

Computational Studies of the Proton-Coupled Metal Ion Transport in the SLC11/NRAMP Family of Transporters

Présentée le 1er juillet 2022

Faculté des sciences de base
Laboratoire de chimie et biochimie computationnelles
Programme doctoral en chimie et génie chimique

pour l'obtention du grade de Docteur ès Sciences

par

Maria Letizia MERLINI

Acceptée sur proposition du jury

Prof. S. Gerber, présidente du jury
Prof. U. Röthlisberger, directrice de thèse
Prof. P. Carloni, rapporteur
Prof. A. Magistrato, rapporteuse
Prof. M. Dal Peraro, rapporteur

"He who opens a school door closes a prison."

Victor Hugo

To Sara and Enrico...

Acknowledgements

I would like to thank first my scientific advisor, Prof. Ursula Röthlisberger. She gave me the opportunity to work as a Ph.D. student in her group, she always kept believing in me and supported me in the many difficulties of which is paved this journey. I am grateful for the challenging project she assigned me as well as for the scientific mindset she contributed to build in our fruitful scientific discussions. Of course, I would also like to thank her loyal partners, Ruby, Quanta and Mocca. Their company as well as the wave of joy they always bring with them have been fundamental so survive hard days.

I also would like to acknowledge Prof. Paolo Carloni and Dr. Alessandra Magistrato for the fruitful discussions on the conceivment of this project as well as for their constant support.

A special thanks also goes to Dr. Elisa Liberatore and Dr. Daniele Narzi for the invaluable help during the first years of my Ph.D..

Finally, I would like to thank all the big LCBC family: Karin (what would LCBC be without her?!) and all of the colleagues with which I had the pleasure to share a long or short part of this journey: Ariadni, Thibaud, Polydefkis, Elisa, Daniele, Marta, Esra, Murat, Siri, Nick, Swarnendu, Martin, Paramvir, Mathias, Guido, Farzaneh, Marko, Pablo, François, Justin, Slava, Simon, Andrea, Irea, Virginia, Nico, Sophie, Lorenzo and Andrej. They all made my stay at LCBC unforgettable.

Lausanne, 18 April 2022

M. L. M.

Abstract

In the last years, experimental evidences have demonstrated a link between the overload of metal ions inside nervous system cells and the onset of severe and eventually lethal neurodegenerative diseases, such as Parkinson and Alzheimer disease. This has prompted the investigation of the structural and functional properties of transporters responsible for the uptake and regulation of metal ions inside cells. In particular, members of the Nrap/SLC11 ("natural resistance-associated macrophage proteins"/"solute carrier 11") family of transporters tightly regulate the influx of divalent transition-metal ions across cellular membranes thus contributing to the preservation of cellular homeostasis. In these secondary active transporters, the translocation of divalent transition metal ions from the extracellular matrix to the cellular cytoplasm is coupled with transport of protons in the same direction (symport). The co-transport is realized by using an "alternate access mechanism", in which the protein is alternately open towards the extracellular matrix, ("outward facing conformation") or towards the cellular cytoplasm ("inward facing conformation").

Although a wide range of functional studies as well as the crystallographic characterization of prokaryotic members of the family provided fundamental insights to the comprehension of this proton-coupled transport, most mechanistic details are still unknown. As an example, the identification of the complete coordination sphere of the divalent transition-metal ion in the protein binding site represents a crucial information for the understanding of the transport mechanism. However, the coordination sphere of the metal ion in the inward-facing conformation of the protein has only been partially resolved by crystallographic experiments and later attempts to provide comprehensive descriptions led to contradictory models. Furthermore, several experimental evidences proved that proton uniport in the absence of the metal ion substrate, is also possible. This suggests a noncanonical symporter behaviour for Nrap transporters, in which the two substrates are not tightly coupled. Despite these recent advances in the comprehension of Nrams proton transport, the underlying mechanistic details remain still ambiguous, as demonstrated by different and inconsistent mechanistic models proposed during the last years.

In this thesis, computational methods have been used in order to provide new molecular level insights in the proton-coupled divalent transition metal ions transport performed by Nrams.

Thanks also to the rapid growth of computer power, computational biochemistry has become nowadays a very useful tool able to provide fundamental details on the mechanistic nature of many biological processes and assist experimental techniques in an new advantageous synergistic approach to research.

In particular, a combination of classical molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) *ab initio* Born-Oppenheimer and Car-Parrinello MD methods have been applied in order to refine the partially resolved coordination sphere of the Mn^{2+} ion in the binding active site of *Staphylococcus Capitis* Divalent Metal-ion Transporter (ScaDMT), which represents the only available crystal structure of a substrate-bound member of the family in an inward-facing conformation. This QM/MM multiscale approach, using an accurate quantum treatment of the active site at the density functional theory (DFT) level and a classical treatment of the surroundings, allowed for state-of-the-art calculations of the entire system that led to the unequivocal identification of the metal's coordination sphere.

Furthermore, MD simulations have been performed for both the inward-facing conformation ScaDMT and the other prokaryotic member of the family *Eremococcus Coleocola* DMT (EcoDMT), whose crystal structure has been obtained in an outward-facing conformation. Since experimental evidences proved that proton transport can occur even in the absence of the metal ion substrate, we performed simulations of both proteins in their apo forms. In particular, in order to assess the potential role played by a conserved histidine residue in the proton transport, for each of the two proteins we performed MD simulations in unprotonated and protonated state of this residue. These simulations allowed us to characterize four amino acidic residues identified as likely key players in the proton transport process and to determine which role they assume in the process. Finally we were able to identify the likely primary proton acceptor and proposed a potential route followed by protons in their influx toward the intracellular cytoplasm.

Key words: Nramp transporters, secondary active transport, transition metal ions, proton transport, symport, uniport, alternate access mechanism, molecular dynamics, QM/MM, Car-Parrinello, Born-Oppenheimer, coordination sphere

Abstract

Negli ultimi anni, molteplici evidenze scientifiche hanno messo in relazione l'accumulo di ioni metallici all'interno delle cellule del sistema nervoso con l'insorgenza di malattie neurodegenerative gravi e talvolta letali, come ad esempio il Morbo di Parkinson e il Morbo di Alzheimer. Ciò ha dato un forte impulso allo studio delle proprietà funzionali e strutturali dei trasportatori responsabili dell'assorbimento e regolazione di ioni metallici all'interno delle cellule. In particolare, i membri della famiglia di trasportatori Nramp/SLC11 ("natural resistance-associated macrophage proteins" / "solute carrier 11") regolano strettamente il trasporto di ioni divalenti dei metalli di transizione attraverso le membrane cellulari contribuendo così al mantenimento dell'omeostasi cellulare. In questi trasportatori attivi secondari, il trasferimento di ioni divalenti dei metalli di transizione dalla matrice extracellulare al citosplasma cellulare è accoppiato col trasporto di protoni nella stessa direzione (simporto). Il co-transporto è realizzato attraverso un meccanismo di "accesso alternato" in cui la proteina è alternativamente aperta verso la matrice extracellulare ("outward facing conformation") o verso il citoplasma ("inward-facing conformation").

Nonostante un'ampia gamma di studi funzionali e la caratterizzazione cristallografica di alcuni membri procariotici della famiglia abbiano fornito indizi fondamentali per la comprensione di questo trasporto accoppiato, molti dettagli meccanicistici restano ancora sconosciuti. Ad esempio, l'identificazione della completa sfera di coordinazione dello ione divalente dei metalli di transizione nel sito attivo della proteina rappresenta un'informazione cruciale per la comprensione del meccanismo di trasporto. Tuttavia, la sfera di coordinazione dello ione metallico nella conformazione inward-facing della proteina è stata solo parzialmente risolta attraverso le tecniche cristallografiche e i successivi tentativi di fornirne una descrizione completa hanno portato alla formulazione di modelli inconsistenti tra loro. Inoltre, alcune osservazioni sperimentali hanno provato che, in assenza del substrato metallico, anche un fenomeno di uniporto protonico è possibile. Ciò suggerisce per i trasportatori Nramp un simporto non canonico in cui i due substrati non sono strettamente accoppiati. Nonostante i recenti passi in avanti nella comprensione del trasporto protonico, i dettagli meccanicistici rimangono sinora ambigui, come testimoniato dai contrastanti modelli meccanicistici proposti negli ultimi anni.

In questa tesi abbiamo utilizzato vari metodi computazionali allo scopo di fornire nuovi indizi a livello molecolare riguardo al trasporto di ioni metallici accoppiato con trasporto protonico effettuato dai trasportatori Nramp. Anche grazie alla rapida crescita delle prestazioni dei computers la biochimica computazionale rappresenta ad oggi uno strumento molto utile, capace di fornire fondamentali dettagli sulla natura meccanicistica di molti processi biologici e in grado di assistere le tecniche sperimentali in un nuovo vantaggioso e sinergistico approccio alla ricerca.

In particolare, una combinazione di dinamica molecolare (MD) e QM/MM basato sulla dinamica molecolare *ab initio* Born-Oppenheimer e Car-Parrinello sono stati utilizzati allo scopo di perfezionare la sfera di coordinazione, solo parzialmente risolta, dello ione Mn^{2+} nel sito attivo del trasportatore *Staphylococcus Capitis* Divalent Metal-ion Transporter (ScaDMT) che rappresenta ad oggi la sola struttura cristallina di un membro della famiglia nella conformazione inward-facing e con il substrato legato. L'approccio multiscale del QM/MM, che utilizza un'accurato trattamento a livello DFT nella zona del sito attivo e un trattamento a livello di dinamica classica per il rimanente sistema, ha permesso calcoli all'avanguardia sull'intero sistema che hanno portato all'inequivocabile identificazione della sfera di coordinazione del metallo.

Inoltre, simulazioni di dinamica molecolare sono state eseguite sia sulla conformazione inward-facing di ScaDMT sia sull'altro membro procariotico della famiglia *Eremococcus Coleocola* DMT (EcoDMT), la cui struttura cristallina è stata ottenuta nella conformazione outward-facing. Dato che molteplici evidenze sperimentali hanno dimostrato che il trasporto protonico può avvenire anche in assenza del substrato metallico, noi abbiamo eseguito simulazioni di entrambe le proteine nella loro forma apo. In particolare, allo scopo di individuare il potenziale ruolo giocato nel trasporto protonico da un residuo istidinico conservato nella famiglia, per ciascuna delle due proteine abbiamo eseguito simulazioni MD nella forma protonata e deprotonata di questo residuo. Queste simulazioni ci hanno permesso di caratterizzare quattro residui amminoacidici identificati come fondamentali nel trasporto protonico e determinare il ruolo da essi giocato nel processo. Infine è stato possibile identificare il probabile accettore primario del protone e il probabile cammino seguito dai protoni nel loro influsso nel citoplasma intracellulare.

Key words: trasportatori Nramp, trasporto secondario attivo, ioni dei metalli di transizione, trasporto protonico, simporto, uniporto, meccanismo ad accesso alternato, dinamica molecolare, QM/MM, Car-Parrinello, Born-Oppenheimer, sfera di coordinazione

Contents

Acknowledgements	i
Abstract (English)	iii
Abstract (Italian)	v
List of figures	ix
List of tables	xix
1 The Nramp family of transporters	1
1.1 The crucial role of Nramp transporters in living organisms	1
1.2 Structural and functional properties of Nramps	5
1.2.1 Substrate selectivity	5
1.2.2 Structural features	6
1.2.3 Currently available crystal structures	6
1.2.4 The proton-coupled alternate-access mechanism	7
1.2.5 Metal-ion binding sites	9
1.2.6 Proton transport in the Nramp family of transporters	13
References	21
2 Theoretical methods	33
2.1 Density Functional Theory	33
2.1.1 Hohenberg-Kohn Theorems	34
2.1.2 Kohn-Sham Equations	35
2.1.3 Exchange-Correlation Functionals	37
2.2 Classical Molecular Dynamics	39
2.3 <i>Ab Initio</i> Molecular Dynamics (AIMD)	44
2.3.1 From the full quantum problem to AIMD	44
2.3.2 Born-Oppenheimer Molecular Dynamics	49
2.3.3 Car-Parrinello Molecular Dynamics	51
2.4 Hybrid Quantum Mechanics/Molecular Mechanics Simulations	53
2.4.1 Theoretical background of QM/MM	56
2.4.2 Pitfalls of the QM/MM method	59
References	63

3	Coordination of Mn(II) in SLC11/NRAMP transporters by QM/MM	73
3.1	Abstract	74
3.2	Introduction	74
3.3	Computational Details	78
3.4	Results	81
3.5	Conclusions	88
	References	91
4	Computational studies provide new insights into Nramps proton transport	99
4.1	Abstract	100
4.2	Introduction	100
4.3	Computational details	103
4.4	Results	107
4.5	Conclusions	115
	References	119
	Conclusions and outlook	125
A	Supporting Information - Coordination of Mn(II) in SLC11/NRAMP transporters uncovered by QM/MM simulations	127
	References	129
B	Supporting Information - Computational studies provide new insights into Nramps proton transport	131
	Bibliography	135
	Curriculum Vitae	135

List of Figures

- 1.1 Schematic representation of primary (**left**) and secondary (**right**) active transport. The lipid bilayer (gray) represents the cellular membrane while the primary and secondary active transporters are depicted in blue and violet, respectively. The primary active transporter powers the transport of its substrate (yellow spheres) by using an external source of energy (e.g. the energy released from the ATP hydrolysis reaction). The electrochemical gradient created by the transport of the driving ions (yellow spheres) is harnessed by the secondary active transporter to realize the cotransport of the intended substrate (green spheres). 2
- 1.2 Schematic representation illustrating the difference between uniport, symport and antiport mechanisms. In uniport a single substrate is transported across the membrane in a unique direction. When two different substrates are instead transported simultaneously through the membrane we refer to as cotransport. Cotransport is typical of secondary active transporters and can be of two types: symport and antiport. In the first one, the two substrates are transported across the membrane in the same direction while in the latter they are transported in opposite directions. 3
- 1.3 Nramp secondary structure displaying the characteristic topology of the Leu-T fold. The two pseudosymmetric and intertwining repeats composed by α -helices 1-5 and 6-10 are represented in magenta and green, respectively. The 11th helix, excluded from the Leu-T-like fold, is represented in orange while the possible 12th helix is represented in gray. The unwound regions of α -helices 1 and 6 host the metal ion binding site, represented as a gray sphere. 6

- 1.4 Schematic representation of the putative transport mechanism of Nramp transporters. The membrane bilayer is represented in gray, the extracellular matrix and the cellular cytoplasm are colored in light blue, the two proteic subdomains are represented in magenta and green while the transition metal ion (Mn^{2+} in this case) is depicted as a gray sphere. Structure 1 represents the metal-bound outward-facing conformation (OFC) (DraDMT- Mn^{2+} , PDB ID: 6BU5 [124]); 3 represents the metal-bound inward-facing conformation (IFC) (ScaDMT- Mn^{2+} , PDB IDs: 4WGW/5M95 [121, 122]); 4 represents the apo IFC (apo-ScaDMT and apo-DraDMT, PDB IDs: 4WGV/5M94 [121, 122] and 5KTE/6D9W [123, 124], respectively); 5 is the apo inward-occluded configuration (IOC) (apo IOC-DraDMT, PDB ID: 6C3I, [124]); 6 represents the apo OFC (apo-EcoDMT and apo-DraDMT, PDB IDs: 5M87 [122] and 6D91 [124], respectively). Structure 2, framed by a dashed line, is the hypothetical outward-occluded conformation (OOC) for which a crystal structure is still missing [124]. 9
- 1.5 Schematic representation of the main rearrangements of α -helices during the conformational change from the OFC to the IFC states. α -helices undergoing more significant reorientations (i.e. 1, 5, 6 and 10) are represented in orange while the relatively stationary ones (i.e. 2, 3, 4, 7, 8 and 9) are represented as light blue blocks. The metal-ion substrate is represented as a gray sphere. 10
- 1.6 Metal-ion binding site of a metal-bound Nramp protein in its inward-facing conformation, as obtained by Ehrnstorfer and colleagues in the IFC-ScaDMT- Mn^{2+} crystal structure [121]. The picture has been obtained with VMD. 11
- 1.7 Schematic representation of the metal-ion binding site of a metal-bound Nramp protein in its outward-facing conformation, as observed by Bozzi and colleagues in the OFC-DraDMT- Mn^{2+} crystal structure [124]. 11
- 1.8 Schematic representation of the metal ion coordination spheres for IFC-ScaDMT as obtained from the theoretical study of Pujol-Giménez and colleagues [144] (**left**) and as speculated by Bozzi and colleagues [124] (**right**). The Cd^{2+} and Mn^{2+} ions are represented as pink and gray spheres, respectively. In both panels TM stands for transmembrane α -helix and the O, N, and S atoms are represented in red, blue and orange, respectively. 13
- 1.9 MD snapshot of a configuration of the IFC ScaDMT protein in complex with its substrate (a Mn^{2+} ion) for which we performed a combination of classical MD and *ab initio* QM/MM MD simulations. The protein α -helices are represented as magenta ribbons, the Mn^{2+} ion is represented as a gray sphere while the four amino acidic residues identified as ligands in the X-ray structure [121] are represented as sticks. The picture has been obtained using VMD [145, 146]. . . 14

- 1.10 Schematic representation of the proton transport mechanism as hypothesized by Ehrnstorfer and colleagues [122]. In the OFC (**top**), Asp51 and His236 are both in close contact with the aqueous cavity (represented in light blue). The proton (represented in orange) binds to His236, which acts as the primary proton acceptor. After the conformational change to the IFC (**bottom**) the proton bound to His236 can be released into the cytoplasm either directly, through the wide inward-directed aqueous cavity or via the narrow channel. The two possible exit pathways are indicated as orange dashed arrows. TM1 and TM3, belonging to the first protein subdomain, are represented in magenta while TM6 and TM9, belonging to the second protein subdomain, are represented in green. The metal ion (Mn^{2+}) is represented as a gray sphere. Within the amino acidic residues, O and N atoms are represented in red and blue, respectively, while C and H atoms are represented in black. The residue numberings correspond to those of EcoDMT and ScaDMT in the **top** and **bottom** panel, respectively. The pictures have been obtained with Biorender [151]. 16
- 1.11 Schematic representation of the overall transport mechanism as described by Pujol-Giménez and colleagues [144]. (a) Both substrates bind in the high-affinity OFC, where Glu127 is in a high-pKa state and accepts the proton. (b) After a conformational switch to an inward-occluded state, the proton is released by Glu127 likely along the narrow water-filled channel established between TM3 and TM9. (c) The consequent negative charge on the Glu127 is the signal that triggers intracellular gate opening while, simultaneously, a partial unwinding of TM8 confers high flexibility to Glu127. (d) In the IFC, the metal ion is solvated and released to the cytoplasm. (e) Glu127 reorients itself into its high-pKa state and the gate closes, thus recovering the initial OFC state (f). The proton and the divalent transition metal ion are represented as orange and gray spheres, respectively. Their respective charge is also indicated. TM1 and TM3 and TM8 and TM9 are represented in magenta and green, respectively. The solvent accessible region of the protein is represented in light blue. The pictures have been obtained using the Biorender software [151]. 18
- 1.12 Proton uniport (**top**) and metal ion/proton symport (**bottom**) as hypothesized by Bozzi and colleagues. In proton uniport, H^+ enter via the external vestibule and are released through the narrow side channel, without the need for the intracellular gate to open (i.e. without the need for a conformational change from the OFC to the IFC). The metal ion/proton symport requires instead the opening of the gate, and, consequently, the conformational switch from OFC to IFC. In this case the two substrates enter both via the external vestibule but, after binding, take two distinct routes towards the cytoplasm. The two proteic subunits are represented in green and magenta, the membrane is represented in gray while the metal ion and the proton are represented as gray and black spheres, respectively. The picture hasv been created with Biorender [151]. . . . 20

1.13	MD snapshots reporting two configurations of the IFC apo-ScaDMT (left) and the OFC apo-EcoDMT (right) proteins used for classical MD simulations [121, 122]. As before, the α -helices of the two proteins are represented as magenta ribbons. The two pictures have been obtained using VMD [145, 146].	20
2.1	Schematic representation of the bonded and nonbonded interaction terms comprised in molecular mechanics force fields. The picture has been obtained using VMD. [36, 37].	41
2.2	Two-dimensional schematic illustration of the Periodic Boundary Condition (PBC) scheme. The purple and green spheres represent the static and moving particles, respectively. The green dashed sphere represents the arrival point of the green sphere and the black arrow represents the movement direction. As it is possible to see, the same system in the central gray-shaded box is replicated in each of the mirror boxes.	44
3.1	Schematic representation of the putative metal ion transport alternate access mechanism [49]. It involves at least six distinct conformational states, five of which have been identified and (at least partially) crystallized for different systems: (1) Mn^{2+} -bound outward-facing conformation (for DraDMT with a G223W mutant) [49]; (2) Mn^{2+} -bound outward-occluded conformation (no crystal structure available) ; (3) Mn^{2+} -bound inward-facing conformation (ScaDMT) [46]; (4) apo inward-facing conformation (apo ScaDMT [46] and apo DraDMT [48]); (5) apo inward-occluded conformation (DraDMT as G45R mutant) [49]; and (6) apo outward-facing conformation (apo DraDMT as G223W mutant [49] and EcoDMT [47]). The conformational state (2) is still missing. The two symmetrically related halves of the protein (α -helices 1-5 and 6-10) are colored in green and magenta, respectively while the Mn^{2+} ion is represented as a blue sphere. The cellular membrane is represented in light gray whereas the extracellular matrix and the cytoplasm are represented in light blue.	76
3.2	ScaDMT- Mn^{2+} topology (a) and structure (b). Protein helices are represented as cylinders while the Mn^{2+} ion is depicted as a blue sphere. The two halves of the protein related by a pseudo-symmetric relationship are colored in magenta (α -helices 1-5) and green (α -helices 6-10) respectively; the additional α -helix 11 is colored in orange. The helix numbers are labeled in (a) in order to schematically show the protein fold. The representation of the structure in (b) has been obtained using VMD [60, 61].	77
3.3	Structure of ScaDMT- Mn^{2+} in its inward-facing conformation (left) and detail of its metal ion binding site (right) as obtained from X-ray crystallography [46]. The protein α -helices are colored in magenta, Mn^{2+} is represented as a blue sphere and the four coordinating amino acid residues are represented in licorice style. C atoms are colored in cyan, O atoms in red, N atoms in blue, S in yellow and H atoms in white. The water accessible volume close to the active site is represented as a cyan mesh. Both pictures have been obtained with VMD [60, 61].	78

