aeruginosa* and implications in Cystic Fibrosis. *J Cyst Fibros*. 2015 Jan;14(1):6-15. doi: 10.1016/j.jcf.2014.04.008. Epub 2014 Jun 3. PMID: 24933711.
8. Amaral MD, Farinha CM. Rescuing mutant CFTR: a multi-task approach to a better outcome in treating cystic fibrosis. *Curr Pharm Des*. 2013;19(19):3497-508. doi: 10.2174/13816128113199990318. PMID: 23331027.
9. Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989 Sep 8;245(4922):1073-80. doi: 10.1126/science.2570460. PMID: 2570460.

10. Lukacs GL, Verkman AS. CFTR: folding, misfolding and correcting the  $\Delta F508$  conformational defect. *Trends Mol Med*. 2012 Feb;18(2):81-91. doi: 10.1016/j.molmed.2011.10.003. Epub 2011 Dec 3. PMID: 22138491; PMCID: PMC3643519.
11. Lopes-Pacheco M. CFTR Modulators: Shedding Light on Precision Medicine for Cystic Fibrosis. *Front Pharmacol*. 2016 Sep 5;7:275. doi: 10.3389/fphar.2016.00275. PMID: 27656143; PMCID: PMC5011145.
12. Infield DT, Strickland KM, Gaggar A, McCarty NA. The molecular evolution of function in the CFTR chloride channel. *J Gen Physiol*. 2021 Dec 6;153(12):e202012625. doi: 10.1085/jgp.202012625. Epub 2021 Oct 14. PMID: 34647973; PMCID: PMC8640958.
13. Cant N, Pollock N, Ford RC. CFTR structure and cystic fibrosis. *Int J Biochem Cell Biol*. 2014 Jul;52:15-25. doi: 10.1016/j.biocel.2014.02.004. Epub 2014 Feb 15. PMID: 24534272.
14. Aubin CN, Linsdell P. Positive charges at the intracellular mouth of the pore regulate anion conduction in the CFTR chloride channel. *J Gen Physiol*. 2006 Nov;128(5):535-45. doi: 10.1085/jgp.200609516. Epub 2006 Oct 16. PMID: 17043152; PMCID: PMC2151590.
15. Kim Chiaw P, Eckford PD, Bear CE. Insights into the mechanisms underlying CFTR channel activity, the molecular basis for cystic fibrosis and strategies for therapy. *Essays Biochem*. 2011 Sep 7;50(1):233-48. doi: 10.1042/bse0500233. PMID: 21967060.
16. Hunt JF, Wang C, Ford RC. Cystic fibrosis transmembrane conductance regulator (ABCC7) structure. *Cold Spring Harb Perspect Med*. 2013 Feb 1;3(2):a009514. doi: 10.1101/cshperspect.a009514. PMID: 23378596; PMCID: PMC3552343.
17. Linsdell P. Architecture and functional properties of the CFTR channel pore. *Cell Mol Life Sci*. 2017 Jan;74(1):67-83. doi: 10.1007/s00018-016-2389-5. Epub 2016 Oct 3. PMID: 27699452.
18. Cheng SH, Rich DP, Marshall J, Gregory RJ, Welsh MJ, Smith AE. Phosphorylation of the R domain by cAMP-dependent protein kinase regulates the CFTR chloride channel. *Cell*. 1991 Sep 6;66(5):1027-36. doi: 10.1016/0092-8674(91)90446-6. PMID: 1716180.
19. Wang W, Wu J, Bernard K, Li G, Wang G, Bevenssee MO, Kirk KL. ATP-independent CFTR channel gating and allosteric modulation by phosphorylation. *Proc Natl Acad Sci U S A*. 2010 Feb 23;107(8):3888-93. doi: 10.1073/pnas.0913001107. Epub 2010 Feb 3. PMID: 20133716; PMCID: PMC2840504.



20. Wei S, Roessler BC, Icyuz M, Chauvet S, Tao B, Hartman JL 4th, Kirk KL. Long-range coupling between the extracellular gates and the intracellular ATP binding domains of multidrug resistance protein pumps and cystic fibrosis transmembrane conductance regulator channels. *FASEB J*. 2016 Mar;30(3):1247-62. doi: 10.1096/fj.15-278382. Epub 2015 Nov 25. PMID: 26606940; PMCID: PMC4750415.
21. Quon BS, Rowe SM. New and emerging targeted therapies for cystic fibrosis. *BMJ*. 2016 Mar 30;352:i859. doi: 10.1136/bmj.i859. PMID: 27030675; PMCID: PMC4817245
22. Amaral MD. Novel personalized therapies for cystic fibrosis: treating the basic defect in all patients. *J Intern Med*. 2015 Feb;277(2):155-166. doi: 10.1111/joim.12314. PMID: 25266997.
23. Bompadre SG, Sohma Y, Li M, Hwang TC. G551D and G1349D, two CF-associated mutations in the signature sequences of CFTR, exhibit distinct gating defects. *J Gen Physiol*. 2007 Apr;129(4):285-98. doi: 10.1085/jgp.200609667. Epub 2007 Mar 12. PMID: 17353351; PMCID: PMC2151620.
24. Ramalho AS, Lewandowska MA, Farinha CM, Mendes F, Gonçalves J, Barreto C, Harris A, Amaral MD. Deletion of CFTR translation start site reveals functional isoforms of the protein in CF patients. *Cell Physiol Biochem*. 2009;24(5-6):335-46. doi: 10.1159/000257426. Epub 2009 Nov 4. PMID: 19910674; PMCID: PMC2793277.
25. De Boeck K, Amaral MD. Progress in therapies for cystic fibrosis. *Lancet Respir Med*. 2016 Aug;4(8):662-674. doi: 10.1016/S2213-2600(16)00023-0. Epub 2016 Apr 1. PMID: 27053340.
26. Miller AC, Comellas AP, Hornick DB, Stoltz DA, Cavanaugh JE, Gerke AK, Welsh MJ, Zabner J, Polgreen PM. Cystic fibrosis carriers are at increased risk for a wide range of cystic fibrosis-related conditions. *Proc Natl Acad Sci U S A*. 2020 Jan 21;117(3):1621-1627. doi: 10.1073/pnas.1914912117. Epub 2019 Dec 27. PMID: 31882447; PMCID: PMC6983448.
27. Polgreen PM, Brown GD, Hornick DB, Ahmad F, London B, Stoltz DA, Comellas AP. CFTR Heterozygotes Are at Increased Risk of Respiratory Infections: A Population-Based Study. *Open Forum Infect Dis*. 2018 Nov 1;5(11):ofy219. doi: 10.1093/ofid/ofy219. PMID: 30397620; PMCID: PMC6210382.
28. Amin R, Ratjen F. Emerging drugs for cystic fibrosis. *Expert Opin Emerg Drugs*. 2014 Mar;19(1):143-55. doi: 10.1517/14728214.2014.882316. Epub 2014 Jan 30. PMID: 24479826.

29. Elborn JS. Cystic fibrosis. *Lancet*. 2016 Nov 19;388(10059):2519-2531. doi: 10.1016/S0140-6736(16)00576-6. Epub 2016 Apr 29. PMID: 27140670.
30. Fajac I, De Boeck K. New horizons for cystic fibrosis treatment. *Pharmacol Ther*. 2017 Feb;170:205-211. doi: 10.1016/j.pharmthera.2016.11.009. Epub 2016 Dec 1. PMID: 27916649
31. Prayle A, Watson A, Fortnum H, Smyth A. Side effects of aminoglycosides on the kidney, ear and balance in cystic fibrosis. *Thorax*. 2010 Jul;65(7):654-8. doi: 10.1136/thx.2009.131532. PMID: 20627927; PMCID: PMC2921289.
32. Miller, J. P.; Drew, L.; Green, O.; Vilella, A.; McEwan, B.; Patel, N.; Qiu, D.; Bhalla, A.; Bastos, C.; Parks, D.; Giuliano, K.; Longo, K.; Ivarsson, M.; Munoz, B.; Lee, P.-S.; Mehmet, H.; Haeberlein, M., CFTR Amplifiers: A New Class of CFTR Modulator that Complements the Substrate Limitations of Other CF Therapeutic Modalities. In C60. ALL ABOUT CYSTIC FIBROSIS, American Thoracic Society: 2016; pp A5574-A5574.
33. Meoli A, Fainardi V, Deolmi M, Chiopris G, Marinelli F, Caminiti C, Esposito S, Pisi G. State of the Art on Approved Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Modulators and Triple-Combination Therapy. *Pharmaceuticals (Basel)*. 2021 Sep 15;14(9):928. doi: 10.3390/ph14090928. PMID: 34577628; PMCID: PMC8471029.
34. Flume PA, Liou TG, Borowitz DS, Li H, Yen K, Ordoñez CL, Geller DE; VX 08-770-104 Study Group. Ivacaftor in subjects with cystic fibrosis who are homozygous for the F508del-CFTR mutation. *Chest*. 2012 Sep;142(3):718-724. doi: 10.1378/chest.11-2672. PMID: 22383668; PMCID: PMC3435140.
35. Mall MA, Sheppard DN. Chronic ivacaftor treatment: getting F508del-CFTR into more trouble? *J Cyst Fibros*. 2014 Dec;13(6):605-7. doi: 10.1016/j.jcf.2014.10.001. Epub 2014 Oct 16. PMID: 25458465.
36. Cholon DM, Esther CR Jr, Gentsch M. Efficacy of lumacaftor-ivacaftor for the treatment of cystic fibrosis patients homozygous for the F508del-CFTR mutation. *Expert Rev Precis Med Drug Dev*. 2016;1(3):235-243. doi: 10.1080/23808993.2016.1175299. Epub 2016 Apr 22. PMID: 27482545; PMCID: PMC4963025.
37. Munck A, Kerem E, Ellemunter H, Campbell D, Wang LT, Ahluwalia N, Owen CA, Wainwright C. Tezacaftor/ivacaftor in people with cystic fibrosis heterozygous for minimal function CFTR

- mutations. *J Cyst Fibros.* 2020 Nov;19(6):962-968. doi: 10.1016/j.jcf.2020.04.015. Epub 2020 Jun 13. PMID: 32546431.
38. Becq F, Mirval S, Carrez T, Lévêque M, Billet A, Coraux C, Sage E, Cantereau A. The rescue of F508del-CFTR by elexacaftor/tezacaftor/ivacaftor (Trikafta) in human airway epithelial cells is underestimated due to the presence of ivacaftor. *Eur Respir J.* 2022 Feb 24;59(2):2100671. doi: 10.1183/13993003.00671-2021. PMID: 34266939
39. Farinha CM, Amaral MD. Most F508del-CFTR is targeted to degradation at an early folding checkpoint and independently of calnexin. *Mol Cell Biol.* 2005 Jun;25(12):5242-52. doi: 10.1128/MCB.25.12.5242-5252.2005. PMID: 15923638; PMCID: PMC1140594.
40. Birault V, Solari R, Hanrahan J, Thomas DY. Correctors of the basic trafficking defect of the mutant F508del-CFTR that causes cystic fibrosis. *Curr Opin Chem Biol.* 2013 Jun;17(3):353-60. doi: 10.1016/j.cbpa.2013.04.020. Epub 2013 May 24. PMID: 23711435.
41. Van Goor F, Hadida S, Grootenhuys PD, Burton B, Stack JH, Straley KS, Decker CJ, Miller M, McCartney J, Olson ER, Wine JJ, Frizzell RA, Ashlock M, Negulescu PA. Correction of the F508del-CFTR protein processing defect in vitro by the investigational drug VX-809. *Proc Natl Acad Sci U S A.* 2011 Nov 15;108(46):18843-8. doi: 10.1073/pnas.1105787108. Epub 2011 Oct 5. PMID: 21976485; PMCID: PMC3219147
42. Clancy JP, Rowe SM, Accurso FJ, Aitken ML, Amin RS, Ashlock MA, Ballmann M, Boyle MP, Bronsveld I, Campbell PW, De Boeck K, Donaldson SH, Dorkin HL, Dunitz JM, Durie PR, Jain M, Leonard A, McCoy KS, Moss RB, Pilewski JM, Rosenbluth DB, Rubenstein RC, Schechter MS, Botfield M, Ordoñez CL, Spencer-Green GT, Vernillet L, Wisseh S, Yen K, Konstan MW. Results of a phase IIa study of VX-809, an investigational CFTR corrector compound, in subjects with cystic fibrosis homozygous for the F508del-CFTR mutation. *Thorax.* 2012 Jan;67(1):12-8. doi: 10.1136/thoraxjnl-2011-200393. Epub 2011 Aug 8. PMID: 21825083; PMCID: PMC3746507.].
43. Burgel, P.R.; Durieu, I.; Chiron, R.; Ramel, S.; Danner-Boucher, I.; Prevotat, A.; Grenet, D.; Marguet, C.; Reynaud-Gaubert, M.; Macey, J.; et al. Rapid Improvement after Starting Elexacaftor-tezacaftor-ivacaftor in Patients with Cystic Fibrosis and Advanced Pulmonary Disease. *Am. J. Respir. Crit. Care Med.* 2021, 204

44. Lopes-Pacheco M. CFTR Modulators: The Changing Face of Cystic Fibrosis in the Era of Precision Medicine. *Front Pharmacol.* 2020 Feb 21;10:1662. doi: 10.3389/fphar.2019.01662. PMID: 32153386; PMCID: PMC7046560.
45. Lopes-Pacheco M, Sabirzhanova I, Rapino D, Morales MM, Guggino WB, Cebotaru L. Correctors Rescue CFTR Mutations in Nucleotide-Binding Domain 1 (NBD1) by Modulating Proteostasis. *Chembiochem.* 2016 Mar 15;17(6):493-505. doi: 10.1002/cbic.201500620. Epub 2016 Feb 16. PMID: 26864378; PMCID: PMC5557405.
46. Fiedorczuk K, Chen J. Molecular structures reveal synergistic rescue of  $\Delta 508$  CFTR by Trikafta modulators. *Science.* 2022 Oct 21;378(6617):284-290. doi: 10.1126/science.ade2216. Epub 2022 Oct 20. PMID: 36264792.
47. Hillenaar T, Yeoh HY, Sahasrabudhe P, Mijnders M, van Willigen M, van der Sluijs P, Braakman I, Im J. ABC-transporter CFTR folds with high fidelity through a modular, stepwise pathway, *bioRxiv*, 2022.07.20.500765; doi: <https://doi.org/10.1101/2022.07.20.500765>.
48. Okiyoneda T, Veit G, Dekkers JF, Bagdany M, Soya N, Xu H, Roldan A, Verkman AS, Kurth M, Simon A, Hegedus T, Beekman JM, Lukacs GL. Mechanism-based corrector combination restores  $\Delta F508$ -CFTR folding and function. *Nat Chem Biol.* 2013 Jul;9(7):444-54. doi: 10.1038/nchembio.1253. Epub 2013 May 12. PMID: 23666117; PMCID: PMC3840170.
49. He L, Kota P, Aleksandrov AA, Cui L, Jensen T, Dokholyan NV, Riordan JR. Correctors of  $\Delta F508$  CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB J.* 2013 Feb;27(2):536-45. doi: 10.1096/fj.12-216119. Epub 2012 Oct 26. Erratum in: *FASEB J.* 2014 Jun;28(6):2737. PMID: 23104983; PMCID: PMC3545534.
50. D'Ursi P, Uggeri M, Urbinati C, Millo E, Paiardi G, Milanesi L, Ford RC, Clews J, Meng X, Bergese P, Ridolfi A, Pedemonte N, Fossa P, Orro A, Rusnati M. Exploitation of a novel biosensor based on the full-length human F508de1-CFTR with computational studies, biochemical and biological assays for the characterization of a new Lumacaftor/Tezacaftor analogue Sensor. *Actuator. B Chem.*, 2019, 301, <https://doi.org/10.1016/j.snb.2019.127131>.
51. Hudson RP, Dawson JE, Chong PA, Yang Z, Millen L, Thomas PJ, Brouillette CG, Forman-Kay JD. Direct Binding of the Corrector VX-809 to Human CFTR NBD1: Evidence of an Allosteric Coupling

- between the Binding Site and the NBD1:CL4 Interface. *Mol Pharmacol*. 2017 Aug;92(2):124-135. doi: 10.1124/mol.117.108373. Epub 2017 May 25. PMID: 28546419.
52. Ren HY, Grove DE, De La Rosa O, Houck SA, Sopha P, Van Goor F, Hoffman BJ, Cyr DM. VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol Biol Cell*. 2013 Oct;24(19):3016-24. doi: 10.1091/mbc.E13-05-0240. Epub 2013 Aug 7. PMID: 23924900; PMCID: PMC3784376.
53. Laselva O, Marzaro G, Vaccarin C, Lampronti I, Tamanini A, Lippi G, Gambari R, Cabrini G, Bear CE, Chilin A, Dechecchi MC. Molecular Mechanism of Action of Trimethylangelicin Derivatives as CFTR Modulators. *Front Pharmacol*. 2018 Jul 4;9:719. doi: 10.3389/fphar.2018.00719. PMID: 30022950; PMCID: PMC6039571.
54. Loo TW, Bartlett MC, Clarke DM. Corrector VX-809 stabilizes the first transmembrane domain of CFTR. *Biochem Pharmacol*. 2013 Sep 1;86(5):612-9. doi: 10.1016/j.bcp.2013.06.028. Epub 2013 Jul 5. PMID: 23835419.
55. Loo TW, Clarke DM. Corrector VX-809 promotes interactions between cytoplasmic loop one and the first nucleotide-binding domain of CFTR. *Biochem Pharmacol*. 2017 Jul 15;136:24-31. doi: 10.1016/j.bcp.2017.03.020. Epub 2017 Mar 31. PMID: 28366727.
56. Baatallah N, Elbahnsi A, Mornon JP, Chevalier B, Pranke I, Servel N, Zelli R, Décout JL, Edelman A, Sermet-Gaudelus I, Callebaut I, Hinzpeter A. Pharmacological chaperones improve intra-domain stability and inter-domain assembly via distinct binding sites to rescue misfolded CFTR. *Cell Mol Life Sci*. 2021 Dec;78(23):7813-7829. doi: 10.1007/s00018-021-03994-5. Epub 2021 Oct 29. PMID: 34714360.
57. Liu F, Zhang Z, Csanády L, Gadsby DC, Chen J. Molecular Structure of the Human CFTR Ion Channel. *Cell*. 2017 Mar 23;169(1):85-95.e8. doi: 10.1016/j.cell.2017.02.024. PMID: 28340353
58. Zhang Z, Liu F, Chen J. Molecular structure of the ATP-bound, phosphorylated human CFTR. *Proc Natl Acad Sci U S A*. 2018 Dec 11;115(50):12757-12762. doi: 10.1073/pnas.1815287115. Epub 2018 Nov 20. PMID: 30459277; PMCID: PMC6294961.
59. Liu F, Zhang Z, Levit A, Levring J, Touhara KK, Shoichet BK, Chen J. Structural identification of a hotspot on CFTR for potentiation. *Science*. 2019 Jun 21;364(6446):1184-1188. doi: 10.1126/science.aaw7611. PMID: 31221859; PMCID: PMC7184887.