- 3.4 Frontal view of the whole computational model comprising the three component CHL-POPC-POPE membrane, the ScaDMT-IFC protein embedded into the lipid bilayer and in complex with the Mn(II) cation and the pool of solvent (H_2O) and monovalent ions K^+ and Cl^- around the membrane (**left**); zoom into the membrane-embedded ScaDMT-IFC- Mn^{2+} protein (**top right**); magnification of the Mn^{2+} active site region (**bottom right**). All the pictures have been obtained using VMD [60, 61]. 80
- 3.5 Distances between Mn^{2+} and the molecules approaching its coordination sphere during the entire classical MD simulation (thermalization: 600ns + production: 1000 ns). Distances are reported in Å, time is reported in ps. The distances between Mn^{2+} and the four ligands identified in the crystal structure Asp49, Asn52, Ala223 and Met226 are represented in orange, yellow, turquoise and magenta, respectively. The fluctuations in their bond length values are smaller with respect to that of the other molecules as these four ligands were kept restrained in their positions during the entire simulation. The four water molecules WAT56324, WAT1701, WAT24524 and WAT3067 are represented as blue, red, light blue and purple lines, respectively. Lastly, distances between Mn^{2+} and Gln389, Gly46 and Asp49($\text{O}^{\delta 1}$) are represented as green, pink and dark violet lines, respectively. 82
- 3.6 Snapshot of the Mn^{2+} coordination sphere as obtained from the classical MD simulations (**left**) and from QM/MM MD simulations (**right**). In both cases ScaDMT α -helices are depicted as magenta ribbons. Asp49, Asn52, Ala223 and Met226 are represented as sticks, water molecules as balls and sticks model while Mn^{2+} is represented as a blue sphere. C atoms are colored in cyan, O in red, N in blue, S in yellow and H in white. 83
- 3.7 Fluctuations of the bond lengths for each of the six Mn^{2+} -coordinating ligands: Mn^{2+} -Asp49($\text{O}^{\delta 2}$) (orange), Mn^{2+} -Asn52($\text{O}^{\delta 1}$) (yellow), Mn^{2+} -Ala223(O) (turquoise), Mn^{2+} -Met226(S^{δ}) (magenta), Mn^{2+} -WAT1 (blue) and Mn^{2+} -WAT2 (red). Mn^{2+} -WAT3 distance is also represented in purple. The straight lines in the same color of the corresponding curve represent the average values during the whole 29 ps long CPMD QM/MM trajectory. 84
- 3.8 (a) Expanded view of the region around the protein active site. The position of the three atoms Gln389($\text{O}^{\epsilon 1}$), Gly46(O) and Asp49($\text{O}^{\delta 1}$) with respect to the Mn^{2+} ion is highlighted by the labels and the black dashed lines. (b) Oscillations of Mn^{2+} -Asp49($\text{O}^{\delta 1}$) (dark violet), Mn^{2+} -Gln389($\text{O}^{\epsilon 1}$) (green) and Mn^{2+} -Gly46(O) (pink) distances over the whole QM/MM CPMD trajectory (29 ps). Horizontal lines represent the average values. 86

- 3.9 (a) Representation of the hydrogen bonds linking WAT1 hydrogens with Gln389(O^{ε1}) and WAT41196(O). Hydrogen bonds are indicated as blue and orange dashed lines, respectively. (b) Representation of the hydrogen bond between WAT1 and one of the Asp49 carboxylic oxygens as a turquoise dashed line. The percentages of persistence of each of these hydrogen bonds are also given in the same color. Mn²⁺ is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style. 87
- 3.10 (a) (b) (c) Representation of the hydrogen bond linking WAT2 and Val45(O) as a red dashed line with the persistence percentage expressed in the same color. Representation of the hydrogen bonds linking WAT2 and either WAT5 (a), WAT9704 (b) or WAT4458 (c) as coral, steel blue and light sea green dashed lines respectively, together with their persistence percentages in the same color. Mn²⁺ is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style. 88
- 3.11 (a) Representation of the hydrogen bonds connecting Gly46 to Asn272 and WAT5 as violet and light blue dashed lines, respectively. (b) Representation of the hydrogen bonds formed by the two carboxylic oxygens of Asp49. In particular, Asp49(O^{δ2}) forms hydrogen bonds with either WAT41196 or WAT42740 (magenta and green dashed lines respectively) while Asp49(O^{δ1}) forms hydrogen bonds with either WAT41196, WAT1319 or the Gly51 residue (magenta, brown and purple dashed lines, respectively). The persistence percentage of each hydrogen bond is given in the correspondent color. Mn²⁺ is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style. 88
- 3.12 Schematic depiction of the Mn²⁺ coordination sphere as described by Pujol-Giménez and co-authors [105] (left) and Bozzi and co-authors (right) [49]. In the former, the Cd²⁺ metal ion has an octahedral coordination with the four residues observed in the X-ray structure (namely Asp49, Asn52, Ala223 and Met226) plus the side chain of Gln389 and the backbone carbonyl oxygen of Gly46 acting as ligands. In the latter instead, besides the four residues observed in the crystal structure the metal ion is coordinated by two water molecules. However, Asp49 acts as a bidentate ligand imparting therefore a heptahedral coordination. The transmembrane α -helices are represented in light gray, the Cd²⁺ and Mn²⁺ ions are represented as pink and gray spheres while N, O, and S atoms are represented in blue, red and orange, respectively. 89

- 4.1 Schematic representation of the conserved network of protonatable residues potentially involved in the proton transfer [7, 10, 22, 23]. Transmembrane α -helices TM1 and TM3, belonging to the first protein subdomain, and TM6 and TM9, belonging to the second protein subdomain, are represented as magenta and green ribbons, respectively. Amino acidic residues are represented in a stick style, with C, N, O and H atoms represented in cyan, blue, red and white, respectively. The Asp residue on the unwound region of TM1, together with the His residue on TM6b and the Glu on TM3 connect the metal binding site, located at the unwound regions of TM1 and TM6, with the narrow hydrophilic channel formed by TM3 and TM9, linked together by two salt bridges between Asp and Glu residues located on TM3 and two Arg located on TM9. The figure has been obtained using VMD [29, 30]. 102
- 4.2 Schematic representation of the proton transport mechanism as hypothesized by Ehrnstorfer and colleagues [22]. The residue numbering corresponds to those of EcoDMT and ScaDMT [25] in the **top** and **bottom** panel, respectively. In the OFC (**top**), Asp51 and His236 are both in close contact with the aqueous cavity (represented in light blue). The proton (represented in orange) binds to His236, which acts as the primary proton acceptor. After the conformational change to the IFC (**bottom**) the proton can be released into the cytoplasm either directly, through the wide inward-directed aqueous cavity, which is also the metal ion exit pathway, or via the narrow water-filled channel formed by TM3 and TM9. The two possible exit pathways are indicated as orange dashed arrows. TM1 and TM3, belonging to the first protein subdomain, are represented in magenta while TM6 and TM9, belonging to the second protein subdomain, are represented in green. The metal ion (Mn^{2+}) is represented as a gray sphere. Within the amino acidic residues, O and N atoms are represented in red and blue, respectively, while C and H atoms are represented in black. The pictures have been obtained with Biorender [31]. 104

- 4.3 Schematic representation of the overall transport mechanism as described by Pujol-Giménez and colleagues [10]. **(a)** The substrates both bind in the high-affinity OFC, where Glu127 is in a high-pKa state and accepts the proton. **(b)** After a conformational switch to an inward-occluded state, the proton is released by Glu127 likely along the narrow channel formed by TM3 and TM9. **(c)** The negative charge on Glu127 after the proton release is the "signal" that triggers the intracellular gate opening while, simultaneously, a partial unwinding of TM8 confers high flexibility to Glu127. **(d)** In the IFC the metal ion is solvated and released into the cytoplasm. **(e)** Glu127 reorients itself into the high-pKa state and the gate closes, thus recovering the initial OFC state **(f)**. The proton and the divalent transition metal ion are represented as orange and gray spheres, respectively. Their charge is reported next to it. TM1 and TM3 and TM8 and TM9 are represented in magenta and green, respectively. The solvent accessible region of the protein is represented in light blue. The pictures have been obtained using Biorender [31]. 105
- 4.4 Examples of water bridges. **(a)** Water bridge formed between two hydrogen bond acceptors; **(b)** water bridge formed between two hydrogen bond donors; **(c)** water bridge formed between a hydrogen bond donor and a hydrogen bond acceptor; **(d)** second order water bridge composed by two water molecules. . . 107
- 4.5 pKa values of residues Asp51 (violet), Glh129 (red), Asp126 (turquoise) and His236 (blue) during the OFC trajectory ($1 \mu s = 1000 ns$) with His236 in its deprotonated form. Time is reported in ns. Residue numbering from OFC. 108
- 4.6 pKa values of residues Asp49 (violet), Glh127 (red), Asp124 (turquoise) and His228 (blue) during IFC trajectory ($1 \mu s = 1000 ns$) with His228 in its deprotonated form. Time is reported in ns. 110
- 4.7 **(left)** Asp51, Glh129, Asp126 and His236 Solvent Accessible Surface Area (SASA) values over time ($1 \mu s = 1000 ns$) in OFC trajectory. **(right)** Asp49, Glh127, Asp124 and His228 SASA values over time in IFC trajectory. SASA values are given in nm^2 while time is reported in ps. 112
- 4.8 Solvent accessible surface area (SASA) over time for the side chain carboxylic groups of the two residues indicated as potential primary proton acceptors, namely Asp51 and Glh129 in OFC, **(left)** and Asp49 and Glh127 in IFC **(right)**. Asp51/49 and Glh129/127 are represented in violet and red, respectively. The SASA values are given in nm^2 while time is given ns. 112
- 4.9 Graphical representation of the accessibility of solvent inside the narrow hydrophilic channel formed between TM3 and TM9 and in contact with Asp51/49 and Glh129/127 in OFC **(a)** and IFC **(b)**, respectively. The two proteins are represented as magenta ribbons while the residues forming the hydrophilic network are represented as sticks with C, H, O, and N atoms colored in cyan, white, red and blue, respectively. The solvent is represented as a gray mesh. Both pictures have been obtained using VMD [29, 30]. 113

- 4.10 Pairs of amino acidic residues investigated to verify the eventual presence of water bridges during OFC and IFC MD simulations. As usual, TM1 and TM3 are colored in magenta while TM6 and TM9 are colored in green. C and H atoms of the residues are indicated in black while O and N are represented in red and blue, respectively. The three couples of residues analysed are represented with orange arrows. They are: (1) Asp51/49-His236/228; (2) Asp51/49-Glh129/127 and (3) His236/228-Asp126/124 in apo-EcoDMT and apo-ScaDMT numbering, respectively. The cyan arrows indicate the two possible release routes of protons: along the narrow hydrophilic channel or in the wide inward-directed vestibule connecting the active site and the cellular cytoplasm. The picture has been obtained using Biorender [31]. 114
- 4.11 Mechanistic model proposed for the proton transport involving both OFC and IFC. The protonated Glu residue located on TM3 (Glh) likely acts as the primary proton acceptor initiating the proton transport process. In the OFC, the proton is transferred from Glh to the Asp residue located on TM1 (**top left**) and from there to the His residue located on TM6b (**top right center**) while the deprotonated Glh residue immediately recovers its protonated state. Once the proton is located on the His residue and the conformational change to the IFC has occurred the proton can be released into the cellular cytoplasm using two distinct routes: it is either directly released through the wide inward-directed vestibule connecting the metal-ion binding site and the cytoplasm (**bottom right**) or it is transported to the Asp residue located on TM3 and finally from there released into the cellular cytoplasm through the narrow hydrophilic channel formed between TM3 and TM9 (**bottom left**). The light blue regions indicate regions accessible to the solvent while the proton is represented in orange. The sequence of events is indicated by the black arrows. The picture has been obtained using Biorender [31].117
- 4.12 Mechanistic model proposed for the proton transport occurring entirely in the OFC. From the protonated Glu residue located on TM3 (Glh) likely acting as the primary proton acceptor, the proton is transferred to the Asp residue on TM1. From there it is transferred to the His residue located on TM6b which, in turn, transfers it to the Asp residue on TM3. Finally, the proton is released into the cellular cytoplasm through the narrow hydrophilic channel formed between TM3 and TM9. The light blue regions indicate regions accessible to the solvent while the proton is represented in orange. The sequence of events is indicated by the black arrows. The picture has been obtained using Biorender [31]. . . . 118

- A.1 Chronological sequence of access of water molecules inside the Mn^{2+} coordination sphere during the equilibration phase of classical MD. (a) At the very beginning of the classical MD simulation ($t = 0$ ps) two water molecules (WAT30467 and WAT41018) are already close to Mn^{2+} . This seems reasonable considering that ScaDMT in its inward-facing conformation is in direct contact with the water composed cytoplasmatic vestibule. The short minimization dynamics allowed water molecules to diffuse inside the cytoplasmatic vestibule and approach the Mn^{2+} binding site. However, in these first moments of the thermalization process water molecules approach Mn^{2+} only transiently; (b) $t = 686$ ps: WAT24524 (light-blue box) enters Mn^{2+} coordination sphere and remains there until it exchanges with WAT3067, at around half of the production run; (c) $t = 220200$ ps: WAT1701 (red box) enters Mn^{2+} coordination sphere and keeps its positions until the end of the production run; (d) $t = 317050$ ps: WAT56324 (blue box) enters Mn^{2+} coordination sphere and keeps its position until the end of the production run; (e) end of the equilibration phase of classical MD ($t = 600000$ ps), WAT24524 (light blue box), WAT1701 (red box) and WAT56324 (blue box) coordinate Mn^{2+} together with the four amino acidic residues Asp49, Asn52, Ala223 and Met226. 128
- A.2 Distances between Mn^{2+} and WAT56324, WAT1701, WAT24524 and WAT3067 during the production run (1000 ns) of classical MD simulation, represented as blue, red, light blue and purple lines respectively (**left panel**). The exchange between WAT24524 and WAT3067 is highlighted by the black box. Magnification of the waters exchange area (**right bottom panel**). Representation of the Mn^{2+} coordination spheres before and after the water molecules exchange (**top right panels**). Asp49, Asn52, Ala223 and Met226 ligands are represented in licorice style while Mn^{2+} is represented as a gray sphere. The water molecules coordinating the metal ion are represented as ball and sticks model in the same colors with which their distances are represented in the graphs. Average values of the Mn^{2+} -waters distances are also given. 129
- B.1 pKa values of residues Asp51 (violet), Glh129 (red), Asp126 (turquoise) and His236 (blue) during apo-EcoDMT (OFC) trajectory ($1 \mu\text{s} = 1000\text{ns}$) with His236 in its protonated form. Time is reported in ns. Residue numbering from apo-EcoDMT. 131
- B.2 pKa values of residues Asp49 (violet), Glh127 (red), Asp124 (turquoise) and His228 (blue) during apo-ScaDMT (IFC) trajectory ($1 \mu\text{s} = 1000\text{ns}$) with His228 in its protonated form. Time is reported in ns. Residue numbering from apo-ScaDMT. 132
- B.3 (**left**) Asp51, Glh129, Asp126 and His236 Solvent Accessible Surface Area (SASA) values over time ($1 \mu\text{s} = 1000$ ns) in apo-EcoDMT (OFC) trajectory with His236 protonated. (**right**) Asp49, Glh127, Asp124 and His228 SASA values over time in apo-ScaDMT (IFC) trajectory with His228 protonated. SASA values are given in nm^2 while time is reported in ps. 133

List of Tables

1.1	Currently available crystal structures of Nramps. For each of the crystal structures the homolog, conformational state, the substrate, the resolution of the X-ray structure and the PDB IDs, as well as the references are given. 5M8K, 5M8J and 5M8A represent crystal structures containing the E129Q, H236A and E129A mutations with respect to the wildtype OFC-apo-EcoDMT protein (5M87), respectively. 6TL2 represents instead the crystal structure of an OFC-EcoDMT protein in complex with its inhibitor benzyl-bis-isothioureia.	8
1.2	Ionizable residues forming the hydrophilic network of residues potentially involved in the proton transfer in ScaDMT [121], EcoDMT [122] and DraDMT [123, 124]. The Asp residue on TM1 is also involved in the coordination of the divalent transition-metal ion. The Asp and Glu(2) residues on TM3 (lines 4 and 5) form two salt bridges with the two facing Arg residues on TM9 (lines 6 and 7), thus linking these two α -helices to shape the narrow hydrophilic channel mentioned above.	15
3.1	(top) Experimental and average QM/MM values of bond lengths for each bond formed by Mn^{2+} and its six ligands. The average distance between Mn^{2+} and the non-coordinating water molecule WAT3 is also reported. (bottom) Experimental and average QM/MM values for the distances between Mn^{2+} and Gln389($O^{\epsilon 1}$), Gly46(O) and Asp49($O^{\delta 1}$) atoms. QM/MM values are presented in Å, together with their standard deviation errors. Furthermore, each bond is indicated in the same color with which it appears in Fig. 3.7 and in Fig. 3.8b, respectively.	86
3.2	Experimental and average QM/MM values of angles formed between Mn^{2+} and its six ligands (in degrees). The QM/MM values are presented together with their standard deviations.	87
4.1	Ionizable residues forming the hydrophilic network in ScaDMT [25], EcoDMT [22] and DraDMT [23, 26] crystallographic structures. Notably, the Asp residue on TM1 is also involved in the coordination of the divalent transition metal ion. The Asp and Glu(2) residues on TM3 (lines 4 and 5) form two salt bridges with two facing Arg residues on TM9 (lines 6 and 7), thus linking these two α -helices to shape a narrow hydrophilic channel.	101

4.2	Total number of CHL, POPC and POPE molecules and final percentage composition of the membrane in each of the four studied systems. The letters u and p stand for unprotonated and protonated His residues, respectively.	105
4.3	Number of total water molecules, K^+ and Cl^- ions, as well as K^+ and Cl^- concentrations in each of the four studied systems (reported in mM).	106
4.4	Average pKa (pK_{aAV}) values with their standard deviations ($\sigma(pKa)$), ΔpKa (defined as the difference between the maximum and minimum pKa values assumed during the simulation) and δpKa (defined as the difference between the average pKa value and the standard pKa value in aqueous solution at physiological pH) for residues Asp51, Glh129, Asp126 and His236 of OFC (lines 1-4) and for residues Asp49, Glh127, Asp124 and His228 of IFC (lines 5-8). Reference standard pKa values in aqueous solution at physiological pH are taken from ref. [72]. Each residue is indicated in the same color used in Fig. 4.5 and 4.6.	111
4.5	Percentage of existence (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the OFC MD simulations (His236 unprotonated).	114
4.6	Percentage of existence of water bridges (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the IFC MD simulation (His228 unprotonated).	114
A.1	Energy values (in a.u.) and energy differences (in $kcal\ mol^{-1}$) of a simple model of the experimental ScaDMT active site in the sextet ($S = 5/2$), quartet ($S = 3/2$) and doublet ($S = 1/2$) multiplicity states, as obtained from DFT calculations. The tested model contains the four amino acidic residues coordinating Mn^{2+} (Asp49, Asn52, Ala223, Met226) properly saturated on the -NH and -CO groups and the Mn^{2+} ion itself, for a total of 62 atoms. DFT calculations were performed using the hybrid B3LYP exchange-correlation functional [1–3], the 6-31G(d,p) basis set [4–9] and the Gaussian 16 software package [10].	127
A.2	Absolute energies (in a.u.)	127
A.3	Energy differences (in $kcal\ mol^{-1}$).	127
A.4	Average distances between Mn^{2+} and Asp49($O^{\delta 2}$), Asn52($O^{\delta 1}$), Ala223(O), Met226(S^{δ}), Gln389($O^{\epsilon 1}$), Gly46(O) and Asp49($O^{\delta 1}$) calculated over the equilibration phase of classical MD (600ns) (first column), the production phase (1000ns) (second column) and the whole classical MD trajectory (equilibration + production) (1600ns) (third column).	127

A.5	Average distances between Mn^{2+} and the four water molecules participating to the coordination sphere during the classical MD simulation. Each average value has been calculated over the total permanence period of that water molecule in the coordination sphere. Mn^{2+} -WAT24524 average value has been calculated by averaging distance values between $t_0 = 686$ ps and $t_F = 1088100$ ps; Mn^{2+} -WAT1701 average value has been calculated by averaging distance values between $t_0 = 220200$ ps and $t_F = 1600000$ ps; Mn^{2+} -WAT56324 average value has been calculated by averaging distance values between $t_0 = 317050$ ps and $t_F = 1600000$ ps; Mn^{2+} -WAT3067 average value has been calculated by averaging distance values between $t_0 = 1088200$ ps and $t_F = 1600000$ ps.	128
B.1	Average pKa (pKa_{AV}) values with their standard deviations ($\sigma(\text{pKa})$) and the (ΔpKa) for residues Asp51, Glh129, Asp126 and His236 of apo-EcoDMT (lines 1-4) and for residues Asp49, Glh127, Asp124 and His228 (lines 5-8). Each residue is indicated in the same color used in Fig. B.1 and B.2.	132
B.2	Percentage of existence (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the apo-EcoDMT OFC MD simulations (His236 protonated).	133
B.3	Percentage of existence of water bridges (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the apo-ScaDMT IFC MD simulation (His228 protonated).	133
B.4	Average distances and corresponding standard deviations (in Å) between the residues forming the pairs analysed in the "water bridge" analysis for apo-EcoDMT with His236 unprotonated.	133
B.5	Average distances and corresponding standard deviations (in Å) between the residues forming the pairs analysed in the "water bridge" analysis for apo-ScaDMT with His228 unprotonated.	133
B.6	Average distances and corresponding standard deviations (in Å) between the residues forming the pairs analysed in the "water bridge" analysis for apo-EcoDMT with His236 protonated.	134
B.7	Average distances and corresponding standard deviations (in Å) between the residues forming the pairs analysed in the "water bridge" analysis for apo-ScaDMT with His228 protonated.	134

1 The Nramp family of transporters

1.1 The crucial role of Nramp transporters in living organisms

Transition metal ions are essential to all living organisms. They are mainly used as cofactors of enzymes catalyzing a wide range of chemical reactions fundamental for cellular metabolism and respiration. In particular, redox-active ions, such as Fe^{+2} , Mn^{+2} , Cu^{+2} or Co^{+2} , are usually bound to enzymes involved in catalytic reactions while the redox-inactive ion Zn^{+2} plays major roles in the stabilization of protein secondary structure. Nevertheless, excess or shortage of transition metal ions, as a result of genetic disorders or malnutrition, can induce the onset of severe eventually lethal diseases [1–3]. As an example, an excess of iron uptake has been proved to cause the hereditary disease hemochromatosis [4–6], while, on the contrary, a poor absorption of the same metal causes anemia [7]. Furthermore, recent studies claim the overload of transition metal ions in nervous system cells to be involved in the onset of neurodegenerative diseases such as Alzheimer, Parkinson or Huntington disease [8–14]. Therefore, metal ion homeostasis, achieved through a set of tightly regulated processes of uptake, storage and secretion, is crucial for the maintenance of regular cellular metabolic functions [15, 16].