60. Fiedorczuk K, Chen J. Mechanism of CFTR correction by type I folding correctors. *Cell*. 2022 Jan 6;185(1):158-168.e11. doi: 10.1016/j.cell.2021.12.009. PMID: 34995514.
61. Faitelson Y, Beigelman A. Can You Teach an Old Drug New Tricks? The Use of Salbutamol for Gastrointestinal Anaphylaxis. *J Allergy Clin Immunol Pract*. 2021 Aug;9(8):3138-3139. doi: 10.1016/j.jaip.2021.05.017. PMID: 34366098.
62. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004 Aug;3(8):673-83. doi: 10.1038/nrd1468. PMID: 15286734.
63. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Williams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019 Jan;18(1):41-58. doi: 10.1038/nrd.2018.168. Epub 2018 Oct 12. PMID: 30310233.
64. Jarada TN, Rokne JG, Alhadj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *J Cheminform*. 2020 Jul 22;12(1):46. doi: 10.1186/s13321-020-00450-7. PMID: 33431024; PMCID: PMC7374666.
65. Anderson SD, Daviskas E, Brannan JD, Chan HK. Repurposing excipients as active inhalation agents: The mannitol story. *Adv Drug Deliv Rev*. 2018 Aug;133:45-56. doi: 10.1016/j.addr.2018.04.003. Epub 2018 Apr 5. PMID: 29626547.
66. Newman SP. Delivering drugs to the lungs: The history of repurposing in the treatment of respiratory diseases. *Adv Drug Deliv Rev*. 2018 Aug;133:5-18. doi: 10.1016/j.addr.2018.04.010. Epub 2018 Apr 11. PMID: 29653129.
67. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006 Jan 1;34(Database issue):D668-72. doi: 10.1093/nar/gkj067. PMID: 16381955; PMCID: PMC1347430.
68. Greenwood JR, Calkins D, Sullivan AP, Shelley JC. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J Comput Aided Mol Des*. 2010 Jun;24(6-7):591-604. Doi: 10.1007/s10822-010-9349-1. Epub 2010 Mar 31. PMID: 20354892.

69. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des.* 2007 Dec;21(12):681-91. Doi: 10.1007/s10822-007-9133-z. Epub 2007 Sep 27. PMID: 17899391.
70. D'Ursi P, Uggeri M, Urbinati C, Millo E, Paiardi G, Milanesi L, Ford RC, Clews J, Meng X, Bergese P, Ridolfi A, Pedemonte N, Fossa P, Orro A, Rusnati M. Exploitation of a novel biosensor based on the full-length human F508de1-CFTR with computational studies, biochemical and biological assays for the characterization of a new Lumacaftor/Tezacaftor analogue Sensor. *Actuator. B Chem.*, 2019, 301, <https://doi.org/10.1016/j.snb.2019.127131>.
71. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics.* 2009 Jun 2;10:168. doi: 10.1186/1471-2105-10-168. PMID: 19486540; PMCID: PMC2700099.
72. Orro A, Uggeri M, Rusnati M, Urbinati C, Pedemonte N, Pesce E, Moscatelli M, Padoan R, Cichero E, Fossa P, D'Ursi P. In silico drug repositioning on F508del-CFTR: A proof-of-concept study on the AIFA library. *Eur J Med Chem.* 2021 Mar 5;213:113186. doi: 10.1016/j.ejmech.2021.113186. Epub 2021 Jan 13. PMID: 33472120.
73. Liu K, Kokubo H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-docking Study. *J Chem Inf Model.* 2017 Oct 23;57(10):2514-2522. doi: 10.1021/acs.jcim.7b00412. Epub 2017 Sep 29. PMID: 28902511.
74. Liu K, Watanabe E, Kokubo H. Exploring the stability of ligand binding modes to proteins by molecular dynamics simulations. *J Comput Aided Mol Des.* 2017 Feb;31(2):201-211. doi: 10.1007/s10822-016-0005-2. Epub 2017 Jan 10. PMID: 28074360.
75. Liu K, Kokubo H. Prediction of ligand binding mode among multiple cross-docking poses by molecular dynamics simulations. *J Comput Aided Mol Des.* 2020 Nov;34(11):1195-1205. doi: 10.1007/s10822-020-00340-y. Epub 2020 Sep 1. PMID: 32869148.
76. Amaretti A, Raimondi S, Leonardi A, Quartieri A, Rossi M. Hydrolysis of the rutinose-conjugates flavonoids rutin and hesperidin by the gut microbiota and bifidobacteria. *Nutrients.* 2015 Apr 14;7(4):2788-800. doi: 10.3390/nu7042788. PMID: 25875120; PMCID: PMC4425173.

77. Shin NR, Moon JS, Shin SY, Li L, Lee YB, Kim TJ, Han NS. Isolation and characterization of human intestinal *Enterococcus avium* EFEL009 converting rutin to quercetin. *Lett Appl Microbiol*. 2016 Jan;62(1):68-74. doi: 10.1111/lam.12512. Epub 2015 Dec 9. PMID: 26505733.
78. Yu B, Jiang Y, Jin L, Ma T, Yang H. Role of Quercetin in Modulating Chloride Transport in the Intestine. *Front Physiol*. 2016 Nov 23;7:549. doi: 10.3389/fphys.2016.00549. PMID: 27932986; PMCID: PMC5120089.
79. Pyle LC, Fulton JC, Sloane PA, Backer K, Mazur M, Prasain J, Barnes S, Clancy JP, Rowe SM. Activation of the cystic fibrosis transmembrane conductance regulator by the flavonoid quercetin: potential use as a biomarker of  $\Delta F508$  cystic fibrosis transmembrane conductance regulator rescue. *Am J Respir Cell Mol Biol*. 2010 Nov;43(5):607-16. doi: 10.1165/rcmb.2009-0281OC. Epub 2009 Dec 30. PMID: 20042712; PMCID: PMC2970857.
80. Lubamba B, Lebacqz J, Reyckler G, Marbaix E, Wallemacq P, Lebecque P, Leal T. Inhaled phosphodiesterase type 5 inhibitors restore chloride transport in cystic fibrosis mice. *Eur Respir J*. 2011 Jan;37(1):72-8. doi: 10.1183/09031936.00013510. Epub 2010 Jun 18. PMID: 20562123
81. Joensuu H. Treatment of inoperable gastrointestinal stromal tumor (GIST) with Imatinib (Gleevec, Gleevec). *Med Klin (Munich)*. 2002 Jan 15;97 Suppl 1:28-30. PMID: 11831069.
82. Pillaiyar T, Meenakshisundaram S, Manickam M, Sankaranarayanan M. A medicinal chemistry perspective of drug repositioning: Recent advances and challenges in drug discovery. *Eur J Med Chem*. 2020 Jun 1;195:112275. doi: 10.1016/j.ejmech.2020.112275. Epub 2020 Apr 2. PMID: 32283298; PMCID: PMC7156148.
83. Lara-Ramirez EE, López-Cedillo JC, Noguera-Torres B, Kashif M, Garcia-Perez C, Bocanegra-Garcia V, Agusti R, Uhrig ML, Rivera G. An in vitro and in vivo evaluation of new potential trans-sialidase inhibitors of *Trypanosoma cruzi* predicted by a computational drug repositioning method. *Eur J Med Chem*. 2017 May 26;132:249-261. doi: 10.1016/j.ejmech.2017.03.063. Epub 2017 Mar 28. PMID: 28364659.
84. Makino T, Shimizu R, Kanemaru M, Suzuki Y, Moriwaki M, Mizukami H. Enzymatically modified isoquercitrin, alpha-oligoglucosyl quercetin 3-O-glucoside, is absorbed more easily than other quercetin glycosides or aglycone after oral administration in rats. *Biol Pharm Bull*. 2009 Dec;32(12):2034-40. doi: 10.1248/bpb.32.2034. PMID: 19952424.



85. Liu F, Zhang Z, Levit A, Levring J, Touhara KK, Shoichet BK, Chen J. Structural identification of a hotspot on CFTR for potentiation. *Science*. 2019 Jun 21;364(6446):1184-1188. doi: 10.1126/science.aaw7611. PMID: 31221859; PMCID: PMC7184887.
86. Devereux G, Steele S, Griffiths K, Devlin E, Fraser-Pitt D, Cotton S, Norrie J, Chrystyn H, O'Neil D. An Open-Label Investigation of the Pharmacokinetics and Tolerability of Oral Cysteamine in Adults with Cystic Fibrosis. *Clin Drug Investig*. 2016 Aug;36(8):605-12. doi: 10.1007/s40261-016-0405-z. PMID: 27153825; PMCID: PMC4951511.
87. Tosco A, De Gregorio F, Esposito S, De Stefano D, Sana I, Ferrari E, Sepe A, Salvadori L, Buonpensiero P, Di Pasqua A, Grassia R, Leone CA, Guido S, De Rosa G, Lusa S, Bona G, Stoll G, Maiuri MC, Mehta A, Kroemer G, Maiuri L, Raia V. A novel treatment of cystic fibrosis acting on-target: cysteamine plus epigallocatechin gallate for the autophagy-dependent rescue of class II-mutated CFTR. *Cell Death Differ*. 2017 Jul;24(7):1305. doi: 10.1038/cdd.2016.43. Epub 2016 Jul 22. Erratum for: *Cell Death Differ*. 2016 Aug;23(8):1380-93. PMID: 27447111; PMCID: PMC5520161.
88. De Stefano D, Vilella VR, Esposito S, Tosco A, Sepe A, De Gregorio F, Salvadori L, Grassia R, Leone CA, De Rosa G, Maiuri MC, Pettoello-Mantovani M, Guido S, Bossi A, Zolin A, Venerando A, Pinna LA, Mehta A, Bona G, Kroemer G, Maiuri L, Raia V. Restoration of CFTR function in patients with cystic fibrosis carrying the F508del-CFTR mutation. *Autophagy*. 2014;10(11):2053-74. doi: 10.4161/15548627.2014.973737. PMID: 25350163; PMCID: PMC4502695.
89. Devereux G, Wrolstad D, Bourke SJ, Daines CL, Doe S, Dougherty R, Franco R, Innes A, Kopp BT, Lascano J, Layish D, MacGregor G, Murray L, Peckham D, Lucidi V, Lovie E, Robertson J, Fraser-Pitt DJ, O'Neil DA. Oral cysteamine as an adjunct treatment in cystic fibrosis pulmonary exacerbations: An exploratory randomized clinical trial. *PLoS One*. 2020 Dec 28;15(12):e0242945. doi: 10.1371/journal.pone.0242945. PMID: 33370348; PMCID: PMC7769283.
90. Vilella VR, Esposito S, Maiuri MC, Raia V, Kroemer G, Maiuri L. Towards a rational combination therapy of cystic fibrosis: How cysteamine restores the stability of mutant CFTR. *Autophagy*. 2013 Sep;9(9):1431-4. doi: 10.4161/auto.25517. Epub 2013 Jun 25. PMID: 23800975.

91. Ferrari E, Monzani R, Villella VR, Esposito S, Saluzzo F, Rossin F, D'Eletto M, Tosco A, De Gregorio F, Izzo V, Maiuri MC, Kroemer G, Raia V, Maiuri L. Cysteamine re-establishes the clearance of *Pseudomonas aeruginosa* by macrophages bearing the cystic fibrosis-relevant F508del-CFTR mutation. *Cell Death Dis.* 2017 Jan 12;8(1):e2544. doi: 10.1038/cddis.2016.476. PMID: 28079883; PMCID: PMC5386380.
92. Charrier C, Rodger C, Robertson J, Kowalczyk A, Shand N, Fraser-Pitt D, Mercer D, O'Neil D. Cysteamine (Lynovex®), a novel mucoactive antimicrobial & antibiofilm agent for the treatment of cystic fibrosis. *Orphanet J Rare Dis.* 2014 Nov 30;9:189. doi: 10.1186/s13023-014-0189-2. PMID: 25433388; PMCID: PMC4260250.
93. Tomati V, Caci E, Ferrera L, Pesce E, Sondo E, Cholon DM, Quinney NL, Boyles SE, Armirotti A, Ravazzolo R, Galiotta LJ, Gentsch M, Pedemonte N. Thymosin  $\alpha$ -1 does not correct F508del-CFTR in cystic fibrosis airway epithelia. *JCI Insight.* 2018 Feb 8;3(3):e98699. doi: 10.1172/jci.insight.98699. Erratum in: *JCI Insight.* 2019 Apr 4;4(7): PMID: 29415893; PMCID: PMC5821210.
94. Armirotti A, Tomati V, Matthes E, Veit G, Cholon DM, Phuan PW, Braccia C, Guidone D, Gentsch M, Lukacs GL, Verkman AS, Galiotta LJV, Hanrahan JW, Pedemonte N. Bioactive Thymosin Alpha-1 Does Not Influence F508del-CFTR Maturation and Activity. *Sci Rep.* 2019 Jul 16;9(1):10310. doi: 10.1038/s41598-019-46639-1. PMID: 31311979; PMCID: PMC6635361.
95. Odera M, Furuta T, Sohma Y, Sakurai M. Molecular dynamics simulation study on the structural instability of the most common cystic fibrosis-associated mutant  $\Delta$ F508-CFTR. *Biophys Physicobiol.* 2018 Feb 6;15:33-44. doi: 10.2142/biophysico.15.0\_33. PMID: 29607278; PMCID: PMC5873040.
96. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W443-7. doi: 10.1093/nar/gkv315. Epub 2015 Apr 14. PMID: 25873628; PMCID: PMC4489249.
97. Prins S, Corradi V, Sheppard DN, Tieleman DP, Vergani P. Can two wrongs make a right? F508del-CFTR ion channel rescue by second-site mutations in its transmembrane domains. *J Biol Chem.* 2022 Mar;298(3):101615. doi: 10.1016/j.jbc.2022.101615. Epub 2022 Jan 21. PMID: 35065958; PMCID: PMC8861112.

98. Parodi A, Righetti G, Pesce E, Salis A, Tasso B, Urbinati C, Tomati V, Damonte G, Rusnati M, Pedemonte N, Cichero E, Millo E. Discovery of novel VX-809 hybrid derivatives as F508del-CFTR correctors by molecular modeling, chemical synthesis and biological assays. *Eur J Med Chem.* 2020 Dec 15;208:112833. doi: 10.1016/j.ejmech.2020.112833. Epub 2020 Sep 12. PMID: 32971410.
99. Fossa P, Uggeri M, Orro A, Urbinati C, Rondina A, Milanesi M, Pedemonte N, Pesce E, Padoan R, Ford RC, Meng X, Rusnati M, D'Ursi P. Virtual Drug Repositioning as a Tool to Identify Natural Small Molecules That Synergize with Lumacaftor in F508del-CFTR Binding and Rescuing. *Int J Mol Sci.* 2022 Oct 14;23(20):12274. doi: 10.3390/ijms232012274. PMID: 36293130; PMCID: PMC9602983.
100. Laselva O, Guerra L, Castellani S, Favia M, Di Gioia S, Conese M. Small-molecule drugs for cystic fibrosis: Where are we now? *Pulm Pharmacol Ther.* 2022 Feb;72:102098. doi: 10.1016/j.pupt.2021.102098. Epub 2021 Nov 15. PMID: 34793977.
101. Martelli G, Giacomini D. Antibacterial and antioxidant activities for natural and synthetic dual-active compounds. *Eur J Med Chem.* 2018 Oct 5;158:91-105. doi: 10.1016/j.ejmech.2018.09.009. Epub 2018 Sep 5. PMID: 30205261.
102. Gaskin KJ. Nutritional care in children with cystic fibrosis: are our patients becoming better? *Eur J Clin Nutr.* 2013 May;67(5):558-64. doi: 10.1038/ejcn.2013.20. Epub 2013 Mar 6. PMID: 23462946.
103. Yang H, Ma T. F508del-cystic fibrosis transmembrane regulator correctors for treatment of cystic fibrosis: a patent review. *Expert Opin Ther Pat.* 2015;25(9):991-1002. doi: 10.1517/13543776.2015.1045878. Epub 2015 May 15. PMID: 25971311.
104. Dey I, Shah K, Bradbury NA. Natural Compounds as Therapeutic Agents in the Treatment Cystic Fibrosis. *J Genet Syndr Gene Ther.* 2016 Feb;7(1):284. doi: 10.4172/2157-7412.1000284. Epub 2016 Jan 30. PMID: 27081574; PMCID: PMC4828912.
105. Han MK, Barreto TA, Martinez FJ, Comstock AT, Sajjan US. Randomised clinical trial to determine the safety of quercetin supplementation in patients with chronic obstructive pulmonary disease. *BMJ Open Respir Res.* 2020 Feb;7(1):e000392. doi: 10.1136/bmjresp-2018-000392. PMID: 32071149; PMCID: PMC7047491.