Cellular uptake of a nutrient against a concentration gradient represents an entropically unfavourable process. To overcome this thermodynamic challenge, living organisms developed two distinct transport mechanisms, known as primary and secondary active transport, respectively [17]. In the first one, membrane proteins perform the translocation of the substrate across the lipid bilayer by coupling the transport process to an unrelated energetically favorable chemical reaction (in most cases ATP hydrolysis). Several primary active transporters use the energy provided from ATP to establish gradients of common ions such as H^{+} , Na^{+} , K^{+} , or Cl^{-} . The combination of these gradients with selective ion-specific channels creates a net charge separation between the two sides of the membrane, referred to as "membrane potential".

Secondary active transport is instead performed by membrane proteins harnessing these preexisting electrochemical gradients to power the thermodynamically disadvantageous trans-

port of a desired substrate, which is, in most cases, a scarce micronutrient [18–21]. The difference in the two active transport mechanisms is shown in Fig. 1.1.

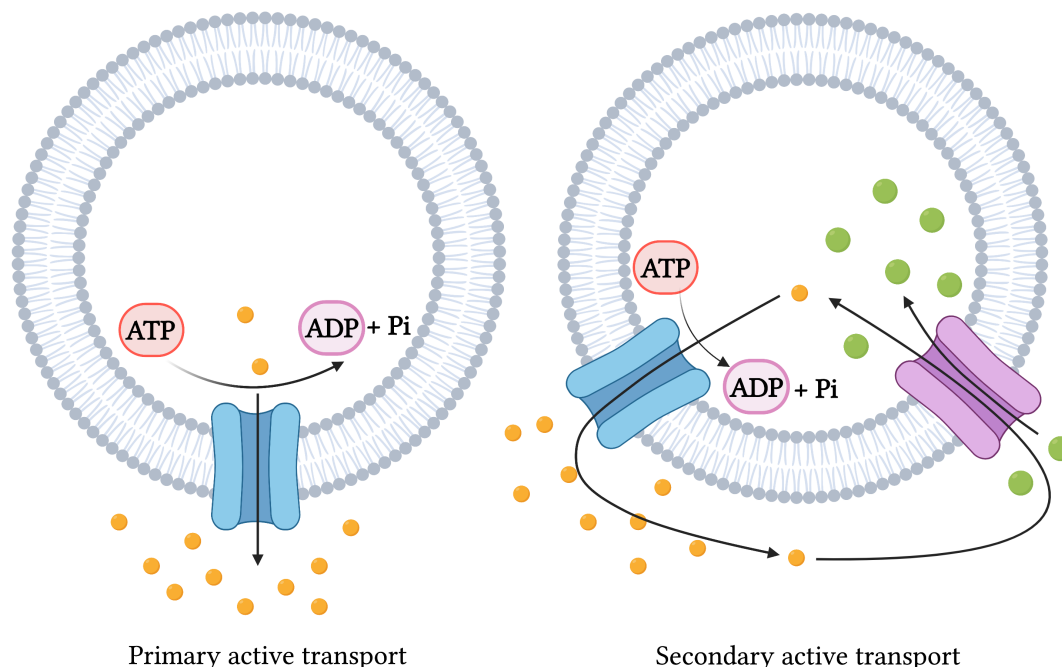


Figure 1.1: Schematic representation of primary (**left**) and secondary (**right**) active transport. The lipid bilayer (gray) represents the cellular membrane while the primary and secondary active transporters are depicted in blue and violet, respectively. The primary active transporter powers the transport of its substrate (yellow spheres) by using an external source of energy (e.g. the energy released from the ATP hydrolysis reaction). The electrochemical gradient created by the transport of the driving ions (yellow spheres) is harnessed by the secondary active transporter to realize the cotransport of the intended substrate (green spheres).

Two types of secondary active (co)transport exist: symport and antiport. In symport, the driving ion and the intended substrate are translocated across the membrane in the same direction, as shown on the right side of Fig. 1.1 and in Fig. 1.2. In antiport instead the protein moves the driving species and the intended substrates in opposite directions. The difference between simple uniport, and symport and antiport cotransport is illustrated in Fig. 1.2. The mutual dependence between primary and driving substrates in both mechanisms of secondary active transport (i.e. symport and antiport) has the purpose to prevent harmful events, such as the backwards flux of the primary substrate down its concentration gradient or purposeless cycles consuming the driving substrate gradient without stimulating the uptake of the primary substrate [22, 23].

Differently from channels, in which substrate movement is completely free, both primary and secondary active transporters usually work through some type of alternating access mechanism in which the protein exposes its substrate binding site to one side of the membrane

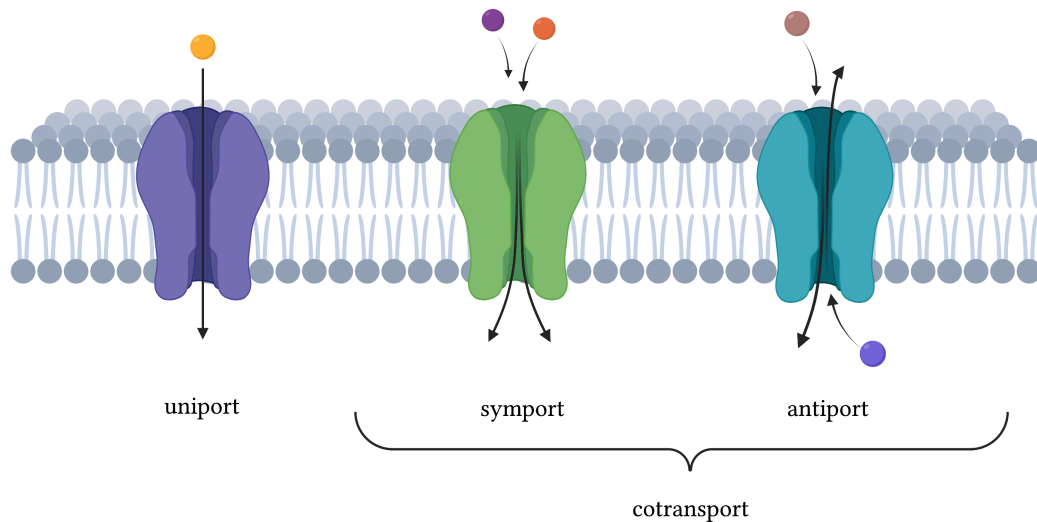


Figure 1.2: Schematic representation illustrating the difference between uniport, symport and antiport mechanisms. In uniport a single substrate is transported across the membrane in a unique direction. When two different substrates are instead transported simultaneously through the membrane we refer to as cotransport. Cotransport is typical of secondary active transporters and can be of two types: symport and antiport. In the first one, the two substrates are transported across the membrane in the same direction while in the latter they are transported in opposite directions.

at a time (i.e. the protein is alternately open towards the extracellular matrix or towards the cellular cytoplasm) [24].

Recently, researchers posed remarkable attention on metal ion transporters present in eukaryotic and prokaryotic cells and on the genes encoding them. As a result, it was discovered that, from bacteria to man, the family of Natural resistance-associated macrophage proteins (Nramps) plays a crucial role in the transport of divalent transition metal ions across cellular membranes [25–30]. These secondary active transporters enable therefore the uptake and trafficking of these micronutrients fundamental for the survival of all living organisms. Nramps work through an alternate-access mechanism in which the transport of the primary substrate (the transition metal ion) is coupled with the transport of H^+ ions (in this case the driving substrate) in the same direction (i.e. from the extracellular matrix to the cellular cytoplasm). They are therefore symporters.

Functional studies have been conducted on numerous Nramp homologs of bacterial species [31–44]. The most important substrate in prokaryotic members of the family resulted to be the Mn^{2+} ion scavenged from the environment. This evidence is supported by the fact

that bacteria often make use of alternative mechanisms for the uptake of iron and zinc and justifies the commonly used name for bacterial Nramps, MntH, that stands for H^+ -driven Mn^{2+} transporter.

Eukaryotic organisms use Nramp homologs for two main purposes. Along with the principal function of uptake and distribution of divalent transition metal ions Nramp proteins also act at the immune system level, through the so called "metal-withholding strategy" [45–47]. This defence mechanism is performed by restricting the access to these nutrients to pathogens. The resulting metal starvation impairs pathogen growth and makes them more susceptible to damages from the simultaneously released reactive oxygen species (ROS) [48–50]. These two complementary roles have likely been developed already in the earliest mono-cellular eukaryotic organisms, as found in the amoeba *Dictyostelium discoideum* [27, 51–53].

Fungi use multiple Nramp homologs for the adsorption and intracellular storage of environmental metal ions, including primarily Fe^{2+} and Mn^{2+} [54–58].

Several Nramp homologs are expressed also in plants [59, 60]. They allow for the uptake, along with the other transition metal ions, of Mn^{2+} , which represents undoubtedly the most important substrate, due to the key role that this metal carries out in chlorophyll photosynthesis. Metal ions, once adsorbed from the soil, are distributed to various tissues or stored so as to ensure survival under low metal conditions [60–66]. Simultaneously to these functions, also in plants, Nramp homologs help in defence against pathogen agents through metal-withholding mechanism [67, 68]. In some cases, plants adapted Nramp genes in order to enable the transport of the non-transition-metal ion Al^{3+} [69, 70].

Fish and many invertebrates express a single or two almost identical Nramp homologs performing both the immune and the metal-uptake roles [71–73].

In higher vertebrates instead, the two roles are exerted by two distinct paralogs, namely NRAMP1 (also known as SLC11A1), and NRAMP2 (also known as SLC11A2 or Divalent Metal Transporter 1, i.e. DMT1). The first one is specialized in the innate immune system's metal-withholding response to the attack of pathogens [30, 47, 74, 75] while the second is responsible for the dietary uptake of essential metal ions and their subsequent distribution to tissues, respectively [76–79].

In particular, mammalian NRAMP1 is expressed solely in phagocytes, where it helps in the killing of engulfed pathogens [80–84]. Mutations in NRAMP1 have indeed been proved to increase the susceptibility to pathogens [29, 45, 47, 81, 85, 86]. On the other side, mammalian NRAMP2 is expressed in many different tissues. However, it carries out its biological function mainly in the small intestines where it is particularly abundant. Here, indeed, it transports the iron ion Fe^{2+} introduced in the body through the diet across the apical membrane into the enterocytes cytosol, thus starting the cascade of reactions leading this metal, fundamental for cellular respiration, into the bloodstream [87–93]. It is furthermore involved in the systemic distribution of iron supplying red blood cells precursors, such as bone marrow erythroblasts. As might be expected, harmful mutations in NRAMP2 typically cause microcytic anemia in

rodents and humans [94–101]. However, it has been shown that mutations in NRAMP2 may also lead to liver iron overload [102–104].

In this paragraph the fundamental role of Nramp transporters in all living organisms has been described. In the following section, the current knowledge of structural and functional properties of Nramps as well as the hypotheses made regarding the transport mechanism will be outlined.

1.2 Structural and functional properties of Nramps

1.2.1 Substrate selectivity

In the landmark study that definitively established Nramps as transition-metal ion transporters, Gunshin et al. [26] proved that Mn^{2+} , Fe^{2+} , Ni^{2+} , Cu^{2+} , Co^{2+} , Zn^{2+} , as well as Cd^{2+} and Pb^{2+} ions were able to stimulate inward currents, due to intracellular cation import, in *Xenopus oocytes* expressing rat NRAMP2 protein. Furthermore, they were able to directly observe Fe^{2+} transport using radioactive labels. The study also showed that the alkaline-earth metal ion Ca^{2+} was selectively excluded from transport [26]. A large number of functional studies followed, in which different Nramp homologs from a wide range of prokaryotic and eukaryotic organisms were used [105–115]. The results coming from these studies can be summarized as follows. The essential micronutrients Mn^{2+} and Fe^{2+} represent the principal substrates for most organisms [33, 116–124]. Nramp transporters show high-affinity ($K_m \approx 1 \mu\text{M}$) for these two substrates likely exceeding the concentration of the ions themselves by an order or magnitude or more in many environments [26, 46, 107–109, 114, 122, 125, 126]. This evidence seems plausible if one considers the biologically important roles of iron and manganese with respect to other transition metal ions. However, a substantial group of other divalent transition metal ions are also transported by Nramp transporters, including the biologically useful Co^{2+} , Ni^{2+} , Cu^{2+} and Zn^{2+} as well as the toxic Cd^{2+} , Pb^{2+} and Hg^{2+} ions [126–134]. This indicates a high promiscuity of these transporters in the choice of the substrates, which likely reflects the difficulty to design a highly selective $\text{Mn}^{2+}/\text{Fe}^{2+}$ binding site [26, 107–109, 135]. The absence of an evolutionary thrust towards a unique binding site finely tuned in its geometry or amino acid composition to selectively transport only Mn^{2+} and Fe^{2+} is likely due to the lower concentration of other transition-metal ions with respect to Mn^{2+} and Fe^{2+} . In contrast, alkaline-earth divalent metal ions Ca^{2+} , Mg^{2+} , Sr^{2+} and Ba^{2+} have been proved in several studies to be selectively excluded from transport. These metal ions are quite abundant in most environments and, therefore, the strong discrimination exhibited by most Nramps against these non transition-metal ions is reputed to reflect the evolution promoted adaptation to obtain scarce micronutrients [26, 107–109, 135].

1.2.2 Structural features

Nrapm transporters usually contain 11 or 12 transmembrane (TM) α -helices with the N-terminus located in the cellular cytoplasm [25] as confirmed by the currently available crystal structures of Nrapm transporters obtained from prokaryotic members of the family *Staphylococcus Capitis* (ScaDMT) [121], *Eremococcus Coleocola* (EcoDMT) [122] and *Deinococcus Radiodurans* (DraDMT) [123, 124]. These crystal structures also confirmed for this protein family a LeuT-like fold [136] previously predicted based on distant sequence homology [27, 137, 138]. This fold, shared by many secondary transporter families, consists of 10 TMs α -helices divided into two pseudosymmetric and intertwining repeats of 5 consecutive helices, namely 1-5 and 6-10 in Nrapms. The first helix of each subdomain (i.e. α -helices 1 and 6, respectively) are broken into two helical segments by unwound regions in which the substrate binding site is located, as shown in Fig. 1.3 [139]. However, some structural features can vary among different Nrapm homologs. As an example, EcoDMT [122] as well as many eukaryotic homologs, possess a twelfth helix which is missing in ScaDMT [121], DraDMT [123, 124] and in most prokaryotic members of the family. Moreover, the N-terminus and the loop regions connecting the helical segments can considerably vary in length among homologs, especially those between α -helices 5-6, 7-8 and 9-10.

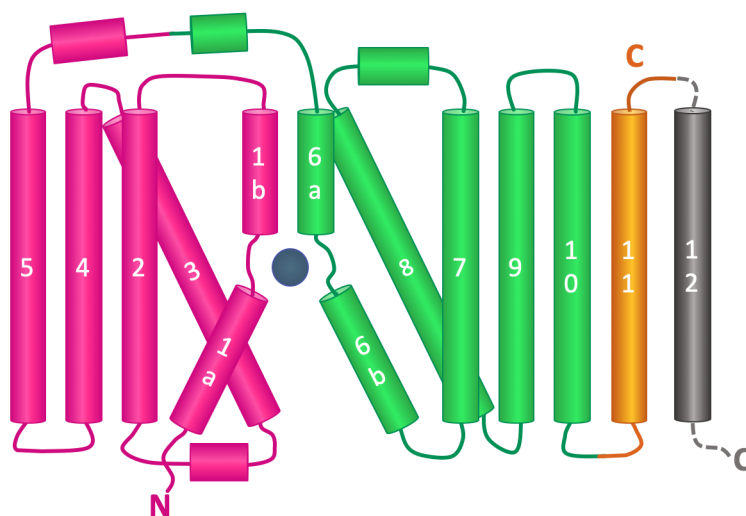


Figure 1.3: Nrapm secondary structure displaying the characteristic topology of the Leu-T fold. The two pseudosymmetric and intertwining repeats composed by α -helices 1-5 and 6-10 are represented in magenta and green, respectively. The 11th helix, excluded from the Leu-T-like fold, is represented in orange while the possible 12th helix is represented in gray. The unwound regions of α -helices 1 and 6 host the metal ion binding site, represented as a gray sphere.

1.2.3 Currently available crystal structures

As mentioned above, the currently available crystal structures of Nrapm proteins have been all obtained from prokaryotic members of the family. In particular, the *Staphylococcus Capitis* Divalent Metal-ion Transporter (ScaDMT) was crystallized in 2014 by Ehrnstorfer and co-

workers in its inward-facing conformation (IFC), both in a substrate-bound form, in complex with a Mn^{2+} ion (IFC-ScaDMT- Mn^{2+}), and in a substrate-free (apo) form (IFC-apo-ScaDMT) (PDB IDs: 4WGW and 4WGV, respectively) [121]. In 2016, the same authors deposited refined crystal structures of IFC-ScaDMT- Mn^{2+} and IFC-apo-ScaDMT (PDB IDs: 5M95 and 5M94, respectively). However, the latter differ from the original ones only for a small number of amino acid residues located in the 11th α -helix, close to the C-terminus, that were not yet resolved in 4WGW and 4WGV, while the rest of the crystal structures, including the metal-ion binding site region, are identical. Successively, in 2017, Dutzler and co-workers provided the crystal structure of the *Eremococcus Coleocola* DMT (EcoDMT) protein in its apo outward-facing conformation (OFC) (OFC-apo-EcoDMT) (PDB ID: 5M87) [122]. Furthermore, Bozzi et al. obtained crystal structures of the *Deinococcus Radiodurans* DMT (DraDMT) prokaryotic homolog in four different conformations. An apo IFC of DraDMT (IFC-apo-DraDMT), equivalent to IFC-apo-ScaDMT, was initially crystallized in 2016 (PDB ID: 5KTE) [123]. In 2019, along with a new version of IFC-apo-DraDMT (PDB ID: 6D9W), a G223W mutant of DraDMT was resolved in its OFC, both in the apo form (OFC-apo-DraDMT), equivalent to OFC-apo-EcoDMT, and in a Mn^{2+} -bound state (OFC-DraDMT- Mn^{2+}), representing the first crystal structure of a substrate-bound outward-facing conformation, (PDB IDs: 6D91 and 6BU5, respectively). Lastly, a novel inward-facing substrate-free intermediate conformation between the apo inward-facing conformation (IFC-apo-ScaDMT and IFC-apo-DraDMT) and the apo outward-facing conformation (OFC-apo-EcoDMT and OFC-apo-DraDMT) was crystallized for the G45R mutant of DraDMT. In this latter conformational state, the metal ion binding site results to be inaccessible to solvent from either the extracellular and the cytoplasmatic sides and, for this reason, it is usually referred to as "inward-occluded conformation" (IOC) (IOC-apo-DraDMT), (PDB ID: 6C3I) [124]. The currently available crystal structures of Nramps are listed in Table 1.1.