106. Belchamber KBR, Donnelly LE. Targeting defective pulmonary innate immunity - A new therapeutic option? *Pharmacol Ther.* 2020 May;209:107500. doi: 10.1016/j.pharmthera.2020.107500. Epub 2020 Feb 13. PMID: 32061706.
107. Centko RM, Carlile GW, Barne I, Patrick BO, Blagojevic P, Thomas DY, Andersen RJ. Combination of Selective PARP3 and PARP16 Inhibitory Analogues of Latonduine A Corrects F508del-CFTR Trafficking. *ACS Omega.* 2020 Sep 29;5(40):25593-25604. doi: 10.1021/acsomega.0c02467. PMID: 33073085; PMCID: PMC7557227.
108. Dekkers JF, Van Mourik P, Vonk AM, Kruisselbrink E, Berkers G, de Winter-de Groot KM, Janssens HM, Bronsveld I, van der Ent CK, de Jonge HR, Beekman JM. Potentiator synergy in rectal organoids carrying S1251N, G551D, or F508del CFTR mutations. *J Cyst Fibros.* 2016 Sep;15(5):568-78. doi: 10.1016/j.jcf.2016.04.007. Epub 2016 May 5. PMID: 27160424.
109. Wang W, Hong JS, Rab A, Sorscher EJ, Kirk KL. Robust Stimulation of W1282X-CFTR Channel Activity by a Combination of Allosteric Modulators. *PLoS One.* 2016 Mar 23;11(3):e0152232. doi: 10.1371/journal.pone.0152232. PMID: 27007499; PMCID: PMC4805204.
110. Cho DY, Zhang S, Lazrak A, Grayson JW, Peña Garcia JA, Skinner DF, Lim DJ, Mackey C, Banks C, Matalon S, Woodworth BA. Resveratrol and ivacaftor are additive G551D CFTR-channel potentiators: therapeutic implications for cystic fibrosis sinus disease. *Int Forum Allergy Rhinol.* 2019 Jan;9(1):100-105. doi: 10.1002/alr.22202. Epub 2018 Aug 27. PMID: 30152192; PMCID: PMC6318032.
111. Pesce E, Gorrieri G, Sirci F, Napolitano F, Carrella D, Caci E, Tomati V, Zegarra-Moran O, di Bernardo D, Galiotta LJ. Evaluation of a systems biology approach to identify pharmacological correctors of the mutant CFTR chloride channel. *J Cyst Fibros.* 2016 Jul;15(4):425-35. doi: 10.1016/j.jcf.2016.02.009. Epub 2016 Mar 10. PMID: 26971626.
112. Lin CC, Hsieh NK, Liou HL, Chen HI. Niacinamide mitigated the acute lung injury induced by phorbol myristate acetate in isolated rat's lungs. *J Biomed Sci.* 2012 Mar 1;19(1):27. doi: 10.1186/1423-0127-19-27. PMID: 22375599; PMCID: PMC3311060.
113. Fernandes CA, Fievez L, Ucakar B, Neyrinck AM, Fillee C, Huaux F, Delzenne NM, Bureau F, Vanbever R. Nicotinamide enhances apoptosis of G(M)-CSF-treated neutrophils and

- attenuates endotoxin-induced airway inflammation in mice. *Am J Physiol Lung Cell Mol Physiol*. 2011 Mar;300(3):L354-61. doi: 10.1152/ajplung.00198.2010. Epub 2010 Dec 3. PMID: 21131399.
114. Stutts MJ, Gabriel SE, Price EM, Sarkadi B, Olsen JC, Boucher RC. Pyridine nucleotide redox potential modulates cystic fibrosis transmembrane conductance regulator Cl<sup>-</sup> conductance. *J Biol Chem*. 1994 Mar 25;269(12):8667-74. PMID: 7510695.
115. Mendoza JL, Schmidt A, Li Q, Nuvaga E, Barrett T, Bridges RJ, Feranchak AP, Brautigam CA, Thomas PJ. Requirements for efficient correction of  $\Delta F508$  CFTR revealed by analyses of evolved sequences. *Cell*. 2012 Jan 20;148(1-2):164-74. doi: 10.1016/j.cell.2011.11.023. PMID: 22265409; PMCID: PMC3266553.
116. Hwang ES, Song SB. Possible Adverse Effects of High-Dose Nicotinamide: Mechanisms and Safety Assessment. *Biomolecules*. 2020 Apr 29;10(5):687. doi: 10.3390/biom10050687. PMID: 32365524; PMCID: PMC7277745.
117. Wang F, Zeltwanger S, Yang IC, Nairn AC, Hwang TC. Actions of genistein on cystic fibrosis transmembrane conductance regulator channel gating. Evidence for two binding sites with opposite effects. *J Gen Physiol*. 1998 Mar;111(3):477-90. doi: 10.1085/jgp.111.3.477. PMID: 9482713; PMCID: PMC2217116.
118. Habib AR, Kajbafzadeh M, Desai S, Yang CL, Skolnik K, Quon BS. A Systematic Review of the Clinical Efficacy and Safety of CFTR Modulators in Cystic Fibrosis. *Sci Rep*. 2019 May 10;9(1):7234. doi: 10.1038/s41598-019-43652-2. PMID: 31076617; PMCID: PMC6510767.
119. Laselva O, Qureshi Z, Zeng ZW, Petrotchenko EV, Ramjeesingh M, Hamilton CM, Huan LJ, Borchers CH, Pomès R, Young R, Bear CE. Identification of binding sites for ivacaftor on the cystic fibrosis transmembrane conductance regulator. *iScience*. 2021 May 15;24(6):102542. doi: 10.1016/j.isci.2021.102542. PMID: 34142049; PMCID: PMC8184517.
120. Shaughnessy CA, Zeitlin PL, Bratcher PE. Elexacaftor is a CFTR potentiator and acts synergistically with ivacaftor during acute and chronic treatment. *Sci Rep*. 2021 Oct 6;11(1):19810. doi: 10.1038/s41598-021-99184-1. Erratum in: *Sci Rep*. 2021 Oct 25;11(1):21295. PMID: 34615919; PMCID: PMC8494914.

121. Bitam S, Elbahnsi A, Creste G, Pranke I, Chevalier B, Berhal F, Hoffmann B, Servel N, Baatalah N, Tondelier D, Hatton A, Moquereau C, Faria Da Cunha M, Pastor A, Lepissier A, Hinzpeter A, Mornon JP, Prestat G, Edelman A, Callebaut I, Gravier-Pelletier C, Sermet-Gaudelus I. New insights into structure and function of bis-phosphinic acid derivatives and implications for CFTR modulation. *Sci Rep.* 2021 Mar 25;11(1):6842. doi: 10.1038/s41598-021-83240-x. Erratum in: *Sci Rep.* 2021 Sep 15;11(1):18709. PMID: 33767236; PMCID: PMC7994384.
122. Fiedorczuk K, Chen J. Molecular structures reveal synergistic rescue of  $\Delta 508$  CFTR by Trikafta modulators. *Science.* 2022 Oct 21;378(6617):284-290. doi: 10.1126/science.ade2216. Epub 2022 Oct 20. PMID: 36264792.
123. Baatallah N, Elbahnsi A, Mornon JP, Chevalier B, Pranke I, Servel N, Zelli R, Décout JL, Edelman A, Sermet-Gaudelus I, Callebaut I, Hinzpeter A. Pharmacological chaperones improve intra-domain stability and inter-domain assembly via distinct binding sites to rescue misfolded CFTR. *Cell Mol Life Sci.* 2021 Dec;78(23):7813-7829. doi: 10.1007/s00018-021-03994-5. Epub 2021 Oct 29. PMID: 34714360.
124. Loo TW, Bartlett MC, Clarke DM. Corrector VX-809 stabilizes the first transmembrane domain of CFTR. *Biochem Pharmacol.* 2013 Sep 1;86(5):612-9. doi: 10.1016/j.bcp.2013.06.028. Epub 2013 Jul 5. PMID: 23835419.
125. Okiyonedo T, Veit G, Dekkers JF, Bagdany M, Soya N, Xu H, Roldan A, Verkman AS, Kurth M, Simon A, Hegedus T, Beekman JM, Lukacs GL. Mechanism-based corrector combination restores  $\Delta F508$ -CFTR folding and function. *Nat Chem Biol.* 2013 Jul;9(7):444-54. doi: 10.1038/nchembio.1253. Epub 2013 May 12. PMID: 23666117; PMCID: PMC3840170.
126. Eckford PD, Ramjeesingh M, Molinski S, Pasyk S, Dekkers JF, Li C, Ahmadi S, Ip W, Chung TE, Du K, Yeger H, Beekman J, Gonska T, Bear CE. VX-809 and related corrector compounds exhibit secondary activity stabilizing active  $F508\text{del}$ -CFTR after its partial rescue to the cell surface. *Chem Biol.* 2014 May 22;21(5):666-78. doi: 10.1016/j.chembiol.2014.02.021. Epub 2014 Apr 10. PMID: 24726831.
127. Laselva O, Molinski S, Casavola V, Bear CE. Correctors of the Major Cystic Fibrosis Mutant Interact through Membrane-Spanning Domains. *Mol Pharmacol.* 2018 Jun;93(6):612-618. doi: 10.1124/mol.118.111799. Epub 2018 Apr 4. PMID: 29618585.