1.2.4 The proton-coupled alternate-access mechanism

The transition-metal ion transport is performed in Nramp secondary active transporters via a proton-coupled alternate-access mechanism in which the two substrates are translocated across the membrane in the same direction, i.e. from the extracellular matrix to the cellular cytoplasm (symport) [121–124, 140, 141]. The currently available crystal structures of Nramp proteins, described above, capture five distinct conformational states of the alternate-access mechanism: metal-bound OFC (OFC-DraDMT- Mn^{2+}) [124]; metal-bound IFC (IFC-ScaDMT- Mn^{2+}) [121]; apo IFC (IFC-apo-ScaDMT and IFC-apo-DraDMT) [121, 123]; apo OFC (OFC-apo-EcoDMT and OFC-apo-DraDMT) and, lastly, the apo inward-occluded conformation (IOC-apo-DraDMT) [124]. The presence of an apo IOC suggests for the transport cycle the presence of at least 6 distinct conformations, with the sixth one being a metal-bound "outward-occluded conformation" (OOC) occurring between the metal-bound OFC and the metal-bound IFC and in which, like in the apo IOC, the metal ion binding site (filled with the substrate) is not accessible from both sides of the membrane. However, such a conformational state has not yet been directly observed and, therefore, it remains for the moment only hypothetical. A

Homolog	Conformation	Ligand	Resolution (Å)	PDB ID	Reference
ScaDMT	Inward-open	None	3.10	5M94	[121]
ScaDMT	Inward-open	Mn ²⁺	3.4	5M95	[121]
ScaDMT	Inward-open	None	3.10	4WGV	[121]
ScaDMT	Inward-open	Mn ²⁺	3.4	4WGW	[121]
EcoDMT	Outward-open	None	3.3	5M87	[122]
EcoDMT	Outward-open	None	3.6	5M8K	[122]
EcoDMT	Outward-open	None	3.7	5M8J	[122]
EcoDMT	Outward-open	None	3.9	5M8A	[122]
EcoDMT	Outward-open	Inhibitor	3.8	6TL2	[140]
DraDMT	Inward-open	None	3.94	5KTE	[123]
DraDMT	Inward-open	None	3.94	6D9W	[124]
DraDMT	Outward-open	None	2.35	6D91	[124]
DraDMT	Outward-open	Mn ²⁺	2.4	6BU5	[124]
DraDMT	Inward-occluded	None	2.95	6C3I	[124]

Table 1.1: Currently available crystal structures of Nramps. For each of the crystal structures the homolog, conformational state, the substrate, the resolution of the X-ray structure and the PDB IDs, as well as the references are given. 5M8K, 5M8J and 5M8A represent crystal structures containing the E129Q, H236A and E129A mutations with respect to the wildtype OFC-apo-EcoDMT protein (5M87), respectively. 6TL2 represents instead the crystal structure of an OFC-EcoDMT protein in complex with its inhibitor benzyl-bis-isothiourea.

schematic representation of the putative transport cycle is represented in Fig. 1.4.

In the outward-facing conformation the substrate (i.e. a divalent transition-metal ion) can access the metal-ion binding site through an extracellular matrix water-filled vestibule formed between α -helices 1b, 3, 6a, 8 and 10 while tight side chain packing seals the internal vestibule, thus preventing contacts between the cytoplasm and the binding site [122, 124]. On the contrary, in the inward-facing conformation, the internal vestibule formed between α -helices 1a, 2, 5, 6b and 7 and filled with cytoplasm enables the substrate release into the cell while the external vestibule is in turn sealed in order to avoid contact between the binding site and the extracellular matrix [121, 123]. In the occluded conformation found in DraDMT, the metal-ion binding site results to be inaccessible to either the extracellular matrix and the cytoplasm, as already discussed before [124].

A comparison between the OFC observed in EcoDMT and DraDMT and the IFC observed in ScaDMT and DraDMT structures reveals that α -helices 1a, 4, 5, 6a and 10 are mostly involved in reorientations when the protein undergoes a conformational change. In particular, starting from the outward-open state, α -helices 1b and 6a tilt towards α -helices 3 and 8 while, simultaneously, the top half of α -helix 10 bends at the position of a conserved proline thus forming a ~ 20 Å long barrier composed of mainly hydrophobic residues above the metal-ion binding site, concurring in the process of sealing of the external vestibule. At the same time, the rigid rearrangement of α -helix 5 (suggested to be linked to the displacement of α -helix 6a through the 5-6 linking helical segment) and the consequent unhindered movement of the helical segment 1a rotating upward perpendicularly to the membrane, contribute to the

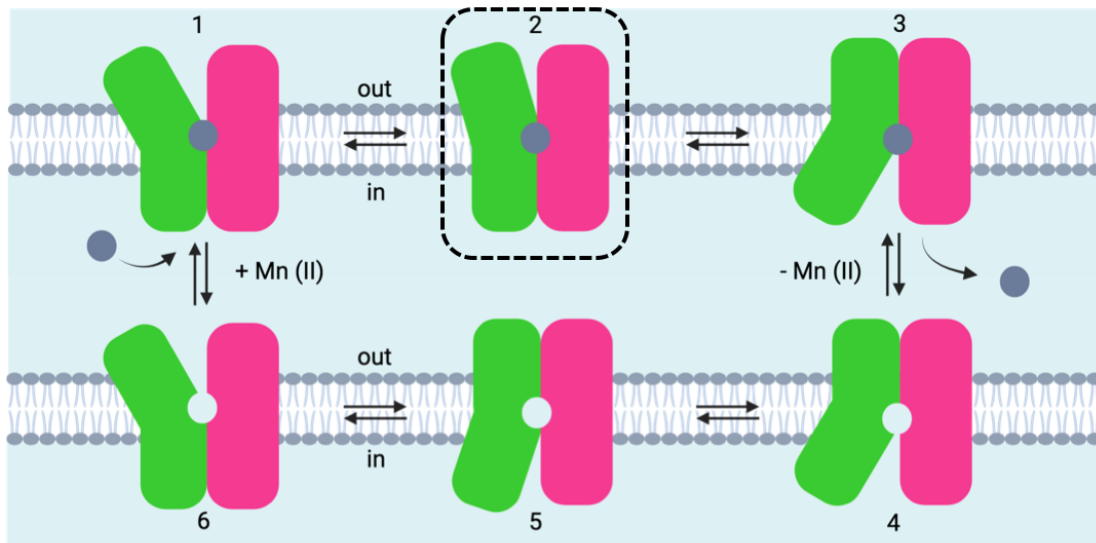


Figure 1.4: Schematic representation of the putative transport mechanism of Nramp transporters. The membrane bilayer is represented in gray, the extracellular matrix and the cellular cytoplasm are colored in light blue, the two proteic subdomains are represented in magenta and green while the transition metal ion (Mn^{2+} in this case) is depicted as a gray sphere. Structure 1 represents the metal-bound outward-facing conformation (OFC) (DraDMT- Mn^{2+} , PDB ID: 6BU5 [124]); 3 represents the metal-bound inward-facing conformation (IFC) (ScaDMT- Mn^{2+} , PDB IDs: 4WGW/5M95 [121, 122]); 4 represents the apo IFC (apo-ScaDMT and apo-DraDMT, PDB IDs: 4WGV/5M94 [121, 122] and 5KTE/6D9W [123, 124], respectively); 5 is the apo inward-occluded configuration (IOC) (apo IOC-DraDMT, PDB ID: 6C3I, [124]); 6 represents the apo OFC (apo-EcoDMT and apo-DraDMT, PDB IDs: 5M87 [122] and 6D91 [124], respectively). Structure 2, framed by a dashed line, is the hypothetical outward-occluded conformation (OOC) for which a crystal structure is still missing [124].

complete opening of the internal vestibule. The described conformational change from the OFC to the IFC is schematically depicted in Fig. 1.5 [141].

1.2.5 Metal-ion binding sites

Nramp transporters possess a highly conserved metal-ion binding site, located within the unwound regions of helices 1 and 6 and representing, as already mentioned, a conserved feature of the Leu-T fold [136, 139].

The first snapshot of a Nramp metal-ion binding site was provided by the IFC-ScaDMT- Mn^{2+} crystal structure obtained in 2014 by Ehrnstorfer and colleagues [121]. In this structure, still representing the only X-ray structure of a substrate-bound Nramp protein in the inward-facing conformation, the Mn^{2+} ion is coordinated by the side chains of an aspartate and an asparagine residue located on helix 1 (Asp49 and Asn52, respectively), and the backbone carbonyl oxygen of an alanine and the side chain of a methionine residue both located on helix 6, (Ala223 and Met226, respectively) (Fig. 1.6). While Asp49, Asn52 and Ala223 coordinate Mn^{2+} through oxygen atoms, acting as hard ligands, Met226 coordinates it through the thioether sulfur atom, acting as a soft ligand. However, the low resolution of the crystal structure (3.4 Å) as well as the highly distorted pseudo-planar coordination geometry does not allow to

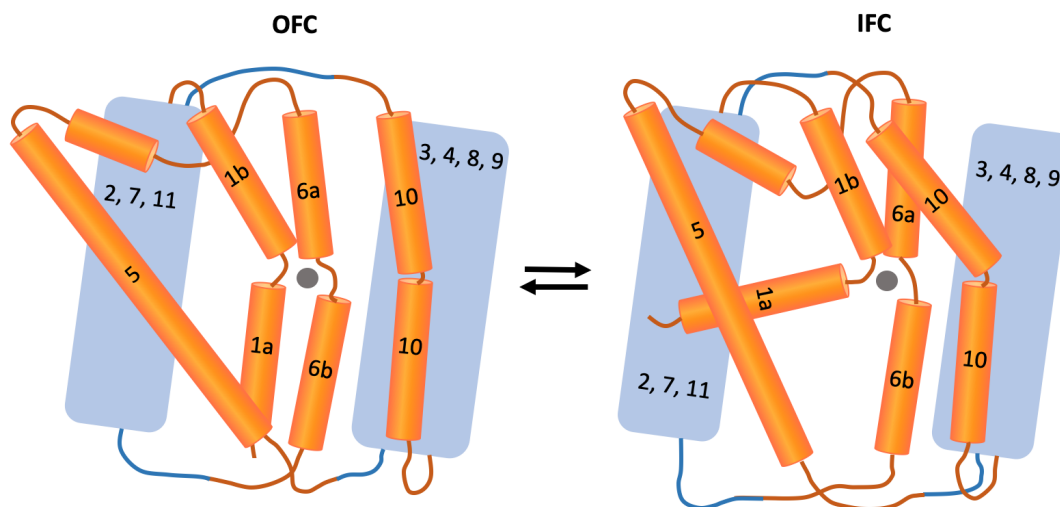


Figure 1.5: Schematic representation of the main rearrangements of α -helices during the conformational change from the OFC to the IFC states. α -helices undergoing more significant reorientations (i.e. 1, 5, 6 and 10) are represented in orange while the relatively stationary ones (i.e. 2, 3, 4, 7, 8 and 9) are represented as light blue blocks. The metal-ion substrate is represented as a gray sphere.

either confidently consider the coordination sphere as completed nor to identify any possible further ligand.

In crystals soaked with other transition-metal ions, Fe^{2+} , Ni^{2+} , Co^{2+} , Cd^{2+} and Pb^{2+} all bind at the same binding site as Mn^{2+} while Cu^{2+} binds at a slightly shifted site, still holding a tight interaction with Met226 [121]. Surprisingly, the Zn^{2+} ion binds in a completely distinct site, where it closely interacts with a H residue (H233) located on helix 6b [121]. As expected, the alkaline-earth metal ions Ca^{2+} , Sr^{2+} and Ba^{2+} do not bind even when soaked at high concentrations (200 mM) [121].

The OFC-DraDMT- Mn^{2+} crystal structure obtained by Bozzi and colleagues in 2019 provided the first (and, so far, the only) snapshot of a substrate-bound metal-ion binding site in the outward-facing conformation for a Nramp protein [124]. Here, similarly to the IFC binding site, the substrate (Mn^{2+} ion) is coordinated by side chains of an aspartate and an asparagine residue located on helix 1 (Asp56 and Asn59, respectively), by the side chain of a methionine residue (namely Met230) located on α -helix 6 and the backbone carbonyl oxygen of an alanine (Ala53) located on α -helix 1, and, therefore, not corresponding to Ala223 of IFC-ScaDMT- Mn^{2+} . The two ordered water molecules additionally coordinating Mn^{2+} and Asp56 acting as bidentate ligand lead to a total heptahedral coordination, as shown schematically in Fig. 1.7.

Several mutagenesis studies confirmed the fundamental roles played by the conserved metal-binding amino acid residues for the primary substrate (i.e. transition-metal ion) transport. The substitution of the α -helix 1 aspartate and/or asparagine with an alanine or more conservative residues was proved to impair metal ion binding in the active site and, therefore, eliminate

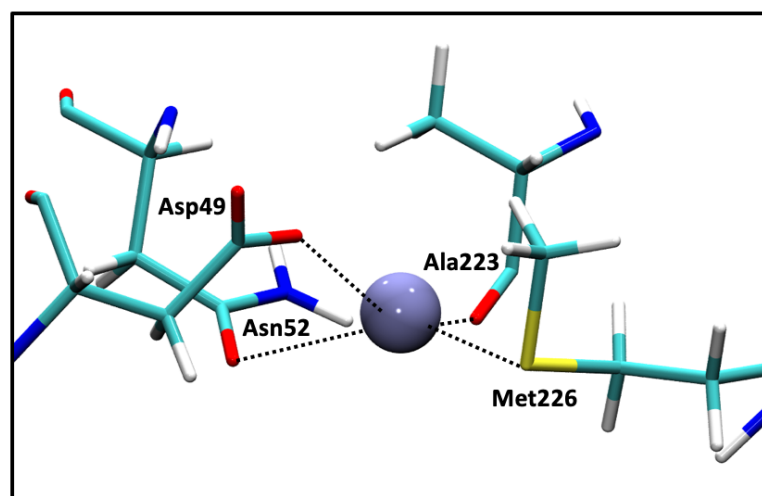


Figure 1.6: Metal-ion binding site of a metal-bound Nramp protein in its inward-facing conformation, as obtained by Ehrnstorfer and colleagues in the IFC-ScaDMT-Mn²⁺ crystal structure [121]. The picture has been obtained with VMD.

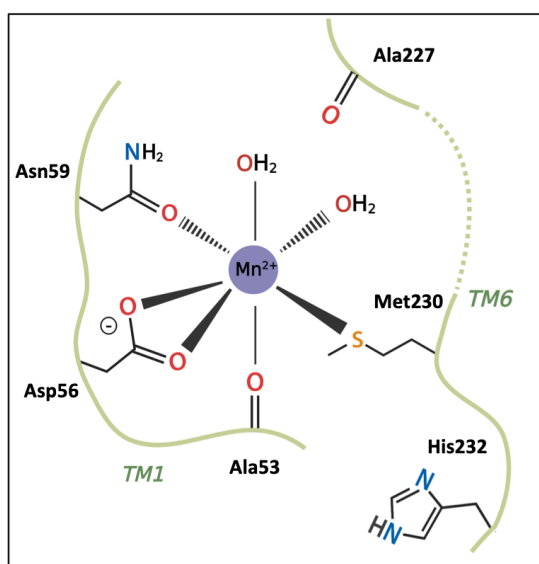


Figure 1.7: Schematic representation of the metal-ion binding site of a metal-bound Nramp protein in its outward-facing conformation, as observed by Bozzi and colleagues in the OFC-DraDMT-Mn²⁺ crystal structure [124].

or severely reduce the metal transport in many investigated homologs [115, 121, 122, 135, 142, 143]. The conserved methionine residue on α -helix 6 similarly was proved to play a crucial role both in metal transport and in the modulation of Nramps substrate selectivity [141, 143]. The substitution of this methionine with several different amino acids (among which alanine, cysteine, glycine, glutamine, and lysine) leads in most cases to the loss or reduction of metal transport [141, 143]. Furthermore, the substitution of methionine with some residues, such as alanine or threonine, enabled the transport of alkaline-earth metal ions such as Ca²⁺ and Mg²⁺, suggesting that the coordinating thio-ether sulfur group of methionine

undergoes semi-covalent interactions which preferentially stabilize transition-metal ions [141, 143]. This explains why alkaline-earth metal ions, establishing solely ionic interactions and thus preferring hard ligands, such as oxygen atoms, become possible Nramp substrates only when methionine is replaced by alanine (that due to its reduced size, leaves space for a water molecule to coordinate the metal ion) or by hydroxyl-bearing residues such as threonine.

The metal-ion coordination sphere in the inward-facing conformation of Nramps: an open issue

As already mentioned, the crystal structure of ScaDMT in its inward-facing conformation and in complex with a Mn^{2+} ion (IFC-ScaDMT- Mn^{2+}), obtained by Ehrnstorfer and colleagues in 2014, still represents the only available snapshot of the metal-ion binding site of a Nramp transporter in the inward-open conformational state [121]. The Mn^{2+} ion resulted to be coordinated by 4 amino acidic residues: namely Asp49, Asn52, Ala223 and Met226. However, the low resolution of the X-ray structure (3.4 Å) and the highly distorted pseudo-planar coordination geometry (see Fig. 1.6) raised many doubts about the completeness of the metal ion coordination sphere.

A theoretical study from Pujol-Giménez [144] and colleagues (2017), performed on the IFC ScaDMT protein and using Cd^{2+} instead of Mn^{2+} as the bound substrate, provided quite a different description of the metal ion coordination sphere with respect to that observed in the crystal structure. Indeed, in the resulting model, along with the 4 amino acidic residues already identified by the X-ray structure, the backbone carbonyl oxygen of a glycine residue located on α -helix 1 (Gly46) and the side chain of a glutamine residue located on α -helix 10 (Gln389) coordinate the metal ion, thus leading to an octahedral coordination geometry. However, as evident in the left panel of Fig. 1.8, the major discrepancy with respect to the crystal structure consists in the fact that Met226 does not coordinate the metal ion and its coordination site is occupied by the second carboxylic oxygen of Asp49 which acts as a bidentate ligand.

More recently, Bozzi and colleagues postulated a heptaedral coordination with two water molecules filling the two axial coordination sites left empty by the crystal structure and the aspartate Asp49 acting as a bidentate ligand for the IFC-ScaDMT coordination sphere (right panel of Fig. 1.8 [124]).

The three different descriptions of the metal ion coordination sphere in an IFC Nramp transporter (i.e. the crystal structure from Ehrnstorfer et al. [121], the theoretical model obtained by Pujol-Giménez et al. [144] and the putative model provided by Bozzi et al. [124]) prove that the precise nature of the metal binding site is still an open issue that needs to be further investigated. In this thesis, we addressed this problem by performing classical molecular dynamics (MD) and mixed quantum mechanics-molecular mechanics *ab initio* molecular dynamics (QM/MM MD) simulations of the IFC-ScaDMT- Mn^{2+} protein (represented in Fig. 1.9), as we will extensively discuss in Chapter 3.

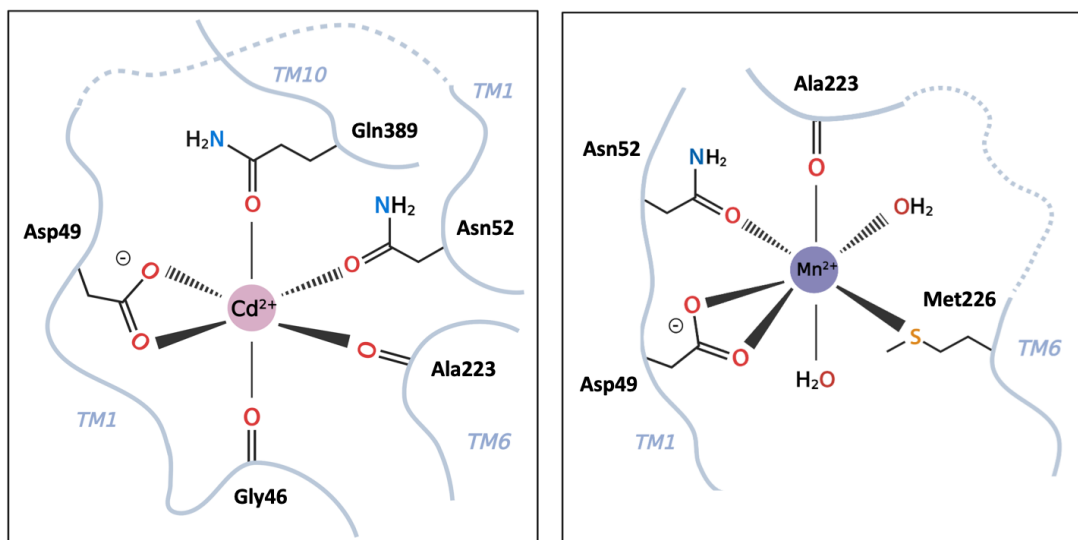


Figure 1.8: Schematic representation of the metal ion coordination spheres for IFC-ScaDMT as obtained from the theoretical study of Pujol-Giménez and colleagues [144] (**left**) and as speculated by Bozzi and colleagues [124] (**right**). The Cd^{2+} and Mn^{2+} ions are represented as pink and gray spheres, respectively. In both panels TM stands for transmembrane α -helix and the O, N, and S atoms are represented in red, blue and orange, respectively.