128. He L, Kota P, Aleksandrov AA, Cui L, Jensen T, Dokholyan NV, Riordan JR. Correctors of  $\Delta F508$  CFTR restore global conformational maturation without thermally stabilizing the mutant protein. *FASEB J.* 2013 Feb;27(2):536-45. doi: 10.1096/fj.12-216119. Epub 2012 Oct 26. Erratum in: *FASEB J.* 2014 Jun;28(6):2737. PMID: 23104983; PMCID: PMC3545534.
129. Lopes-Pacheco M, Silva IAL, Turner MJ, Carlile GW, Sondo E, Thomas DY, Pedemonte N, Hanrahan JW, Amaral MD. Characterization of the mechanism of action of RDR01752, a novel corrector of F508del-CFTR. *Biochem Pharmacol.* 2020 Oct;180:114133. doi: 10.1016/j.bcp.2020.114133. Epub 2020 Jul 3. PMID: 32628927.
130. D'Ursi P, Uggeri M, Urbinati C, Millo E, Paiardi G, Milanese L, Ford RC, Clews J, Meng X, Bergese P, Ridolfi A, Pedemonte N, Fossa P, Orro A, Rusnati M. Exploitation of a novel biosensor based on the full-length human F508del-CFTR with computational studies, biochemical and biological assays for the characterization of a new Lumacaftor/Tezacaftor analogue Sensor. *Actuator. B Chem.*, 2019, 301, <https://doi.org/10.1016/j.snb.2019.127131>.
131. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W443-7. doi: 10.1093/nar/gkv315. Epub 2015 Apr 14. PMID: 25873628; PMCID: PMC4489249.
132. Rusnati M, D'Ursi P, Pedemonte N, Urbinati C, Ford RC, Cichero E, Uggeri M, Orro A, Fossa P. Recent Strategic Advances in CFTR Drug Discovery: An Overview. *Int J Mol Sci.* 2020 Mar 31;21(7):2407. doi: 10.3390/ijms21072407. PMID: 32244346; PMCID: PMC7177952.
133. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010 Jan 30;31(2):455-61. doi: 10.1002/jcc.21334. PMID: 19499576; PMCID: PMC3041641.
134. Dickson CJ, Madej BD, Skjevik AA, Betz RM, Teigen K, Gould IR, Walker RC. Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput.* 2014 Feb 11;10(2):865-879. doi: 10.1021/ct4010307. Epub 2014 Jan 30. PMID: 24803855; PMCID: PMC3985482
135. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE, Darden TA 3th, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg

- A, Sagui C, Salomon-Ferrer R, Seabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, Kollman PA. AMBER 14, University of California, 2014, San Francisco.
136. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009 Jun 2;10:168. doi: 10.1186/1471-2105-10-168. PMID: 19486540; PMCID: PMC2700099.
137. Fossa P, Uggeri M, Orro A, Urbinati C, Rondina A, Milanesi M, Pedemonte N, Pesce E, Padoan R, Ford RC, Meng X, Rusnati M, D'Ursi P. Virtual Drug Repositioning as a Tool to Identify Natural Small Molecules That Synergize with Lumacaftor in F508del-CFTR Binding and Rescuing. *Int J Mol Sci*. 2022 Oct 14;23(20):12274. doi: 10.3390/ijms232012274. PMID: 36293130; PMCID: PMC9602983.
138. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem*. 2005 Dec;26(16):1668-88. doi: 10.1002/jcc.20290. PMID: 16200636; PMCID: PMC1989667.
139. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015 Aug 11;11(8):3696-713. doi: 10.1021/acs.jctc.5b00255. Epub 2015 Jul 23. PMID: 26574453; PMCID: PMC4821407.
140. Newman WG, Woolf AS. Urofacial Syndrome. 2013 Aug 22 [updated 2018 Jun 7]. In: Adam MP, Everman DB, Mirzaa GM, Pagon RA, Wallace SE, Bean LH, Gripp KW, Amemiya A, editors. *GeneReviews*<sup>®</sup> [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2022. PMID: 23967498.
141. Ochoa B, Gorlin RJ. Urofacial (ochoa) syndrome. *Am J Med Genet*. 1987 Jul;27(3):661-7. doi: 10.1002/ajmg.1320270320. PMID: 3631137.
142. Woolf AS, Lopes FM, Ranjzad P, Roberts NA. Congenital Disorders of the Human Urinary Tract: Recent Insights From Genetic and Molecular Studies. *Front Pediatr*. 2019 Apr 11;7:136. doi: 10.3389/fped.2019.00136. PMID: 31032239; PMCID: PMC6470263.
143. Roberts NA, Woolf AS. Heparanase 2 and Urofacial Syndrome, a Genetic Neuropathy. *Adv Exp Med Biol*. 2020;1221:807-819. doi: 10.1007/978-3-030-34521-1\_35. PMID: 32274739.



144. Roberts NA, Woolf AS, Stuart HM, Thuret R, McKenzie EA, Newman WG, Hilton EN. Heparanase 2, mutated in urofacial syndrome, mediates peripheral neural development in *Xenopus*. *Hum Mol Genet*. 2014 Aug 15;23(16):4302-14. doi: 10.1093/hmg/ddu147. Epub 2014 Apr 1. PMID: 24691552; PMCID: PMC4103677.
145. Daly SB, Urquhart JE, Hilton E, McKenzie EA, Kammerer RA, Lewis M, Kerr B, Stuart H, Donnai D, Long DA, Burgu B, Aydogdu O, Derbent M, Garcia-Minaur S, Reardon W, Gener B, Shalev S, Smith R, Woolf AS, Black GC, Newman WG. Mutations in *HPSE2* cause urofacial syndrome. *Am J Hum Genet*. 2010 Jun 11;86(6):963-9. doi: 10.1016/j.ajhg.2010.05.006. Erratum in: *Am J Hum Genet*. 2010 Aug 13;87(2):309. PMID: 20560210; PMCID: PMC3032078.
146. Stuart HM, Roberts NA, Burgu B, Daly SB, Urquhart JE, Bhaskar S, Dickerson JE, Mermerkaya M, Silay MS, Lewis MA, Olondriz MB, Gener B, Beetz C, Varga RE, Gülpınar O, Süer E, Soygür T, Ozçakar ZB, Yalçinkaya F, Kavaz A, Bulum B, Gücük A, Yue WW, Erdogan F, Berry A, Hanley NA, McKenzie EA, Hilton EN, Woolf AS, Newman WG. *LRIG2* mutations cause urofacial syndrome. *Am J Hum Genet*. 2013 Feb 7;92(2):259-64. doi: 10.1016/j.ajhg.2012.12.002. Epub 2013 Jan 11. PMID: 23313374; PMCID: PMC3567269.
147. Aydogdu O, Burgu B, Demirel F, Soygur T, Ozcakar ZB, Yalcinkaya F, Tekgul S. Ochoa syndrome: a spectrum of urofacial syndrome. *Eur J Pediatr*. 2010 Apr;169(4):431-5. doi: 10.1007/s00431-009-1042-9. Epub 2009 Aug 11. PMID: 19669792.
148. Guo D, Holmlund C, Henriksson R, Hedman H. The *LRIG* gene family has three vertebrate paralogs widely expressed in human and mouse tissues and a homolog in *Ascidacea*. *Genomics*. 2004 Jul;84(1):157-65. doi: 10.1016/j.ygeno.2004.01.013. PMID: 15203213.
149. Simion C, Cedano-Prieto ME, Sweeney C. The *LRIG* family: enigmatic regulators of growth factor receptor signaling. *Endocr Relat Cancer*. 2014;21(6):R431-43. doi: 10.1530/ERC-14-0179. Epub 2014 Sep 2. PMID: 25183430; PMCID: PMC4182143.
150. Abraira VE, Satoh T, Fekete DM, Goodrich LV. Vertebrate *Lrig3*-ErbB interactions occur *in vitro* but are unlikely to play a role in *Lrig3*-dependent inner ear morphogenesis. *PLoS One*. 2010 Feb 1;5(2):e8981. doi: 10.1371/journal.pone.0008981. PMID: 20126551; PMCID: PMC2813878.

151. Xu Y, Soo P, Walker F, Zhang HH, Redpath N, Tan CW, Nicola NA, Adams TE, Garrett TP, Zhang JG, Burgess AW. LRIG1 extracellular domain: structure and function analysis. *J Mol Biol.* 2015 May 22;427(10):1934-48. doi: 10.1016/j.jmb.2015.03.001. Epub 2015 Mar 9. PMID: 25765764.
152. Xiao Q, Tan Y, Guo Y, Yang H, Mao F, Xie R, Wang B, Lei T, Guo D. Soluble LRIG2 ectodomain is released from glioblastoma cells and promotes the proliferation and inhibits the apoptosis of glioblastoma cells in vitro and in vivo in a similar manner to the full-length LRIG2. *PLoS One.* 2014 Oct 29;9(10):e111419. doi: 10.1371/journal.pone.0111419. PMID: 25353163; PMCID: PMC4213030.
153. Xiao Q, Dong M, Cheng F, Mao F, Zong W, Wu K, Wang H, Xie R, Wang B, Lei T, Guo D. LRIG2 promotes the proliferation and cell cycle progression of glioblastoma cells in vitro and in vivo through enhancing PDGFR $\beta$  signaling. *Int J Oncol.* 2018 Sep;53(3):1069-1082. doi: 10.3892/ijo.2018.4482. Epub 2018 Jul 16. Erratum in: *Int J Oncol.* 2019 Jun;54(6):2257. Erratum in: *Int J Oncol.* 2022 Feb;60(2): PMID: 30015847; PMCID: PMC6065455.
154. Holmlund C, Haapasalo H, Yi W, Raheem O, Brännström T, Bragge H, Henriksson R, Hedman H. Cytoplasmic LRIG2 expression is associated with poor oligodendroglioma patient survival. *Neuropathology.* 2009 Jun;29(3):242-7. doi: 10.1111/j.1440-1789.2008.00970.x. Epub 2009 Oct 20. PMID: 18992012.
155. Hedman H, Lindström AK, Tot T, Stendahl U, Henriksson R, Hellberg D. LRIG2 in contrast to LRIG1 predicts poor survival in early-stage squamous cell carcinoma of the uterine cervix. *Acta Oncol.* 2010 Aug;49(6):812-5. doi: 10.3109/0284186X.2010.492789. PMID: 20553099.
156. Wang B, Han L, Chen R, Cai M, Han F, Lei T, Guo D. Downregulation of LRIG2 expression by RNA interference inhibits glioblastoma cell (GL15) growth, causes cell cycle redistribution, increases cell apoptosis and enhances cell adhesion and invasion in vitro. *Cancer Biol Ther.* 2009 Jun;8(11):1018-23. doi: 10.4161/cbt.8.11.8375. Epub 2009 Jun 10. PMID: 19421009.
157. Hoesl C, Fröhlich T, Hundt JE, Kneitz H, Goebeler M, Wolf R, Schneider MR, Dahlhoff M. The transmembrane protein LRIG2 increases tumor progression in skin carcinogenesis. *Mol Oncol.* 2019 Nov;13(11):2476-2492. doi: 10.1002/1878-0261.12579. Epub 2019 Oct 21. PMID: 31580518; PMCID: PMC6822252.

158. van Erp S, van den Heuvel DMA, Fujita Y, Robinson RA, Hellemons AJCGM, Adolfs Y, Van Battum EY, Blokhuis AM, Kuijpers M, Demmers JAA, Hedman H, Hoogenraad CC, Siebold C, Yamashita T, Pasterkamp RJ. Lrig2 Negatively Regulates Ectodomain Shedding of Axon Guidance Receptors by ADAM Proteases. *Dev Cell*. 2015 Dec 7;35(5):537-552. doi: 10.1016/j.devcel.2015.11.008. PMID: 26651291.
159. Xu Y, Soo P, Walker F, Zhang HH, Redpath N, Tan CW, Nicola NA, Adams TE, Garrett TP, Zhang JG, Burgess AW. LRIG1 extracellular domain: structure and function analysis. *J Mol Biol*. 2015 May 22;427(10):1934-48. doi: 10.1016/j.jmb.2015.03.001. Epub 2015 Mar 9. PMID: 25765764.
160. Stuart HM, Roberts NA, Burgu B, Daly SB, Urquhart JE, Bhaskar S, Dickerson JE, Mermerkaya M, Silay MS, Lewis MA, Olondriz MB, Gener B, Beetz C, Varga RE, Gülpınar O, Sürer E, Soygür T, Ozçakar ZB, Yalçınkaya F, Kavaz A, Bulum B, Gücük A, Yue WW, Erdogan F, Berry A, Hanley NA, McKenzie EA, Hilton EN, Woolf AS, Newman WG. LRIG2 mutations cause urofacial syndrome. *Am J Hum Genet*. 2013 Feb 7;92(2):259-64. doi: 10.1016/j.ajhg.2012.12.002. Epub 2013 Jan 11. PMID: 23313374; PMCID: PMC3567269.
161. Fadda A, Butt F, Tomei S, Deola S, Lo B, Robay A, Al-Shakaki A, Al-Hajri N, Crystal R, Kambouris M, Wang E, Marincola FM, Fakhro KA, Cugno C. Two hits in one: whole genome sequencing unveils LIG4 syndrome and urofacial syndrome in a case report of a child with complex phenotype. *BMC Med Genet*. 2016 Nov 17;17(1):84. doi: 10.1186/s12881-016-0346-7. PMID: 27855655; PMCID: PMC5114772.
162. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011 Oct 11;7:539. doi: 10.1038/msb.2011.75. PMID: 21988835; PMCID: PMC3261699.
163. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018 Jul 2;46(W1):W296-W303. doi: 10.1093/nar/gky427. PMID: 29788355; PMCID: PMC6030848.

164. Esquivel-Rodriguez J, Filos-Gonzalez V, Li B, Kihara D. Pairwise and multimeric protein-protein docking using the LZerD program suite. *Methods Mol Biol.* 2014;1137:209-34. doi: 10.1007/978-1-4939-0366-5\_15. PMID: 24573484.
165. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model.* 2011 Oct 24;51(10):2778-86. doi: 10.1021/ci200227u. Epub 2011 Oct 5. PMID: 21919503.
166. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE 3th, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Huang Y, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, SalomonFerrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM Kollman PA, AMBER 2018, University of California, San Francisco, 2018.
167. Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci.* 2015 Jan;72(1):137-51. doi: 10.1007/s00018-014-1661-9. Epub 2014 Jun 18. PMID: 24939692.
168. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn.* 2012;30(2):137-49. doi: 10.1080/07391102.2012.675145. PMID: 22702725.
169. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. *Proteins.* 2003;53 Suppl 6:566-72. doi: 10.1002/prot.10532. PMID: 14579347.
170. DeForte S, Uversky VN. Resolving the ambiguity: Making sense of intrinsic disorder when PDB structures disagree. *Protein Sci.* 2016 Mar;25(3):676-88. doi: 10.1002/pro.2864. Epub 2016 Jan 9. PMID: 26683124; PMCID: PMC4815412.
171. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW,

- Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model*. 2001;19(1):26-59. doi: 10.1016/s1093-3263(00)00138-8. PMID: 11381529.
172. Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics*, 2019, 7:10. <https://doi.org/10.3389/fphy.2019.00010>.
173. Mammen M, Choi SK, Whitesides GM. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew Chem Int Ed Engl*. 1998 Nov 2;37(20):2754-2794. doi: 10.1002/(SICI)1521-3773(19981102)37:20<2754::AID-ANIE2754>3.0.CO;2-3. PMID: 29711117.
174. Romero P, Obradovic Z, Dunker AK. Natively disordered proteins: functions and predictions. *Appl Bioinformatics*. 2004;3(2-3):105-13. doi: 10.2165/00822942-200403020-00005. PMID: 15693736.
175. Beck M, Hurt E. The nuclear pore complex: understanding its function through structural insight. *Nat Rev Mol Cell Biol*. 2017 Feb;18(2):73-89. doi: 10.1038/nrm.2016.147. Epub 2016 Dec 21. PMID: 27999437.
176. Kalverda B, Pickersgill H, Shloma VV, Fornerod M. Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. *Cell*. 2010 Feb 5;140(3):360-71. doi: 10.1016/j.cell.2010.01.011. PMID: 20144760.
177. Liang Y, Hetzer MW. Functional interactions between nucleoporins and chromatin. *Curr Opin Cell Biol* 2011 Feb;23(1):65-70. doi: 10.1016/j.ceb.2010.09.008. Epub 2010 Oct 26. PMID: 21030234 PMCID: PMC3753814.
178. Pascual-Garcia P, Debo B, Aleman JR, Talamas JA, Lan Y, Nguyen NH, Won KJ, Capelson. Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. *Mol Cell*. 2017 Apr 6;66(1):63-76.e6. doi: 10.1016/j.molcel.2017.02.020. Epub 2017 Mar 30. PMID: 28366641.
179. Wozniak R, Burke B, Doye V. Nuclear transport and the mitotic apparatus: an evolving relationship. *Cell Mol Life Sci*. 2010 Jul;67(13):2215-30. doi: 10.1007/s00018-010-0325-7. Epub 2010 Apr 8. PMID: 20372967.