1.2.6 Proton transport in the Nramp family of transporters

As previously described, in Nramp secondary active transporters, the translocation of divalent transition metal ions (representing the primary substrates) against their concentration gradient (i.e. from the extracellular matrix to the cellular cytoplasm), is coupled with the transport of protons (driving ions) in the same direction (symport). However, transport of protons and metal ions is not tightly coupled and their respective stoichiometry can vary depending on several factors, such as the metal ion itself, the pH gradient or the transmembrane voltage [26, 55, 122, 124, 141, 144, 147].

A first evidence of the unconventional proton transport occurring in Nramps was provided in the seminal work of Gunshin and colleagues [26]. In particular, the application of a lower external pH and a negative membrane potential to *Xenopus oocytes* cells expressing rat NRAMP2 led to the observation of an inward current, indicating cations influx, even in the absence of metal ions, while the simultaneous drop in the intracellular pH confirmed that the inward current was due to a proton uniport phenomenon. The addition of Fe^{2+} to the system enhanced both the inward current and the intracellular acidification, suggesting therefore a coupled proton-metal cotransport.

A plethora of functional studies, performed on a wide range of eukaryotic and prokaryotic homologs, followed [26, 55, 110, 115, 117, 119, 122, 124, 125, 131, 135, 137, 144, 147–150]. These studies altogether seem to point to the same general features for Nramp transporters: protons are transported from the extracellular matrix to the cellular cytoplasm, down their electrochemical gradient, either in the absence or presence of the metal-ion substrate ("proton



Figure 1.9: MD snapshot of a configuration of the IFC ScaDMT protein in complex with its substrate (a Mn^{2+} ion) for which we performed a combination of classical MD and *ab initio* QM/MM MD simulations. The protein α -helices are represented as magenta ribbons, the Mn^{2+} ion is represented as a gray sphere while the four amino acidic residues identified as ligands in the X-ray structure [121] are represented as sticks. The picture has been obtained using VMD [145, 146].

uniport" or "proton-metal symport", respectively). However, it is important to underline that, in most cases, the presence of metal ions seems to enhance the proton transport and vice-versa.

Recently, several attempts to outline Nramps proton transport mechanism have been made. In this regard, fundamental insights were provided from the crystallographic characterization of Nramp prokaryotic homologs, such as ScaDMT [121], EcoDMT [122] and DraDMT [123, 124], referred to as ScaDMT, EcoDMT and DraDMT, respectively. Although crystallographic experiments in general do not allow to directly observe protons, the derived structural data represented a good starting point in the understanding of the process as it allowed for the identification of the ionizable amino acid residues potentially involved in the process.

In particular, Nramp transporters feature a conserved network of protonatable residues located in the core of the protein which is absent in structurally related LeuT-fold outgroups, including those known to perform proton transport [121, 122, 124, 141]. Within this network, a trio composed of an Asp residue located on α -helix 1 (TM1), a Glu residue on TM3 and a His residue on TM6b connects the metal binding site (exposed to the bulk solvent from both the

IFC and the OFC), with a pair of salt bridges linking TM3 and TM9. These two salt bridges occur between an Asp and a Glu located on TM3 and two Arg located on TM9 and are placed at the center of the lipid bilayer and close to the cytosol, respectively. Furthermore, a group of conserved polar and charged residues flanks TM3, TM9 and the close TM4 and TM8 helices so as to form a narrow hydrophilic channel likely filled with water molecules and, therefore, suitable for proton transfer. The conserved network of ionizable residues potentially involved in the proton transfer mechanism is schematically represented in Fig. 4.1. It is also important to mention that the Asp residue on TM1 is the same that also coordinates the transition metal ion substrate. For the sake of clarity, the specific residues forming the hydrophilic network in ScaDMT, EcoDMT and DraDMT are listed in Table 1.2.

Residue	ScaDMT	EcoDMT	DraDMT
Asp on TM1	Asp49	Asp51	Asp56
His on TM6b	His228	His236	His232
Glu(1) on TM3	Glu127	Glu129	Glu134
Asp on TM3	Asp124	Asp126	Asp131
Glu(2) on TM3	Glu117	Glu119	Glu124
Arg(1) on TM9	Arg355	Arg368	Arg352
Arg(2) on TM9	Arg360	Arg373	Arg353

Table 1.2: Ionizable residues forming the hydrophilic network of residues potentially involved in the proton transfer in ScaDMT [121], EcoDMT [122] and DraDMT [123, 124]. The Asp residue on TM1 is also involved in the coordination of the divalent transition-metal ion. The Asp and Glu(2) residues on TM3 (lines 4 and 5) form two salt bridges with the two facing Arg residues on TM9 (lines 6 and 7), thus linking these two α -helices to shape the narrow hydrophilic channel mentioned above.

Proposed proton transport mechanisms

Ehrnstorfer and colleagues were the first to formulate a hypothesis on the proton transport mechanism, based on the functional and structural information obtained from the ScaDMT and EcoDMT crystal structures [121, 122]. As possible primary proton acceptors, they identified two EcoDMT protonatable residues, Glu129 and His236, both of them close to the metal ion binding site and expected to be susceptible to changes in their access to the substrate-binding region when passing from the IFC to the OFC and vice-versa. However, while the *in silico* computed pKa value for His236 resulted to be in a suitable range to accept and release a proton, the estimated value of Glu129 was too low. In order to obtain new insights, they also performed transport assays on EcoDMT mutants. While the conservative mutant E129Q seemed not to affect both Mn^{2+} and H^+ transport, the H236A mutant displayed remarkably altered transport properties, with robust Mn^{2+} transport but no detectable H^+ coupling. This led them to indicate His236 as the primary proton acceptor.

In EcoDMT OFC [122], His236 is placed in close contact with the aqueous cavity connecting the metal ion binding site with the extracellular matrix, whereas in ScaDMT IFC, this residue is located at the crossroad of two potential proton release pathways [121]. On one side the residue is indeed exposed towards the wide inward-directed cavity constituting the metal

ion release route, while, on the other side, it faces the narrow hydrophilic channel formed between TM3 and TM9, as shown in Fig. 1.10. Summarizing, Ehrnstorfer and colleagues formulated the following hypothesis for the proton transport mechanism: the H residue represents the primary acceptor that binds the proton in the OFC (top panel of Fig. 1.10). After the conformational change to the IFC (bottom panel of Fig. 1.10), the proton is released either through the wide inward-directed cavity, taking the same exit pathway as the metal ion, or via the narrow side channel.

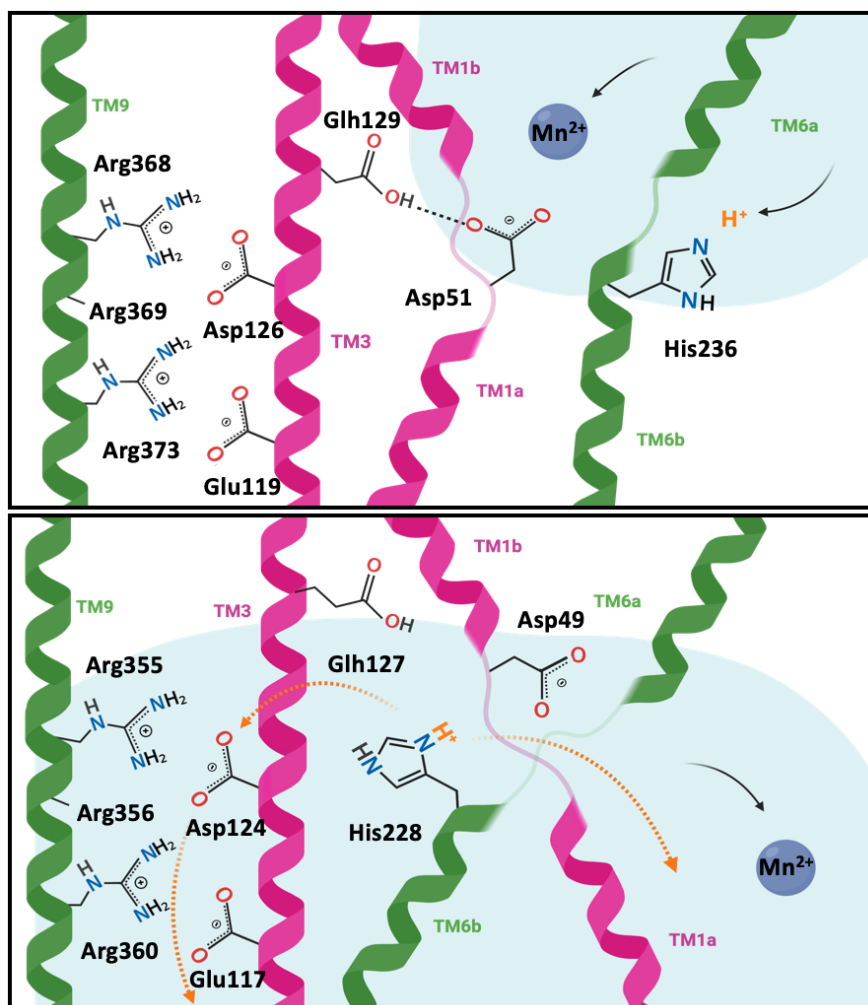


Figure 1.10: Schematic representation of the proton transport mechanism as hypothesized by Ehrnstorfer and colleagues [122]. In the OFC (**top**), Asp51 and His236 are both in close contact with the aqueous cavity (represented in light blue). The proton (represented in orange) binds to His236, which acts as the primary proton acceptor. After the conformational change to the IFC (**bottom**) the proton bound to His236 can be released into the cytoplasm either directly, through the wide inward-directed aqueous cavity or via the narrow channel. The two possible exit pathways are indicated as orange dashed arrows. TM1 and TM3, belonging to the first protein subdomain, are represented in magenta while TM6 and TM9, belonging to the second protein subdomain, are represented in green. The metal ion (Mn^{2+}) is represented as a gray sphere. Within the amino acidic residues, O and N atoms are represented in red and blue, respectively, while C and H atoms are represented in black. The residue numberings correspond to those of EcoDMT and ScaDMT in the **top** and **bottom** panel, respectively. The pictures have been obtained with Biorender [151].

A second hypothesis for the proton transport mechanism was proposed in 2017 by Pujol-Giménez and colleagues, based on a combined computational and experimental work [144]. Amino acidic residues potentially involved in proton binding were assessed through *in silico* pKa predictions performed on ScaDMT crystal structure. Among residues owning remarkable pKa shifts, Glu127 was selected also due to its strategic position, in close contact with the metal ion binding site and the intracellular water-filled cavity. Independently of their pKa values, three other residues were selected. Asp49, which is a metal-ion binding residue conserved within the whole Nramp family, was chosen based on both its position and its theoretical ability to receive a proton. Despite their low pKa values, His228 and His233 were chosen as they were previously proved to be closely connected to proton transport [114, 122]. In order to verify whether protonation of any of the four selected residues (Asp49, Glu127, His228 and H233) would lead to structurally relevant conformational changes, molecular dynamics simulations were performed on ScaDMT X-ray structures in which each residue was individually protonated. Secondary structure analysis showed loss of helicity in transmembrane α -helix TM8 in those conformations in which Glu127 was deprotonated, while this effect was not observed in presence of protonated Glu127. This evidence further pointed to this residue as being involved in relevant structural rearrangements of the transporter. For this reason, the conservative E127D and the detrimental E127A mutations were engineered. Iron transport assays revealed for both of them an affinity for Fe^{2+} similar to that of the wild-type (WT) protein, although with a much slower transport kinetics. In proton transport assays, the I-V relationship obtained from the H^+ -evoked currents was found to be almost identical for WT and E127D while in E127A mutant the current was much lower. This suggested a crucial role for the negative charge, present in both the WT protein and the E127D mutant, for the proton transport process.

The overall transport mechanism proposed by Pujol-Giménez and colleagues can be summarized as follows [144]. Glu127 represents the most likely proton acceptor. Both protons and metal ions bind in the OFC, in the high-pKa state of Glu127, as illustrated in Fig. 1.11(a). Immediately after, a conformational change leads the system to an inward-occluded state, in which Glu127 switches to a low-pKa state thus releasing the proton into the narrow channel, as shown in Fig. 1.11(b). The authors suggest an allosteric coupling between Glu127 side chain and Asp49, established through direct Coulombic interaction. The signal triggering gate opening is the return of Glu127 to the negatively charged state. Simultaneously with the intracellular gate opening a partial unwinding of TM8 occurs, conferring high flexibility to Glu127 side chain (Fig. 1.11(c)). Once the gate is open and the system is in an IFC, the metal ion is released, (Fig. 1.11 (d)) and Glu127 side chain reorients into its high-pKa proton-acceptor state, (Fig. 1.11 (e)). Finally, the initial outward-open state is recovered (Fig. 1.11 (f)).

In 2019, Bozzi and colleagues formulated a new hypothesis for Nramps transport mechanism [124]. From *in silico* pKa predictions performed on DraDMT and ScaDMT X-ray structures, the metal-ion binding residue Asp56 (DraDMT numbering) resulted to have pKa values shifted upwards with respect to the reference values, consistently with the expected behaviour of amino acidic residues located at protein's core. The calculated pKa values for Asp56 (6.7

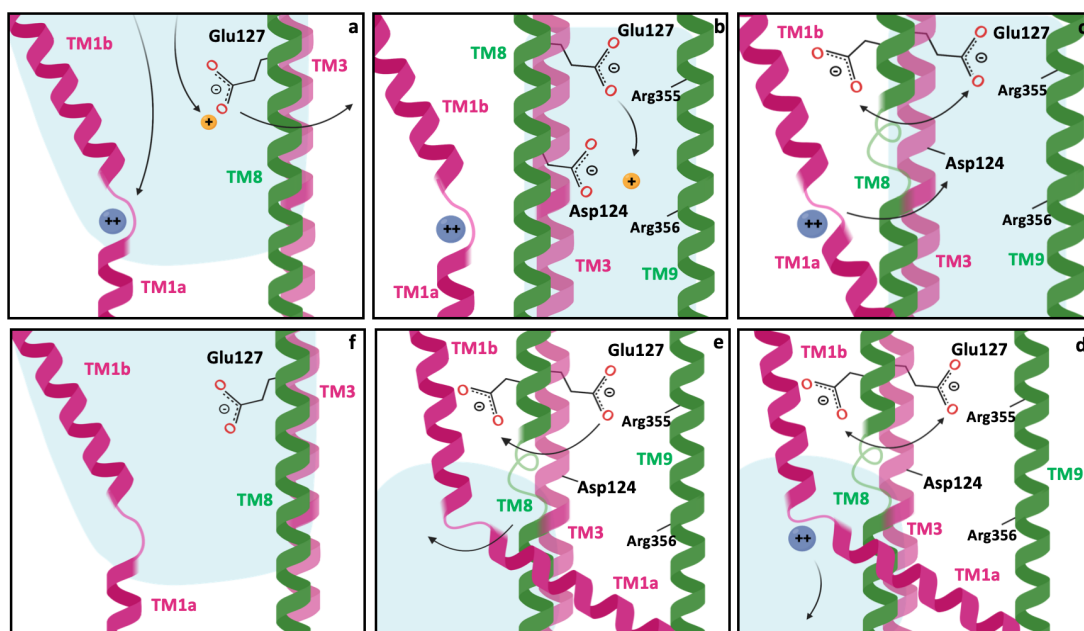


Figure 1.11: Schematic representation of the overall transport mechanism as described by Pujol-Giménez and colleagues [144]. (a) Both substrates bind in the high-affinity OFC, where Glu127 is in a high-pKa state and accepts the proton. (b) After a conformational switch to an inward-occluded state, the proton is released by Glu127 likely along the narrow water-filled channel established between TM3 and TM9. (c) The consequent negative charge on the Glu127 is the signal that triggers intracellular gate opening while, simultaneously, a partial unwinding of TM8 confers high flexibility to Glu127. (d) In the IFC, the metal ion is solvated and released to the cytoplasm. (e) Glu127 reorients itself into its high-pKa state and the gate closes, thus recovering the initial OFC state (f). The proton and the divalent transition metal ion are represented as orange and gray spheres, respectively. Their respective charge is also indicated. TM1 and TM3 and TM8 and TM9 are represented in magenta and green, respectively. The solvent accessible region of the protein is represented in light blue. The pictures have been obtained using the Biorender software [151].

and 6.8 in DraDMT and ScaDMT, respectively) were in an optimal range for rapid protonation/deprotonation at physiological conditions. Furthermore, in transport assays performed on D56A and D56N DraDMT mutants, both H^+ uniport and Mn^{2+} - H^+ symport were removed, while, in transport assays conducted on EcoDMT (expressed in protoliposomes) and *in vivo* EcoliDMT, the corresponding mutations were found to interrupt the Mn^{2+} -stimulated proton transport. Such evidences are consistent with the hypothesis of a crucial role of Asp56 in H^+ transport, and, therefore, Bozzi and colleagues proposed this residue as the primary proton acceptor. They proposed a model in which protonated Asp56 is able to optimize the binding site for the incoming metal ion by forming a hydrogen bond with the metal binding Asn59 residue, which, in turn, due to this interaction orients its carbonyl oxygen opportunely to bind the metal ion. Furthermore, in proteoliposomes transport assays, the H232Q mutant was proved to preserve metal ion transport while H^+ transport, as well as any kind of pH dependence in Mn^{2+} transport kinetics, was eliminated. Therefore, Bozzi and colleagues hypothesized for this residue a role both in stabilizing the proton transfer from Asp56 to Asp131 and as the hinge point for the conformational switch from the OFC to the IFC.

Summarizing, the following mechanism for Nramps metal-proton cotransport was proposed [124]. Residue Asp56 represents the likely primary proton acceptor. Its protonation in the OFC likely optimize the binding site through an hydrogen bond with the Asn59 residue that, in turn, reorients opportunely to bind the incoming metal ion. The proton, displaced from Asp56 by the metal ion binding, is transferred to Asp131 either via Glu134 or via His232. From Asp131 it is finally released, presumably into the water-filled narrow channel formed by TM3 and TM9. In this model, the two key residues involved in proton transport are therefore Asp56 and Asp131, that sequentially protonate and deprotonate, while Glu134 and His232 only stabilize the proton shuttle from Asp56 to Asp131. The simultaneous metal binding and proton transfer trigger the conformational switch from the OFC to the IFC, resulting in intracellular gate opening and metal ion release to the cytoplasm.

Further information were obtained from functional studies performed on conformationally locked DraDMT mutants [124]. In particular, in outward-open locked DraDMT mutants proton uniport was observed, contrarily to metal ion transport. In analogous inward-locked DraDMT mutants such kind of transport was instead not observed. These evidences suggest that while metal transport needs the transporter to undergo conformational change from the OFC to the IFC, proton transport does not necessarily require sampling of the IFC with proton uniport eventually occurring also directly in the OFC as shown in Fig. 1.12. Furthermore, according to this mechanism, the two cosubstrates likely enter the protein through the same wide cavity connecting the binding site to the extracellular ambient but take two distinct pathways to the cytosol: while the metal ion is indeed released through the wide vestibule connecting the binding site to the cytosol, protons likely exit trough the narrow side channel formed by TM3 and TM9, which is open in both the IFC and the OFC (Fig. 1.12).

Despite the fundamental structural and functional evidences provided by a wide range of experimental studies, the proton transport mechanism occurring in Nramp transporters remains still unclear, as evidenced by the three different and inconsistent mechanistic models proposed by Ehrnstorfer et al. [122], Pujol-Giménez et al. [144] and Bozzi et al. [124]. However, the understanding of such mechanistic details would be of fundamental importance in order to explain the reasons underlying the highly variable proton/metal transport stoichiometry and the still undetermined thermodynamic coupling between the two substrates. In this thesis, we addressed this issue by performing classical molecular dynamics (MD) simulations on both ScaDMT (IFC) and EcoDMT (OFC) proteins in their apo forms (represented on the left and right panels of Fig. 1.13, respectively), as we will comprehensively discuss in Chapter 4.

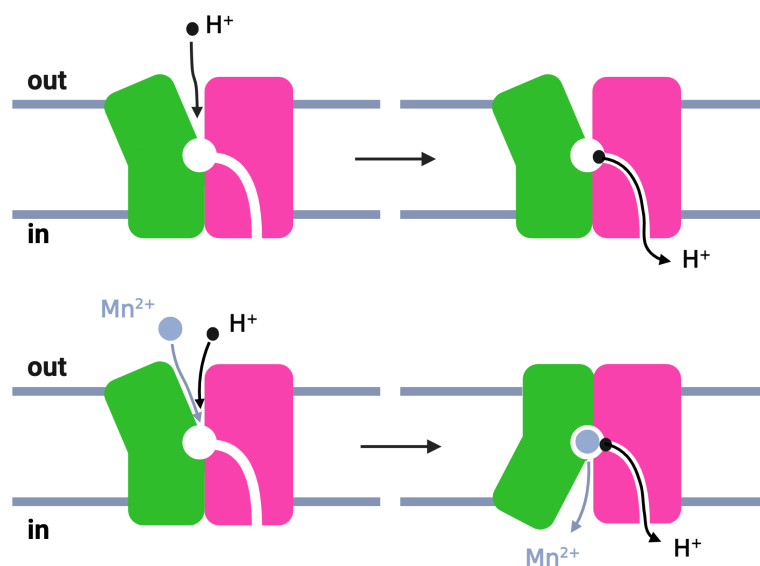


Figure 1.12: Proton uniport (**top**) and metal ion/proton symport (**bottom**) as hypothesized by Bozzi and colleagues. In proton uniport, H^+ enter via the external vestibule and are released through the narrow side channel, without the need for the intracellular gate to open (i.e. without the need for a conformational change from the OFC to the IFC). The metal ion/proton symport requires instead the opening of the gate, and, consequently, the conformational switch from OFC to IFC. In this case the two substrates enter both via the external vestibule but, after binding, take two distinct routes towards the cytoplasm. The two proteic subunits are represented in green and magenta, the membrane is represented in gray while the metal ion and the proton are represented as gray and black spheres, respectively. The picture has been created with Biorender [151].

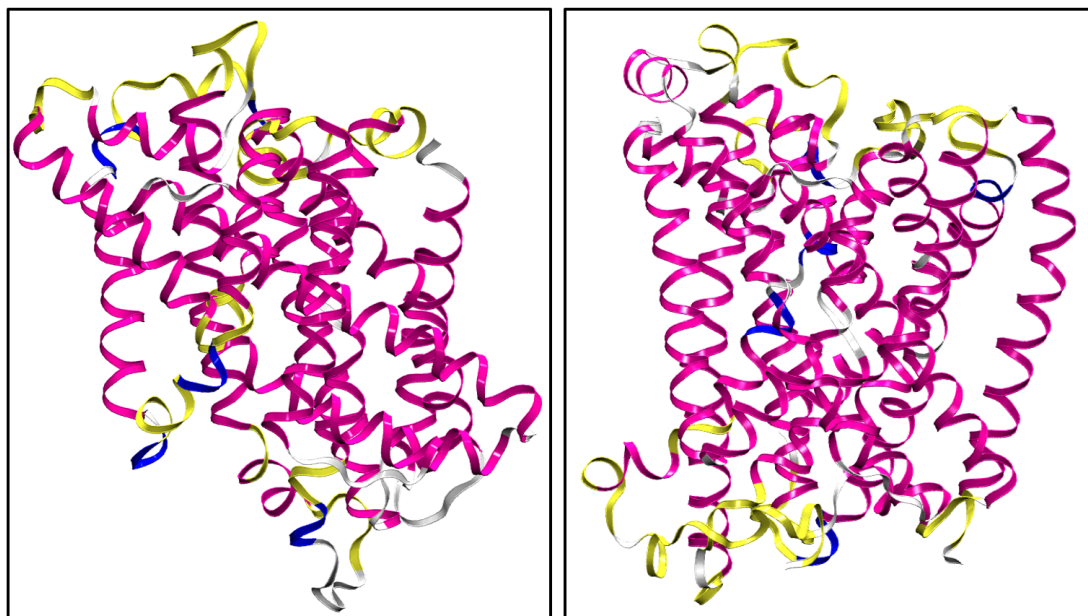


Figure 1.13: MD snapshots reporting two configurations of the IFC apo-ScaDMT (**left**) and the OFC apo-EcoDMT (**right**) proteins used for classical MD simulations [121, 122]. As before, the α -helices of the two proteins are represented as magenta ribbons. The two pictures have been obtained using VMD [145, 146].

References

- [1] Y. Nevo and N. Nelson. “The NRAMP family of metal-ion transporters”. In: *Biochim. Biophys. Acta* 1763.7 (2006), pp. 609–620. DOI: <https://doi.org/10.1016/j.bbamcr.2006.05.007>.
- [2] T. Herrmann et al. “Iron overload in adult Hfe-deficient mice independent of changes in the steady-state expression of the duodenal iron transporters DMT1 and Ireg1/ferroportin”. In: *J. Mol. Med.* 82 (2004), pp. 39–48.
- [3] P. Fergelot et al. “Over-expression of wild-type and mutant HFE in a human melanocytic cell line reveals an intracellular bridge between MHC class I pathway and transferrin iron uptake”. In: *Biol. Cell* 95.5 (2003), pp. 243–255. DOI: [https://doi.org/10.1016/S0248-4900\(03\)00057-1](https://doi.org/10.1016/S0248-4900(03)00057-1).
- [4] A. Pietrangelo. “Hereditary hemochromatosis”. In: *Biochim. Biophys. Acta* 1763.7 (2006), pp. 700–710. DOI: <https://doi.org/10.1016/j.bbamcr.2006.05.013>.
- [5] M. Muckenthaler et al. “Regulatory defects in liver and intestine implicate abnormal hepcidin and Cybrd1 expression in mouse hemochromatosis”. In: *Nat. Gen.* 34 (2003), pp. 102–107. DOI: <https://doi.org/10.1038/ng1152>.
- [6] N. C. Andrews and J. E. Levy. “Iron Is Hot: An Update on the Pathophysiology of Hemochromatosis”. In: *Blood* 92.6 (1998), pp. 1845–1851. DOI: <https://doi.org/10.1182/blood.V92.6.1845>.
- [7] M. D. Fleming et al. “Nramp2 is mutated in the anemic Belgrade (b) rat: Evidence of a role for Nramp2 in endosomal iron transport”. In: *Proc. Natl. Acad. Sci.* 95.3 (1998), pp. 1148–1153. DOI: [10.1073/pnas.95.3.1148](https://doi.org/10.1073/pnas.95.3.1148).
- [8] W. Li et al. “Enhanced Expression of Natural Resistance-Associated Macrophage Protein 1 in Atherosclerotic Lesions May Be Associated with Oxidized Lipid-Induced Apoptosis”. In: *Ann. New York Acad. Sci.* 1030.1 (2004), pp. 202–207. DOI: <https://doi.org/10.1196/annals.1329.026>.
- [9] T. Moos and E. H. Morgan. “The Metabolism of Neuronal Iron and Its Pathogenic Role in Neurological Disease: Review”. In: *Ann. New York Acad. Sci.* 1012.1 (2004), pp. 14–26. DOI: <https://doi.org/10.1196/annals.1306.002>.
- [10] C. Askwith and J. Kaplan. “Iron and copper transport in yeast and its relevance to human disease”. In: *Trends Biochem. Sci.* 23.4 (1998), pp. 135–138. DOI: [https://doi.org/10.1016/S0968-0004\(98\)01192-X](https://doi.org/10.1016/S0968-0004(98)01192-X).
- [11] R. R. Crichton, D. T. Dexter, and R. J. Ward. “Metal based neurodegenerative diseases—From molecular mechanisms to therapeutic strategies”. In: *Coord. Chem. Rev.* 252.10 (2008), pp. 1189–1199. DOI: <https://doi.org/10.1016/j.ccr.2007.10.019>.
- [12] P. Chen, M. R. Miah, and M. Aschner. “Metals and Neurodegeneration”. In: *F1000Research* 5.366 (2016). DOI: [10.12688/f1000research.7431.1](https://doi.org/10.12688/f1000research.7431.1).

- [13] P. Zatta et al. "The role of metals in neurodegenerative processes: aluminum, manganese, and zinc". In: *Brain Res. Bull.* 62.1 (2003), pp. 15–28. DOI: [https://doi.org/10.1016/S0361-9230\(03\)00182-5](https://doi.org/10.1016/S0361-9230(03)00182-5).
- [14] C. E. Cicero et al. "Metals and neurodegenerative diseases. A systematic review". In: *Environ. Res.* 159 (2017), pp. 82–94. DOI: <https://doi.org/10.1016/j.envres.2017.07.048>.
- [15] M. A. Hediger. "Membrane permeability. The diversity of transmembrane transport processes". In: *Curr. Opin. Cell Biol.* 9.4 (1997), pp. 543–546. DOI: [https://doi.org/10.1016/S0955-0674\(97\)80031-2](https://doi.org/10.1016/S0955-0674(97)80031-2).
- [16] N. Nelson. "Metal ion transporters and homeostasis". In: *EMBOJ.* 18.16 (1999), pp. 4361–4371. DOI: <https://doi.org/10.1093/emboj/18.16.4361>.
- [17] G. Rudnick. "How do transporters couple solute movements?" In: *Mol. Membr. Biol.* 30.7 (2013), pp. 355–359. DOI: [10.3109/09687688.2013.842658](https://doi.org/10.3109/09687688.2013.842658).
- [18] O. Boudker and G. Verdon. "Structural perspectives on secondary active transporters". In: *Trends Pharmacol. Sci.* 31.9 (2010), pp. 418–426. DOI: <https://doi.org/10.1016/j.tips.2010.06.004>.
- [19] Y. Shi. "Common Folds and Transport Mechanisms of Secondary Active Transporters". In: *Ann. Rev. Biophys.* 42.1 (2013), pp. 51–72. DOI: [10.1146/annurev-biophys-083012-130429](https://doi.org/10.1146/annurev-biophys-083012-130429).
- [20] L. R. Forrest, R. Krämer, and C. Ziegler. "The structural basis of secondary active transport mechanisms". In: *Biochim. Biophys. Acta - Bioenergetics* 1807.2 (2011), pp. 167–188. DOI: <https://doi.org/10.1016/j.bbabi.2010.10.014>.
- [21] I. C. West. "Energy coupling in secondary active transport". In: *Biochim. Biophys. Acta - Reviews on Biomembranes* 604 (1980), pp. 91–126. DOI: [https://doi.org/10.1016/0304-4157\(80\)90005-2](https://doi.org/10.1016/0304-4157(80)90005-2).
- [22] R. K. Henderson, K. Fendler, and B. Poolman. "Coupling efficiency of secondary active transporters". In: *Curr. Opin. Biotech.* 58 (2019), pp. 62–71. DOI: <https://doi.org/10.1016/j.copbio.2018.11.005>.
- [23] M. V. LeVine et al. "Allosteric Mechanisms of Molecular Machines at the Membrane: Transport by Sodium-Coupled Symporters". In: *Chem. Rev.* 116.11 (2016), pp. 6552–6587. DOI: [10.1021/acs.chemrev.5b00627](https://doi.org/10.1021/acs.chemrev.5b00627).
- [24] O. Jardetzky. "Simple Allosteric Model for Membrane Pumps". In: *Nature* 211 (1966), pp. 969–970. DOI: <https://doi.org/10.1038/211969a0>.
- [25] M. Cellier et al. "Nramp defines a family of membrane proteins." In: *Proc. Natl. Acad. Sci.* 92.22 (1995), pp. 10089–10093. DOI: [10.1073/pnas.92.22.10089](https://doi.org/10.1073/pnas.92.22.10089).
- [26] H. Gunshin et al. "Cloning and characterization of a mammalian proton-coupled metal-ion transporter". In: *Nature* 388 (1997), pp. 482–488. DOI: <https://doi.org/10.1038/41343>.

- [27] M. F. M. Cellier. "Chapter Ten - Nramp: from Sequence to Structure and Mechanism of Divalent Metal Import". In: *Metal Transporters*. Vol. 69. Curr. Topics Membr. Academic Press, 2012, pp. 249–293. DOI: <https://doi.org/10.1016/B978-0-12-394390-3.00010-0>.
- [28] F. Supek et al. "A yeast manganese transporter related to the macrophage protein involved in conferring resistance to mycobacteria." In: *Proc. Natl. Acad. Sci. USA* 93.10 (1996), pp. 5105–5110. DOI: 10.1073/pnas.93.10.5105.
- [29] J. R. Forbes and P. Gros. "Divalent-metal transport by NRAMP proteins at the interface of host–pathogen interactions". In: *Trends Microbiol.* 9.8 (2001), pp. 397–403. DOI: [https://doi.org/10.1016/S0966-842X\(01\)02098-4](https://doi.org/10.1016/S0966-842X(01)02098-4).
- [30] B. Mackenzie and M. A. Hediger. "SLC11 family of H⁺-coupled metal-ion transporters NRAMP1 and DMT1". In: *Pflügers Archiv*. 447 (2004), pp. 571–579. DOI: <https://doi.org/10.1007/s00424-003-1141-9>.
- [31] D. Agranoff et al. "*Mycobacterium tuberculosis* Expresses a Novel Ph-Dependent Divalent Cation Transporter Belonging to the Nramp Family". In: *J. Exp. Med.* 190.5 (1999), pp. 717–724. DOI: <https://doi.org/10.1084/jem.190.5.717>.
- [32] T. E. Kehl-Fie et al. "MntABC and MntH Contribute to Systemic *Staphylococcus aureus* Infection by Competing with Calprotectin for Nutrient Manganese". In: *Infection and Immunity* 81.9 (2013), pp. 3395–3405. DOI: 10.1128/IAI.00420-13.
- [33] D. G. Kehres et al. "The NRAMP proteins of *Salmonella typhimurium* and *Escherichia coli* are selective manganese transporters involved in the response to reactive oxygen". In: *Mol. Microbiol.* 36.5 (2000), pp. 1085–1100. DOI: 10.1046/j.1365-2958.2000.01922.x.
- [34] H. Makui et al. "Identification of the *Escherichia coli* K-12 Nramp orthologue (MntH) as a selective divalent metal ion transporter". In: *Mol. Microbiol.* 35.5 (2000), pp. 1065–1078. DOI: 10.1046/j.1365-2958.2000.01774.x.
- [35] S. Shabayek et al. "A streptococcal NRAMP homologue is crucial for the survival of *Streptococcus agalactiae* under low pH conditions". In: *Mol. Microbiol.* 100 (4 2016), pp. 589–606. DOI: <https://doi.org/10.1111/mmi.13335>.
- [36] S. Haiyan and T. Baoming. "Function analysis of two Mn (II) ion transporter genes (DR1709 and DR2523) in *Deinococcus radiodurans*". In: *Afr. J. Biotechnol.* 9.19 (2010), pp. 2742–2747.
- [37] J. K. Kajfasz et al. "Manganese Uptake, Mediated by SloABC and MntH, Is Essential for the Fitness of *Streptococcus mutans*". In: *mSphere* 5.1 (2020), e00764–19. DOI: 10.1128/mSphere.00764-19.
- [38] T. H. Hohle and M. R. O'Brian. "The *mntH* gene encodes the major Mn²⁺ transporter in *Bradyrhizobium japonicum* and is regulated by manganese via the Fur protein". In: *Mol. Microbiol.* 72 (2 2009), pp. 399–409. DOI: <https://doi.org/10.1111/j.1365-2958.2009.06650.x>.

- [39] R. D. Perry et al. "Manganese transporters Yfe and MntH are Fur-regulated and important for the virulence of *Yersinia pestis*". In: *Microbiology* 158.3 (2012), pp. 804–815. DOI: <https://doi.org/10.1099/mic.0.053710-0>.
- [40] O. L. Champion et al. "*Yersinia pseudotuberculosis* mntH functions in intracellular manganese accumulation, which is essential for virulence and survival in cells expressing functional Nramp1". In: *Microbiology* 157.4 (2011), pp. 1115–1122. DOI: <https://doi.org/10.1099/mic.0.045807-0>.
- [41] M. S. Turner, Y. P. Tan, and P. M. Giffard. "Inactivation of an Iron Transporter in *Lactococcus lactis* Results in Resistance to Tellurite and Oxidative Stress". In: *Appl. Environ. Microbiol.* 73.19 (2007), pp. 6144–6149. DOI: 10.1128/AEM.00413-07.
- [42] E. S. Anderson et al. "The Manganese Transporter MntH Is a Critical Virulence Determinant for *Brucella abortus* 2308 in Experimentally Infected Mice". In: *Infection and Immunity* 77.8 (2009), pp. 3466–3474. DOI: 10.1128/IAI.00444-09.
- [43] E. Boyer et al. "Acquisition of Mn(II) in Addition to Fe(II) Is Required for Full Virulence of *Salmonella enterica* Serovar Typhimurium". In: *Infection and Immunity* 70.11 (2002), pp. 6032–6042. DOI: 10.1128/IAI.70.11.6032-6042.2002.
- [44] I. Reeve et al. "Overexpression, purification, and site-directed spin labeling of the Nramp metal transporter from *Mycobacterium leprae*". In: *Proc. Natl. Acad. Sci. USA* 99.13 (2002), pp. 8608–8613. DOI: 10.1073/pnas.142287699.
- [45] X. W. Li et al. "SLC11A1 (NRAMP1) Polymorphisms and Tuberculosis Susceptibility: Updated Systematic Review and Meta-Analysis". In: *PLoS ONE* 6 (2011), pp. 1–8. DOI: 10.1371/journal.pone.0015831.
- [46] J. R. Forbes and P. Gros. "Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane". In: *Blood* 102.5 (2003), pp. 1884–1892. DOI: <https://doi.org/10.1182/blood-2003-02-0425>.
- [47] M. Wessling-Resnick. "Nramp1 and Other Transporters Involved in Metal Withholding during Infection". In: *J. Biol. Chem.* 290.31 (2015), pp. 18984–18990. DOI: <https://doi.org/10.1074/jbc.R115.643973>.
- [48] O. Cunrath and D. Bumann. "Host resistance factor SLC11A1 restricts *Salmonella* growth through magnesium deprivation". In: *Science* 366.6468 (2019), pp. 995–999. DOI: 10.1126/science.aax7898.
- [49] J. P. Lisher and D. Giedroc. "Manganese acquisition and homeostasis at the host-pathogen interface". In: *Front. Cell. Infect. Microbiol.* 3 (2013), p. 91. DOI: 10.3389/fcimb.2013.00091.
- [50] M. Indriati Hood and E. P. Skaar. "Nutritional immunity: transition metals at the pathogen-host interface". In: *Nat. Rev. Microbiol.* 10.8 (2012), pp. 525–537. DOI: <https://doi.org/10.1038/nrmicro2836>.

- [51] S. Bozzaro, S. Buracco, and B. Peracino. “Iron metabolism and resistance to infection by invasive bacteria in the social amoeba *Dictyostelium discoideum*”. In: *Front. Cell. Infect. Microbiol.* 3 (2013). DOI: 10.3389/fcimb.2013.00050.
- [52] B. Peracino et al. “Function and mechanism of action of *Dictyostelium* Nramp1 (Slc11a1) in bacterial infection”. In: *Traffic* 7.1 (2006), pp. 22–38. DOI: <https://doi.org/10.1111/j.1600-0854.2005.00356.x>.
- [53] Y. Brenz et al. “Nramp1 and NrampB Contribute to Resistance against *Francisella* in *Dictyostelium*”. In: *Front. Cell. Infect. Microbiol.* 7 (2017). DOI: 10.3389/fcimb.2017.00282.
- [54] D. Agranoff et al. “The Nramp orthologue of *Cryptococcus neoformans* is a pH-dependent transporter of manganese, iron, cobalt and nickel”. In: *Biochem. J.* 385.1 (2005), pp. 225–232. DOI: <https://doi.org/10.1042/BJ20040836>.
- [55] X. Z. Chen et al. “Yeast SMF1 Mediates H⁺-coupled Iron Uptake with Concomitant Uncoupled Cation Currents”. In: *J. Biol. Chem.* 274.49 (1999), pp. 35089–35094. DOI: <https://doi.org/10.1074/jbc.274.49.35089>.
- [56] A. Cohen, H. Nelson, and N. Nelson. “The Family of SMF Metal Ion Transporters in Yeast Cells”. In: *J. Biol. Chem.* 275.43 (2000), pp. 33388–33394. DOI: <https://doi.org/10.1074/jbc.M004611200>.
- [57] E. Pinner et al. “Functional Complementation of the Yeast Divalent Cation Transporter Family SMF by NRAMP2, a Member of the Mammalian Natural Resistance-associated Macrophage Protein Family”. In: *J. Biol. Chem.* 272.46 (1997), pp. 28933–28938. DOI: <https://doi.org/10.1074/jbc.272.46.28933>.
- [58] M. E. Portnoy, X. F. Liu, and V. Cizewski Culotta. “*Saccharomyces cerevisiae* Expresses Three Functionally Distinct Homologues of the Nramp Family of Metal Transporters”. In: *Mol. Cell. Biol.* 20.21 (2000), pp. 7893–7902. DOI: 10.1128 / MCB.20.21.7893-7902.2000.
- [59] A. Belouchi, T. Kwan, and Gros P. “Cloning and characterization of the OsNramp family from *Oryza sativa*, a new family of membrane proteins possibly implicated in the transport of metal ions”. In: *Plant Mol. Biol.* 33.6 (1997), pp. 1085–1092. DOI: 10.1023/A:1005723304911.
- [60] S. Thomine et al. “Cadmium and iron transport by members of a plant metal transporter family in *Arabidopsis* with homology to *Nramp* genes”. In: *Proc. Natl. Acad. Sci.* 97.9 (2000), pp. 4991–4996. DOI: 10.1073/pnas.97.9.4991.
- [61] R. Cailliatte et al. “High-Affinity Manganese Uptake by the Metal Transporter NRAMP1 Is Essential for *Arabidopsis* Growth in Low Manganese Conditions”. In: *The Plant Cell* 22.3 (2010), pp. 904–917. DOI: <https://doi.org/10.1105/tpc.109.073023>.
- [62] C. Curie et al. “Involvement of NRAMP1 from *Arabidopsis thaliana* in iron transport”. In: *Biochem. J.* 347.3 (2000), pp. 749–755. DOI: <https://doi.org/10.1042/bj3470749>.