180. Chatel G, Fahrenkrog B. Dynamics and diverse functions of nuclear pore complex proteins. *Nucleus*. 2012 Mar 1;3(2):162-71. doi: 10.4161/nucl.19674. Epub 2012 Mar 1. PMID: 22555605; PMCID: PMC3383572.
181. Griffis ER, Altan N, Lippincott-Schwartz J, Powers MA. Nup98 is a mobile nucleoporin with transcription-dependent dynamics. *Mol Biol Cell*. 2002 Apr;13(4):1282-97. doi: 10.1091/mbc.01-11-0538. PMID: 11950939; PMCID: PMC102269.
182. Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, Tyagi S, Clarke J, Shammas SL, Blackledge M, Gräter F, Lemke EA. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell*. 2015 Oct 22;163(3):734-45. doi: 10.1016/j.cell.2015.09.047. Epub 2015 Oct 8. PMID: 26456112; PMCID: PMC4622936.
183. Lyngdoh DL, Nag N, Uversky VN, Tripathi T. Prevalence and functionality of intrinsic disorder in human FG-nucleoporins. *Int J Biol Macromol*. 2021 Apr 1;175:156-170. doi: 10.1016/j.ijbiomac.2021.01.218. Epub 2021 Feb 3. PMID: 33548309.
184. Jühlen R, Fahrenkrog B. Moonlighting nuclear pore proteins: tissue-specific nucleoporin function in health and disease. *Histochem Cell Biol*. 2018 Dec;150(6):593-605. doi: 10.1007/s00418-018-1748-8. Epub 2018 Oct 25. PMID: 30361777.
185. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol*. 2015 Jun;32:102-12. doi: 10.1016/j.sbi.2015.03.008. Epub 2015 Apr 2. PMID: 25863585; PMCID: PMC4512920.
186. Krishnan VV, Lau EY, Yamada J, Denning DP, Patel SS, Colvin ME, Rexach MF. Intramolecular cohesion of coils mediated by phenylalanine-glycine motifs in the natively unfolded domain of a nucleoporin. *PLoS Comput Biol*. 2008 Aug 8;4(8):e1000145. doi: 10.1371/journal.pcbi.1000145. PMID: 18688269; PMCID: PMC2475668.
187. Milles S, Mercadante D, Aramburu IV, Jensen MR, Banterle N, Koehler C, Tyagi S, Clarke J, Shammas SL, Blackledge M, Gräter F, Lemke EA. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell*. 2015 Oct 22;163(3):734-45. doi: 10.1016/j.cell.2015.09.047. Epub 2015 Oct 8. PMID: 26456112; PMCID: PMC4622936.
188. Pulianmackal AJ, Kanakousaki K, Flegel K, Grushko OG, Gourley E, Rozich E, Buttitta LA. Misregulation of Nucleoporins 98 and 96 leads to defects in protein synthesis that promote

- hallmarks of tumorigenesis. *Dis Model Mech.* 2022 Mar 1;15(3):dmm049234. doi: 10.1242/dmm.049234. Epub 2022 Mar 16. PMID: 35107131; PMCID: PMC8938402.
189. Thibodeau ML, Steinraths M, Brown L, Zong Z, Shomer N, Taubert S, Mungall KL, Ma YP, Mueller R, Birol I, Lehman A. Genomic and Cytogenetic Characterization of a Balanced Translocation Disrupting NUP98. *Cytogenet Genome Res.* 2017;152(3):117-121. doi: 10.1159/000479463. Epub 2017 Aug 31. PMID: 28854430.
190. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
191. Fleming PJ, Fleming KG. HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties. *Biophys J.* 2018 Feb 27;114(4):856-869. doi: 10.1016/j.bpj.2018.01.002. PMID: 29490246; PMCID: PMC5984988.
192. Kumar M, Gromiha MM, Raghava GP. Prediction of RNA binding sites in a protein using SVM and PSSM profile. *Proteins.* 2008 Apr;71(1):189-94. doi: 10.1002/prot.21677. PMID: 17932917.
193. Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des.* 2019 Jan;93(1):12-20. doi: 10.1111/cbdd.13388. Epub 2018 Oct 8. PMID: 30187647
194. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000 Jan 1;28(1):235-42. doi: 10.1093/nar/28.1.235. PMID: 10592235; PMCID: PMC102472.
195. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modeling of protein structures and complexes. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W296-W303. doi: 10.1093/nar/gky427. PMID: 29788355; PMCID: PMC6030848
196. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol.* 1993 Dec 5;234(3):779-815. doi: 10.1006/jmbi.1993.1626. PMID: 8254673.
197. RN Sahoo, S Pattanaik, G Pattnaik, S Mallick, R Mohapatra. Review on the use of Molecular Docking as the First Line Tool in Drug Discovery and Development. *Indian J Pharm Sci* 2022;84(5):1334-1337. doi: 10.36468/pharmaceutical-sciences.1031

198. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009 Jun 2;10:168. doi: 10.1186/1471-2105-10-168. PMID: 19486540; PMCID: PMC2700099
199. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*. 2011 Jun;7(2):146-57. doi: 10.2174/157340911795677602. PMID: 21534921; PMCID: PMC3151162
200. Yang C, Chen EA, Zhang Y. Protein-Ligand Docking in the Machine-Learning Era. *Molecules*. 2022 Jul 18;27(14):4568. doi: 10.3390/molecules27144568. PMID: 35889440; PMCID: PMC9323102
201. Spyrakis F, Cavasotto CN. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration, and active site water molecules description. *Arch Biochem Biophys*. 2015 Oct 1;583:105-19. doi: 10.1016/j.abb.2015.08.002. Epub 2015 Aug 10. PMID: 26271444
202. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron*. 2018 Sep 19;99(6):1129-1143. doi: 10.1016/j.neuron.2018.08.011. PMID: 30236283; PMCID: PMC6209097
203. Sinha S, Tam B, Wang SM. Applications of Molecular Dynamics Simulation in Protein Study. *Membranes (Basel)*. 2022 Aug 29;12(9):844. doi: 10.3390/membranes12090844. PMID: 36135863; PMCID: PMC9505860
204. Lazim R, Suh D, Choi S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int J Mol Sci*. 2020 Sep 1;21(17):6339. doi: 10.3390/ijms21176339. PMID: 32882859; PMCID: PMC7504087
205. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J Chem Phys*. 2004 Jun 22;120(24):11919-29. doi: 10.1063/1.1755656. PMID: 15268227
206. Gentile F, Yaacoub JC, Gleave J, Fernandez M, Ton AT, Ban F, Stern A, Cherkasov A. Artificial intelligence-enabled virtual screening of ultra-large chemical libraries with deep docking. *Nat Protoc*. 2022 Mar;17(3):672-697. doi: 10.1038/s41596-021-00659-2. Epub 2022 Feb 4. PMID: 35121854.



207. Kimber TB, Chen Y, Volkamer A. Deep Learning in Virtual Screening: Recent Applications and Developments. *Int J Mol Sci.* 2021 Apr 23;22(9):4435. doi: 10.3390/ijms22094435. PMID: 33922714; PMCID: PMC8123040.
208. Kozlovskii I, Popov P. Spatiotemporal identification of druggable binding sites using deep learning. *Commun Biol.* 2020 Oct 27;3(1):618. doi: 10.1038/s42003-020-01350-0. PMID: 33110179; PMCID: PMC7591901.
209. Dhakal A, McKay C, Tanner JJ, Cheng J. Artificial intelligence in the prediction of protein-ligand interactions: recent advances and future directions. *Brief Bioinform.* 2022 Jan 17;23(1):bbab476. doi: 10.1093/bib/bbab476. PMID: 34849575; PMCID: PMC8690157
210. Mouchlis VD, Afantitis A, Serra A, Fratello M, Papadiamantis AG, Aidinis V, Lynch I, Greco D, Melagraki G. Advances in de Novo Drug Design: From Conventional to Machine Learning Methods. *Int J Mol Sci.* 2021 Feb 7;22(4):1676. doi: 10.3390/ijms22041676. PMID: 33562347; PMCID: PMC7915729
211. Yang C, Chen EA, Zhang Y. Protein-Ligand Docking in the Machine-Learning Era. *Molecules.* 2022 Jul 18;27(14):4568. doi: 10.3390/molecules27144568. PMID: 35889440; PMCID: PMC9323102
212. Lazim R, Suh D, Choi S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int J Mol Sci.* 2020 Sep 1;21(17):6339. doi: 10.3390/ijms21176339. PMID: 32882859; PMCID: PMC7504087