- [63] A. Sasaki et al. “Nramp5 Is a Major Transporter Responsible for Manganese and Cadmium Uptake in Rice”. In: *Plant Cell* 24 (5 2012), pp. 2155–2167. DOI: <https://doi.org/10.1105/tpc.112.096925>.
- [64] I. Ullah et al. “Evolution, and functional analysis of Natural Resistance-Associated Macrophage Proteins (NRAMPs) from *Theobroma cacao* and their role in cadmium accumulation”. In: *Sci. rep.* 8.1 (2018), pp. 1–15. DOI: <https://doi.org/10.1038/s41598-018-32819-y>.
- [65] N. Yamaji et al. “A node-based switch for preferential distribution of manganese in rice”. In: *Nat. Commun.* 4.1 (2013), pp. 1–11. DOI: <https://doi.org/10.1038/ncomms3442>.
- [66] S. Alejandro et al. “Intracellular distribution of manganese by the trans-Golgi network transporter NRAMP2 is critical for photosynthesis and cellular redox homeostasis”. In: *Plant Cell* 29.12 (2017), pp. 3068–3084. DOI: <https://doi.org/10.1105/tpc.17.00578>.
- [67] C. Peris-Peris et al. “Two NRAMP6 Isoforms Function as Iron and Manganese Transporters and Contribute to Disease Resistance in Rice”. In: *Mol. Plant-Microbe Interact.* 30.5 (2017), pp. 385–398. DOI: 10.1094/MPMI-01-17-0005-R.
- [68] D. Segond et al. “NRAMP genes function in *Arabidopsis thaliana* resistance to *Erwinia chrysanthemi* infection”. In: *Plant J.* 58.2 (2009), pp. 195–207. DOI: <https://doi.org/10.1111/j.1365-3113X.2008.03775.x>.
- [69] J. Y. Li et al. “Natural variation underlies alterations in Nramp aluminum transporter (*NRAT1*) expression and function that play a key role in rice aluminum tolerance”. In: *Proc. Natl. Acad. Sci.* 111.17 (2014), pp. 6503–6508. DOI: 10.1073/pnas.1318975111.
- [70] J. Xia et al. “Plasma membrane-localized transporter for aluminum in rice”. In: *Proc. Natl. Acad. Sci.* 107.43 (2010), pp. 18381–18385. DOI: 10.1073/pnas.1004949107.
- [71] C. Au et al. “SMF-1, SMF-2 and SMF-3 DMT1 Orthologues Regulate and Are Regulated Differentially by Manganese Levels in *C. elegans*”. In: *PLoS ONE* 4.11 (2009), pp. 1–17. DOI: 10.1371/journal.pone.0007792.
- [72] J. Martínez-Barnette et al. “Cloning and functional characterization of the *Anopheles albimanus* DMT1/NRAMP homolog: Implications in iron metabolism in mosquitoes”. In: *Insect Biochem. Mol. Biol.* 37.6 (2007), pp. 532–539. DOI: <https://doi.org/10.1016/j.ibmb.2007.02.009>.
- [73] J. V. Neves et al. “Natural history of SLC11 genes in vertebrates: tales from the fish world”. In: *BMC Evol. Biol.* 11 (2011), p. 106. DOI: <https://doi.org/10.1186/1471-2148-11-106>.
- [74] M. F. Cellier, P. Courville, and C. Campion. “Nramp1 phagocyte intracellular metal withdrawal defense”. In: *Microb. Infect.* 9.14 (2007), pp. 1662–1670. DOI: <https://doi.org/10.1016/j.micinf.2007.09.006>.
- [75] E. E. Johnson and M. Wessling-Resnick. “Iron metabolism and the innate immune response to infection”. In: *Microbes and Infection* 14.3 (2012), pp. 207–216. DOI: <https://doi.org/10.1016/j.micinf.2011.10.001>.

- [76] N. Montalbetti et al. "Mammalian iron transporters: Families SLC11 and SLC40". In: *Mol. Asp. Med.* 34.2 (2013), pp. 270–287. DOI: <https://doi.org/10.1016/j.mam.2013.01.002>.
- [77] N. C. Andrews. "Forging a field: the golden age of iron biology". In: *Blood* 112.2 (2008), pp. 219–230. DOI: <https://doi.org/10.1182/blood-2007-12-077388>.
- [78] R. Coffey and T. Ganz. "Iron homeostasis: An anthropocentric perspective". In: *J. Biol. Chem.* 292.31 (2017), pp. 12727–12734. DOI: <https://doi.org/10.1074/jbc.R117.781823>.
- [79] M. D. Knutson. "Iron transport proteins: Gateways of cellular and systemic iron homeostasis". In: *J. Biol. Chem.* 292.31 (2017), pp. 12735–12743. DOI: <https://doi.org/10.1074/jbc.R117.786632>.
- [80] M. Cellier et al. "Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression." In: *J. Exp. Med.* 180.5 (1994), pp. 1741–1752. DOI: <https://doi.org/10.1084/jem.180.5.1741>.
- [81] S. M. Vidal et al. "Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg". In: *Cell* 73.3 (1993), pp. 469–485. DOI: [https://doi.org/10.1016/0092-8674\(93\)90135-D](https://doi.org/10.1016/0092-8674(93)90135-D).
- [82] J. R. Forbes and P. Gros. "Iron, manganese, and cobalt transport by Nramp1 (Slc11a1) and Nramp2 (Slc11a2) expressed at the plasma membrane". In: *Blood* 102.5 (2003), pp. 1884–1892. DOI: <https://doi.org/10.1182/blood-2003-02-0425>.
- [83] N. Jabado et al. "Natural resistance to intracellular infections: natural resistance-associated macrophage protein 1 (Nramp1) functions as a pH-dependent manganese transporter at the phagosomal membrane". In: *J. Exp. Med.* 192.9 (2000), pp. 1237–1248. DOI: <https://doi.org/10.1084/jem.192.9.1237>.
- [84] G. Govoni et al. "Functional Expression of Nramp1 In Vitro in the Murine Macrophage Line RAW264.7". In: *Infect. Immun.* 67.5 (1999), pp. 2225–2232. DOI: [10.1128/IAI.67.5.2225-2232.1999](https://doi.org/10.1128/IAI.67.5.2225-2232.1999).
- [85] E. Skamene et al. "Genetic regulation of resistance to intracellular pathogens". In: *Nature* 297.5866 (1982), pp. 506–509. DOI: <https://doi.org/10.1038/297506a0>.
- [86] S.M. Vidal et al. "Natural Resistance to Intracellular Infections: Nramp1 Encodes a Membrane Phosphoglycoprotein Absent in Macrophages from Susceptible (Nramp1 D169) Mouse Strains". In: *J. Immunol.* 157.8 (1996), pp. 3559–3568. URL: [URL%20=%20%7Bhttps://www.jimmunol.org/content/157/8/3559%7D,](https://www.jimmunol.org/content/157/8/3559%7D)
- [87] F. Canonne-Hergaux et al. "The Nramp2/DMT1 iron transporter is induced in the duodenum of microcytic anemia mk mice but is not properly targeted to the intestinal brush border". In: *Blood* 96.12 (2000), pp. 3964–3970. DOI: <https://doi.org/10.1182/blood.V96.12.3964>.

- [88] F. Canonne-Hergaux et al. "Cellular and Subcellular Localization of the Nramp2 Iron Transporter in the Intestinal Brush Border and Regulation by Dietary Iron". In: *Blood* 93.12 (1999), pp. 4406–4417. DOI: <https://doi.org/10.1182/blood.V93.12.4406>.
- [89] A. Donovan et al. "Positional cloning of zebrafish *ferroportin1* identifies a conserved vertebrate iron exporter". In: *Nature* 403.6771 (2000), pp. 776–781. DOI: <https://doi.org/10.1038/35001596>.
- [90] A. T. McKie et al. "A Novel Duodenal Iron-Regulated Transporter, IREG1, Implicated in the Basolateral Transfer of Iron to the Circulation". In: *Mol. Cell* 5.2 (2000), pp. 299–309. DOI: [https://doi.org/10.1016/S1097-2765\(00\)80425-6](https://doi.org/10.1016/S1097-2765(00)80425-6).
- [91] C. D. Vulpe et al. "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron transport, is defective in the *sla* mouse". In: *Nat. gen.* 21.2 (1999), pp. 195–199. DOI: <https://doi.org/10.1038/5979>.
- [92] K. Gkouvatsos, G. Papanikolaou, and K. Pantopoulos. "Regulation of iron transport and the role of transferrin". In: *Biochim. Biophys. Acta* 1820.3 (2012), pp. 188–202. DOI: <https://doi.org/10.1016/j.bbagen.2011.10.013>.
- [93] R. S. Ohgami et al. "Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells". In: *Nat. Gen.* 37 (2005), pp. 1264–1269. DOI: <https://doi.org/10.1038/ng1658>.
- [94] M. Knöpfel, L. Zhao, and M. D. Garrick. "Transport of Divalent Transition-Metal Ions Is Lost in Small-Intestinal Tissue of *b/b* Belgrade Rats". In: *Biochem.* 44.9 (2005), pp. 3454–3465. DOI: 10.1021/bi048768+.
- [95] A. Iolascon et al. "Microcytic anemia and hepatic iron overload in a child with compound heterozygous mutations in DMT1 (SCL11A2)". In: *Blood* 107.1 (2006), pp. 349–354. DOI: <https://doi.org/10.1182/blood-2005-06-2477>.
- [96] M. D. Fleming et al. "*Nramp2* is mutated in the anemic Belgrade (*b*) rat: Evidence of a role for Nramp2 in endosomal iron transport". In: *Proc. Natl. Acad. Sci.* 95.3 (1998), pp. 1148–1153. DOI: 10.1073/pnas.95.3.1148.
- [97] M. D. Fleming et al. "Microcytic anaemia mice have a mutation in *Nramp2*, a candidate iron transporter gene". In: *Nat. gen.* 16.4 (1997), pp. 383–386. DOI: <https://doi.org/10.1038/ng0897-383>.
- [98] F. Canonne-Hergaux et al. "Characterization of the iron transporter DMT1 (NRAMP2/DCT1) in red blood cells of normal and anemic *mk/mk* mice". In: *Blood* 98.13 (2001), pp. 3823–3830. DOI: <https://doi.org/10.1182/blood.V98.13.3823>.
- [99] M. A. Su et al. "The G185R Mutation Disrupts Function of the Iron Transporter Nramp2". In: *Blood* 92.6 (1998), pp. 2157–2163. DOI: <https://doi.org/10.1182/blood.V92.6.2157>.
- [100] E. Blanco et al. "Not all DMT1 mutations lead to iron overload". In: *Blood Cells Mol. Dis.* 43.2 (2009), pp. 199–201. DOI: <https://doi.org/10.1016/j.bcmed.2009.05.003>.

- [101] M. P. Mims et al. "Identification of a human mutation of *DMT1* in a patient with microcytic anemia and iron overload". In: *Blood* 105.3 (2005), pp. 1337–1342. DOI: <https://doi.org/10.1182/blood-2004-07-2966>.
- [102] E. Bardou-Jacquet et al. "A novel N491S mutation in the human *SLC11A2* gene impairs protein trafficking and in association with the G212V mutation leads to microcytic anemia and liver iron overload". In: *Blood Cells Mol. Dis.* 47.4 (2011), pp. 243–248. DOI: <https://doi.org/10.1016/j.bcmd.2011.07.004>.
- [103] C. Beaumont et al. "Two new human *DMT1* gene mutations in a patient with microcytic anemia, low ferritinemia, and liver iron overload". In: *Blood* 107.10 (2006), pp. 4168–4170. DOI: <https://doi.org/10.1182/blood-2005-10-4269>.
- [104] S. Lam-Yuk-Tseung et al. "A novel R416C mutation in human *DMT1* (*SLC11A2*) displays pleiotropic effects on function and causes microcytic anemia and hepatic iron overload". In: *Blood Cells Mol. Dis.* 36.3 (2006), pp. 347–354. DOI: <https://doi.org/10.1016/j.bcmd.2006.01.011>.
- [105] A. Cohen, Y. Nevo, and N. Nelson. "The first external loop of the metal ion transporter DCT1 is involved in metal ion binding and specificity". In: *Proc. Natl. Acad. Sci.* 100.19 (2003), pp. 10694–10699. DOI: [10.1073/pnas.1934572100](https://doi.org/10.1073/pnas.1934572100).
- [106] Y. Nevo. "Site-directed mutagenesis investigation of coupling properties of metal ion transport by DCT1". In: *Biochim. Biophys. Acta (BBA) - Biomembranes* 1778.1 (2008), pp. 334–341. DOI: <https://doi.org/10.1016/j.bbamem.2007.10.007>.
- [107] V. Picard et al. "Nramp 2 (DCT1/DMT1) Expressed at the Plasma Membrane Transports Iron and Other Divalent Cations into a Calcein-accessible Cytoplasmic Pool". In: *J. Biol. Chem.* 275.46 (2000), pp. 35738–35745. DOI: <https://doi.org/10.1074/jbc.M005387200>.
- [108] A. C. Illing et al. "Substrate Profile and Metal-ion Selectivity of Human Divalent Metal-ion Transporter-1". In: *J. Biol. Chem.* 287.36 (2012), pp. 30485–30496. DOI: <https://doi.org/10.1074/jbc.M112.364208>.
- [109] P. Marciani et al. "Modulation of DMT1 activity by redox compounds". In: *J. Membr. Biol.* 197.2 (2004), pp. 91–99. DOI: <https://doi.org/10.1007/s00232-003-0644-9>.
- [110] Y. Nevo and N. Nelson. "The Mutation F227I Increases the Coupling of Metal Ion Transport in DCT1". In: *J. Biol. Chem.* 279.51 (2004), pp. 53056–53061. DOI: <https://doi.org/10.1074/jbc.M408398200>.
- [111] M. Pottier et al. "Identification of mutations allowing Natural Resistance Associated Macrophage Proteins (NRAMP) to discriminate against cadmium". In: *Plant J.* 83.4 (2015), pp. 625–637. DOI: <https://doi.org/10.1111/tpj.12914>.
- [112] J. Li et al. "A Functional Study Identifying Critical Residues Involving Metal Transport Activity and Selectivity in Natural Resistance-Associated Macrophage Protein 3 in *Arabidopsis thaliana*". In: *Int. J. Mol. Sci.* 19.5 (2018), p. 1430. DOI: [10.3390/ijms19051430](https://doi.org/10.3390/ijms19051430).
- [113] M. Gantner et al. "Electrophysiology Measurements of Metal Transport by MntH2 from *Enterococcus faecalis*". In: *Membranes* 10.10 (2020), p. 255. DOI: [10.3390/membranes10100255](https://doi.org/10.3390/membranes10100255).

- [114] B. Mackenzie et al. "Divalent metal-ion transporter DMT1 mediates both H^+ -coupled Fe^{2+} transport and uncoupled fluxes". In: *Pflügers Arch.* 451.4 (2006), pp. 544–558. DOI: <https://doi.org/10.1007/s00424-005-1494-3>.
- [115] W. Lan et al. "A facile transport assay for H^+ coupled membrane transport using fluorescence probes". In: *Anal. Methods* 4.1 (2012), pp. 44–46. DOI: <https://doi.org/10.1039/C1AY05549F>.
- [116] N. A. Wolff et al. "A role for divalent metal transporter (DMT1) in mitochondrial uptake of iron and manganese". In: *Sci. rep.* 8.1 (2018), pp. 1–12. DOI: <https://doi.org/10.1038/s41598-017-18584-4>.
- [117] S. Tandy et al. "Nramp2 Expression Is Associated with pH-dependent Iron Uptake across the Apical Membrane of Human Intestinal Caco-2 Cells". In: *J. Biol. Chem.* 275.2 (2000), pp. 1023–1029. DOI: <https://doi.org/10.1074/jbc.275.2.1023>.
- [118] M. T. Worthington et al. "Functional properties of transfected human DMT1 iron transporter". In: *Am. J. Physiol.-Gastroint. Liver Physiol.* 279.6 (2000), G1265–G1273. DOI: [10.1152/ajpgi.2000.279.6.G1265](https://doi.org/10.1152/ajpgi.2000.279.6.G1265).
- [119] S. Buracco et al. "*Dictyostelium* Nramp1, which is structurally and functionally similar to mammalian DMT1 transporter, mediates phagosomal iron efflux". In: *J. cell sci.* 128.17 (2015), pp. 3304–3316. DOI: <https://doi.org/10.1242/jcs.173153>.
- [120] K. Yokosho, N. Yamaji, and J. F. Ma. "Buckwheat FeNramp5 Mediates High Manganese Uptake in Roots". In: *Plant Cell Physiol.* 62.4 (2021), pp. 600–609. DOI: [10.1093/pcp/pcaa153](https://doi.org/10.1093/pcp/pcaa153).
- [121] I. A. Ehrnstorfer et al. "Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport". In: *Nat. Struct. Mol. Biol.* 21.11 (2014), pp. 990–996. DOI: [10.1038/nsmb.2904](https://doi.org/10.1038/nsmb.2904).
- [122] I. A. Ehrnstorfer et al. "Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family". In: *Nature Commun.* 8.1 (2017), p. 14033. DOI: [10.1038/ncomms14033](https://doi.org/10.1038/ncomms14033).
- [123] A. T. Bozzi et al. "Crystal Structure and Conformational Change Mechanism of a Bacterial Nramp-Family Divalent Metal Transporter". In: *Structure* 24.12 (2016), pp. 2102–2114. DOI: <https://doi.org/10.1016/j.str.2016.09.017>.
- [124] A. T. Bozzi et al. "Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter". In: *eLife* 8 (2019), e41124. DOI: [10.7554/eLife.41124](https://doi.org/10.7554/eLife.41124).
- [125] B. Mackenzie et al. "Functional properties of multiple isoforms of human divalent metal-ion transporter 1 (DMT1)". In: *Biochem. J.* 403.1 (2007), pp. 59–69. DOI: <https://doi.org/10.1042/BJ20061290>.
- [126] D. I. Bannon et al. "Uptake of lead and iron by divalent metal transporter 1 in yeast and mammalian cells". In: *Biochem. Biophys. Res. Commun.* 295.4 (2002), pp. 978–984. DOI: [https://doi.org/10.1016/S0006-291X\(02\)00756-8](https://doi.org/10.1016/S0006-291X(02)00756-8).

- [127] M. Okubo et al. "Cadmium transport by human Nramp 2 expressed in *Xenopus laevis* oocytes". In: *Toxicol. App. Pharmacol.* 187.3 (2003), pp. 162–167. DOI: [https://doi.org/10.1016/S0041-008X\(02\)00078-9](https://doi.org/10.1016/S0041-008X(02)00078-9).
- [128] D. I. Bannon et al. "Effect of DMT1 knockdown on iron, cadmium, and lead uptake in Caco-2 cells". In: *Am. J. Physiol.-Cell Physiol.* 284.1 (2003), pp. C44–C50. DOI: [10.1152/ajpcell.00184.2002](https://doi.org/10.1152/ajpcell.00184.2002).
- [129] M. Vázquez et al. "Participation of divalent cation transporter DMT1 in the uptake of inorganic mercury". In: *Toxicol.* 331 (2015), pp. 119–124. DOI: <https://doi.org/10.1016/j.tox.2015.03.005>.
- [130] M. Arredondo et al. "Mouse divalent metal transporter 1 is a copper transporter in HEK293 cells". In: *BioMetals* 27.1 (2014), pp. 115–123. DOI: <https://doi.org/10.1007/s10534-013-9691-6>.
- [131] A. Shawki and B. Mackenzie. "Interaction of calcium with the human divalent metal-ion transporter-1". In: *Biochem. Biophys. Res. Commun.* 393.3 (2010), pp. 471–475. DOI: <https://doi.org/10.1016/j.bbrc.2010.02.025>.
- [132] T. Mizuno et al. "Cloning of three ZIP/Nramp transporter genes from a Ni hyperaccumulator plant *Thlaspi japonicum* and their Ni²⁺-transport abilities". In: *Plant Physiol. Biochem.* 43.8 (2005), pp. 793–801. DOI: <https://doi.org/10.1016/j.plaphy.2005.07.006>.
- [133] M. Tiwari et al. "Expression in *Arabidopsis* and cellular localization reveal involvement of rice NRAMP, OsNRAMP 1, in arsenic transport and tolerance". In: *Plant, Cell Environ.* 37.1 (2014), pp. 140–152. DOI: <https://doi.org/10.1111/pce.12138>.
- [134] T. Ueki, N. Furuno, and H. Michibata. "A novel vanadium transporter of the Nramp family expressed at the vacuole of vanadium-accumulating cells of the ascidian *Ascidia sydneiensis samea*". In: *Biochim. Biophys. Acta* 1810.4 (2011), pp. 457–464. DOI: <https://doi.org/10.1016/j.bbagen.2010.12.006>.
- [135] R. Chaloupka et al. "Identification of Functional Amino Acids in the Nramp Family by a Combination of Evolutionary Analysis and Biophysical Studies of Metal and Proton Cotransport in Vivo". In: *Biochem.* 44.2 (2005), pp. 726–733. DOI: [10.1021/bi048014v](https://doi.org/10.1021/bi048014v).
- [136] A. Yamashita et al. "Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters". In: *Nature* 437.7056 (2005), pp. 215–223. DOI: <https://doi.org/10.1038/nature03978>.
- [137] P. Courville et al. "Solute Carrier 11 Cation Symport Requires Distinct Residues in Transmembrane Helices 1 and 6". In: *J. Biol. Chem.* 283.15 (2008), pp. 9651–9658. DOI: <https://doi.org/10.1074/jbc.M709906200>.
- [138] M. F. M. Cellier. "Nutritional Immunity: Homology Modeling of Nramp Metal Import". In: *Current Topics in Innate Immunity II*. New York: Springer New York, 2012, pp. 335–351. DOI: [10.1007/978-1-4614-0106-3_19](https://doi.org/10.1007/978-1-4614-0106-3_19).
- [139] A. Vastermark et al. In: *Proteins: Structure, Function, and Bioinformatics* 82.10 (2014), pp. 2797–2811. DOI: <https://doi.org/10.1002/prot.24643>.

- [140] C. Manatschal et al. “Mechanistic basis of the inhibition of SLC11/NRAMP-mediated metal ion transport by bis-isothiourea substituted compounds”. In: *eLife* 8 (2019), e51913. DOI: 10.7554/eLife.51913.
- [141] A. T. Bozzi and R. Gaudet. “Molecular Mechanism of Nramp-Family Transition Metal Transport”. In: *J. Mol. Biol.* 433.16 (2021), p. 166991. DOI: <https://doi.org/10.1016/j.jmb.2021.166991>.
- [142] S. Lam-Yuk-Tseung et al. “Iron transport by Nramp2/DMT1: pH regulation of transport by 2 histidines in transmembrane domain 6”. In: *Blood* 101.9 (2003), pp. 3699–3707. DOI: <https://doi.org/10.1182/blood-2002-07-2108>.
- [143] A. T. Bozzi et al. “Conserved methionine dictates substrate preference in Nramp-family divalent metal transporters”. In: *Proc. Natl. Acad. Sci.* 113.37 (2016), pp. 10310–10315. DOI: 10.1073/pnas.1607734113.
- [144] J. Pujol-Giménez, M. A. Hediger, and G. Gyimesi. “A novel proton transfer mechanism in the SLC11 family of divalent metal ion transporters”. In: *Sci. Rep.* 7.1 (2017), p. 6194. DOI: 10.1038/s41598-017-06446-y.
- [145] W. Humphrey, A. Dalke, and K. Schulten. “VMD: Visual molecular dynamics”. In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [146] *VMD Visual Molecular Dynamics*. URL: <https://www.ks.uiuc.edu/Research/vmd/>.
- [147] A. T. Bozzi et al. “Unique structural features in an Nramp metal transporter impart substrate-specific proton cotransport and a kinetic bias to favor import”. In: *J. Gen. Physiol.* 151.12 (2019), pp. 1413–1429. DOI: <https://doi.org/10.1085/jgp.201912428>.
- [148] A. Sacher, A. Cohen, and N. Nelson. “Properties of the mammalian and yeast metal-ion transporters DCT1 and Smf1p expressed in *Xenopus laevis* oocytes”. In: *J. Exp. Biol.* 204.6 (2001), pp. 1053–1061. DOI: <https://doi.org/10.1242/jeb.204.6.1053>.
- [149] N. Nelson, A. Sacher, and H. Nelson. “The significance of molecular slips in transport systems”. In: *Nat. Rev. Mol. Cell. Biol.* 3 (2002), pp. 876–881. DOI: <https://doi.org/10.1038/nrm955>.
- [150] P. Courville et al. “Determination of Transmembrane Topology of the *Escherichia coli* Natural Resistance-associated Macrophage Protein (Nramp) Ortholog”. In: *J. Biol. Chem.* 279.5 (2004), pp. 3318–3326. DOI: <https://doi.org/10.1074/jbc.M309913200>.
- [151] *BioRender*. URL: <https://biorender.com/>.

2 Theoretical methods

In this chapter, the main theoretical methods used in this thesis will be presented together with their underlying theory.

2.1 Density Functional Theory

Density Functional Theory (DFT) differs from wavefunction based electronic structure methods in that it uses the electron density $\rho(\mathbf{r})$ as the central quantity. For a system containing N electrons, it is defined as

$$\rho(\mathbf{r}) = \int \cdots \int |\Psi(r_1, r_2, r_3, \dots, r_N)|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N \quad (2.1)$$

where $\Psi(r_1, r_2, r_3, \dots, r_N)$ is the (suitably normalized) many-electron wavefunction in the position representation. $\rho(\mathbf{r})$ determines the probability to find any of the N electrons at the position \mathbf{r} . It is a non-negative function of only three variables (x, y, z) and integrates to the total number of electrons:

$$\int \rho(\mathbf{r}) d\mathbf{r} = N \quad (2.2)$$

The main advantage of using the electron density over the wavefunction is the major reduction in the dimensionality of the problem: regardless of the number of electrons contained in the system the density is always three-dimensional. This allows for the description of large systems (containing up to thousands of atoms) that *ab initio* methods would not be able to treat. Partly because of this, DFT has become the most widely used electronic structure approach today. For authoritative and extensive discussions on DFT, the author can refer to a

wide range of review articles [1–4] and textbooks [5].

2.1.1 Hohenberg-Kohn Theorems

Density Functional Theory was born in 1964 with the seminal paper by Hohenberg and Kohn [6]. The two key results of this paper are known as "Hohenberg-Kohn theorems" and are at the basis of the DFT method. The authors start by considering an arbitrary number N of electrons moving under the influence of an external potential $v(\mathbf{r})$ (due to the nuclei) and of the mutual Coulomb repulsion.

The first theorem states that the external potential $v(\mathbf{r})$ is uniquely determined by $\rho(\mathbf{r})$, apart from an additive constant. This can be proved in a simple yet extremely elegant manner by using the principle of *reductio ad absurdum*. We start by assuming to know the exact ground-state electron density of the system $\rho(\mathbf{r})$ with an external potential $v(\mathbf{r})$, the Hamiltonian \hat{H} and the associated wavefunction Ψ . Let's suppose that another external potential $v'(\mathbf{r})$, with Hamiltonian \hat{H}' and ground-state wavefunction Ψ' , determines the same electron density $\rho(\mathbf{r})$. Now Ψ and Ψ' cannot clearly be equal since they are solution to different Schrödinger equations. If we denote the eigenvalues associated to Ψ and Ψ' with E_0 and E'_0 , respectively, we have:

$$\begin{aligned} E_0 < \langle \Psi' | \hat{H} | \Psi' \rangle &= \langle \Psi' | \hat{H}' | \Psi' \rangle + \langle \Psi' | \hat{H} - \hat{H}' | \Psi' \rangle \\ &= E'_0 + \int \rho(\mathbf{r}) [v(\mathbf{r}) - v'(\mathbf{r})] d\mathbf{r} \end{aligned} \quad (2.3)$$

Similarly, we can get:

$$\begin{aligned} E'_0 < \langle \Psi | \hat{H}' | \Psi \rangle &= \langle \Psi | \hat{H} | \Psi \rangle + \langle \Psi | \hat{H}' - \hat{H} | \Psi \rangle \\ &= E_0 - \int \rho(\mathbf{r}) [v(\mathbf{r}) - v'(\mathbf{r})] d\mathbf{r} \end{aligned} \quad (2.4)$$

Now, by adding eq. 2.3 and 2.4, we get the absurd result:

$$E_0 + E'_0 < E'_0 + E_0 \quad (2.5)$$

Therefore, there are no two different external potentials $v(\mathbf{r})$ and $v'(\mathbf{r})$ that can give the same electron density $\rho(\mathbf{r})$, that is $\rho(\mathbf{r})$ uniquely determines $v(\mathbf{r})$ (up to an additive constant), the associated Hamiltonian and ground state wavefunction and thus all the related ground state

properties.

Now the total energy can be expressed as an explicit functional of $\rho(\mathbf{r})$ as follows:

$$\begin{aligned} E[\rho] &= T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \\ &= \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{HK}[\rho] \end{aligned} \quad (2.6)$$

where, for the correct $\rho(\mathbf{r})$, $E[\rho]$ yields the ground state energy E_0 . The functional F_{HK} :

$$F_{HK}[\rho] = T[\rho] + V_{ee}[\rho] \quad (2.7)$$

is independent of any external potential and, therefore, is a universal functional of $\rho(\mathbf{r})$.

The second Hohenberg-Kohn theorem states that the ground state energy can be obtained variationally, with the electron density that minimizes the total energy being the exact ground state density:

$$E_0[\rho_0] \leq E[\rho] \quad (2.8)$$

It can be proved as follows: let's assume that the ground state wavefunction Ψ and its related electron density is ρ , uniquely define the external potential $v(\mathbf{r})$. If we now consider a new wavefunction Ψ' , different from Ψ , and whose electron density is ρ' , then we obtain:

$$\langle \Psi' | \hat{H} | \Psi' \rangle = \int \rho'(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{HK}[\rho'] = E[\rho'] > E_0[\rho_0] \quad (2.9)$$

2.1.2 Kohn-Sham Equations

From the Hohenberg-Kohn theorems we know that it is possible to obtain the ground state energy of the system by minimizing the energy functional:

$$E_0[\rho(\mathbf{r})] = \min_{\rho \rightarrow N} \left(\int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r} + F_{HK}[\rho(\mathbf{r})] \right) \quad (2.10)$$

Although with their theorems Hohenberg and Kohn proved that in principle the total energy of a system can be obtained from its ground state electron density, it was still unknown how to evaluate both $\rho(\mathbf{r})$ and $F_{HK}[\rho(\mathbf{r})]$. The issue was addressed in 1965 by Kohn and Sham [7]. They realized that the main failure of the seminal Thomas-Fermi model [8, 9] mainly arises from a bad description of the electronic kinetic energy. To solve the problem, they introduced the idea of a fictitious system of non-interacting electrons generating the same electron density as the real system of interacting particles and that can be expanded in terms of single-particle orbitals:

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i^{KS}(\mathbf{r})|^2 \quad (2.11)$$

These considerations led them to the development of a set of N single-particle, Schrödinger-like equations:

$$\left(-\frac{1}{2} \nabla_i^2 + v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad (2.12)$$

where ε_i represents the energy of the KS single-particle orbital while the terms on the left side of eq. 2.12 represent the kinetic energy of the non-interacting electrons, the external potential, the Hartree potential and the exchange-correlation potential, respectively. The latter includes the non-classical contributions to the electron-electron interactions and the difference between the real system kinetic energy and that of the non-interacting fictitious one. The exchange-correlation potential $v_{xc}(\mathbf{r})$ is defined as:

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho]}{\rho(\mathbf{r})} \quad (2.13)$$

with $E_{xc}[\rho]$ being the so called exchange-correlation functional.

If we now define an effective potential v_{eff} as:

$$v_{eff} = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{xc}(\mathbf{r}) \quad (2.14)$$

eq. 2.12 can be rewritten in the compact form:

$$\left(-\frac{1}{2}\nabla_i^2 + v_{eff}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad (2.15)$$

Similar to the case of the single-particle Hartree-Fock equations, this set of coupled KS single-particle equations has to be solved iteratively. The total energy can be finally determined from the resulting single-particle eigenvalues and the density through the equation:

$$E = \sum_{i=1}^N \varepsilon_i - \frac{1}{2} \iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho] - \int v_{xc}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} \quad (2.16)$$

2.1.3 Exchange-Correlation Functionals

If each term of the Kohn-Sham energy functional would be known we would theoretically be able to compute the exact ground state electron density, total ground state energy as well as all other ground state exchange-correlation term E_{xc} is not yet known. Therefore, since the birth of DFT, approximations for E_{xc} have been used and nowadays a seemingly endless list of approximated exchange-correlation functionals with varying levels of complexity have been developed. In the following, some of the most commonly used types of exchange-correlation functionals will be briefly presented. A comprehensive treatment of the topic can be found in ref. [10, 11].

Local Density Approximation (LDA)

The Local Density Approximation (LDA) already proposed by Kohn and Sham represents the simplest approximation for the exchange-correlation functional. In its general formulation, E_{xc}^{LDA} can be written in the following form:

$$E_{xc}^{LDA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r}) \varepsilon_{xc}^{LDA}(\rho(\mathbf{r})) d\mathbf{r} \quad (2.17)$$

where ε_{xc}^{LDA} is the exchange-correlation energy per particle of a uniform electron gas of density ρ . The exchange-correlation energy of an inhomogeneous system is approximated at every point in space as that of a uniform electron gas with the same local density $\rho(\mathbf{r})$. The exchange energy of a uniform electron gas is exactly known while the correlation energy is usually obtained by fitting to Quantum Monte Carlo results on uniform electron gases with various densities [12, 13]. Strictly, the LDA is valid only for uniform or at least slowly varying densities and LDA functionals work indeed surprisingly well for metals. Modern LDA

functionals tend to be quite similar, differing only in how their correlation contributions have been fitted to the many-body uniform electron gas data. A partial explanation for the surprising success of the LDA even for rather inhomogeneous systems is a systematic error cancellation between an overestimation of correlation and an underestimation of exchange contributions. Furthermore, for any density, the LDA satisfies a number of so-called “sum rules” [14–16]. Among the most famous LDA exchange-correlation functionals we can list the PZ81 [17], the PW92 [18], CP [19] and the VWN [20].

Generalized Gradient Approximation (GGA)

It was realized very early on that the local uniform density of a homogeneous electron gas (HEG) does not always represent a reasonable approximation for the rapidly varying electron densities of many materials and molecules, and that the gradient of the density $\nabla\rho(\mathbf{r})$ needs to be included as well. A first attempt to include the density gradient resulted in the so-called “gradient-expansion approximations” (GEA) [6]. Although it was found not to work well in practise, it provided the basis for the birth of the “generalised gradient approximation” (GGA) semi-local functionals, whose general formulation is:

$$E_{xc}^{GGA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r}) \varepsilon_{xc}^{GGA}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})) d\mathbf{r}. \quad (2.18)$$

Here the gradient of the density as well as the density itself is taken into account at each point in space, thus providing a correction to the LDA and at the same time ensuring the fulfilment of some of the above mentioned “sum rules”. For geometries and ground state energies of molecules and solids, GGAs work better than the LDA. Some of the most used GGAs functionals are the PW91 [21–23], the PBE [24, 25] and its several off-springs: revPBE [26], RPBE [27], PBE-WC [28] and PBEsol [29]. The so called meta-GGA functionals use as additional degrees of freedom the second derivative of the density $\nabla^2\rho(\mathbf{r})$ or the kinetic energy density:

$$\tau_{\sigma}(\mathbf{r}) = \frac{1}{2} \sum_i |\nabla\phi_i(\mathbf{r})|^2 \quad (2.19)$$

An example of a meta-GGA functional is TPSS [30].

Hybrid functionals

In principle, the exchange-correlation energy could be obtained by simply combining a density functional for correlation with the Hartree-Fock exact exchange energy. This is the principle

behind the development of the so called "hybrid functionals" in which the exchange energy is obtained as a mixture between the exact HF expression and a DFT contribution while the correlation energy is evaluated at the DFT level. Usually, both the non-HF contribution to the exchange energy and the correlation energy are treated at the generalized gradient approximation level. A general formulation for this latter case is the following:

$$E_{xc}^{hybrid} = c_x E_x^{HF} + (1 - c_x) E_x^{GGA} + E_c^{GGA} \quad (2.20)$$

As an example, the hybrid functional PBE0 [31] is using the GGA functional PBE. Undoubtedly, the most popular hybrid functional, especially among the quantum chemistry community, is the B3LYP functional composed of the Becke's 3 parameters functional for the exchange energy [32] and the Lee, Yang and Parr (LYP) functional for the correlation energy [33].

2.2 Classical Molecular Dynamics

Molecules are composed of interacting electrons and nuclei and chemical bonds can be described as electronic clouds formed by electrons shared and delocalized between nuclei. However, the large time scale difference of nuclear and electronic motions due to the high ratio between their masses, justifies the separation of electronic and nuclear variables. The electron cloud equilibrates quickly for each instantaneous configuration of the nuclei, which results to be quasi-static on the time scale of electronic motion. The nuclei, in turn, move in the field of the averaged electron densities. This is known as Born-Oppenheimer approximation. In principle, assuming that the Born-Oppenheimer approximation is valid, we know how to exactly determine the potential energy surface of the system. We need indeed to solve the time-independent electronic Schrödinger equation on a dense enough mesh of points in configuration space \mathbf{R} . Subsequently, we need to fit the set of numerical values in order to get an appropriate analytical form. However, it is easy to understand that such a procedure is practically unfeasible for systems containing more than few atoms.

Classical molecular dynamics (MD) is based on empirically parameterized potential energy surfaces, referred to as "force fields", in which the essential physics underlying interatomic interactions is captured and condensed into a simple analytic form. The parameters are derived as much as possible from experimental data or *ab initio* calculations, so that the empirical potential properly mimics the *ab initio* Born-Oppenheimer surface [34]. As a consequence, the construction of a good force field represents a nontrivial task [35]. On the other side it leads to tremendous computational simplifications. A force field typically used for classical molecular dynamics simulations has this form:

$$\begin{aligned}
U(\mathbf{r}_1, \dots, \mathbf{r}_N) = & \sum_{bonds} \frac{k_b}{2} (l_i - l_{i0})^2 + \sum_{angles} \frac{k_\theta}{2} (\theta_i - \theta_{i0})^2 \\
& + \sum_{torsions} k_d [1 + \cos(n\omega_i - \gamma_i)] \\
& + \sum_{atompairs} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
& + \sum_{atompairs} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.21}
\end{aligned}$$

where $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$ represents the potential energy of N interacting atoms as a function of their positions $\mathbf{r}_i = (x_i, y_i, z_i)$.

The first three terms of the equation describe the so called "bonded interactions" and the summation indices run indeed over all the bonds, angles and dihedrals defined by the covalent bonds existing in the system, respectively. In particular, the first two terms describe the potential energies for the deformation of the bonds lengths l_i (bond stretching) and angles θ_i (angle bending) from their respective equilibrium values l_{i0} and θ_{i0} . The harmonic form of these two terms, with force constants for bond stretching and angle bending k_b and k_θ respectively, describes chemical bonds but does not allow for the description of events such as bond breaking or formation. The third term defines a torsional potential for the twisting of two adjacent planes of atoms against each other. It has a periodic form with periodicity characterized by n and heights of rotational barriers k_d .

The last two terms of eq. 2.21 describe the so-called "nonbonded interactions". The summation indices run over all the pairs of not directly chemically bonded atoms separated by distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$. In particular, the fourth term describes the van der Waals repulsive and attractive interatomic interactions by using a Lennard-Jones 12-6 potential. The fifth term describes instead the Coulomb interactions, with q_i and q_j representing effective atomic point charges and r_{ij} their distance. Effects due to specific environments can be taken into account using properly modified values for the partial charges as well as for the dielectric constant and the Van der Waals parameters ϵ_{ij} and σ_{ij} .

A schematic representation of the bonded and nonbonded interactions described in common force fields is given in Fig. 2.1.

Force fields used for simulations of biological systems can be classified in "all atom" or "united atom". In "all atom" force fields all atoms are explicitly treated while in "united atom" force fields only heavy (non-hydrogen) atoms and polar hydrogens are treated explicitly. Nonpolar hydrogen atoms are instead included into the description of the carbon atoms to which they are bound to by using an augmented Van der Waals radius [38]. Among all atom force fields we can list CHARMM [39, 40], AMBER [41, 42] and OPLS [43, 44]. An example of a united atom force field is instead GROMOS [45, 46]. Despite the continuous advances in the accuracy

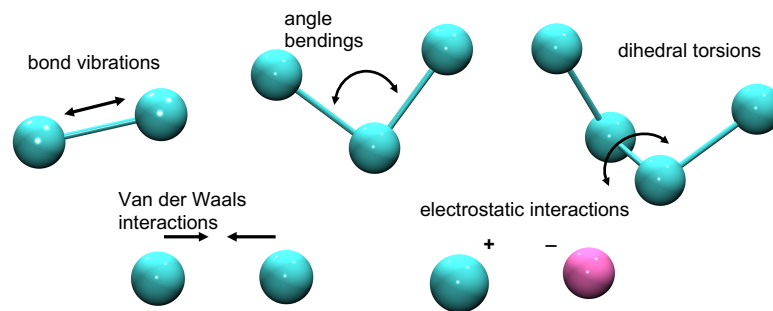


Figure 2.1: Schematic representation of the bonded and nonbonded interaction terms comprised in molecular mechanics force fields. The picture has been obtained using VMD. [36, 37].

of force fields one of the major pitfalls is represented by the failure in the description of polarization effects since atoms described by fixed effective point charges. For this reason, recently a huge effort has been devoted to the development of polarizable models such as AMOEBA [47–49].

The force acting upon the i th particle (i.e. atom) of the system at time t is given by the negative gradient (i.e. first derivatives vector) of the potential $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$ with respect to the particle position at that time $\mathbf{r}_i(t) = (x_i(t), y_i(t), z_i(t))$, as shown in eq. 2.22:

$$\mathbf{f}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \dots, \mathbf{r}_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i} \right) \quad (2.22)$$

At the same time, \mathbf{F}_i is given by Newton's equation of motion shown in eq. 2.23:

$$\mathbf{f}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2} \quad (2.23)$$

with m_i being the i th particle mass.

In MD simulations, the time evolution of the particles composing the system is followed by solving Newton's equations of motion in eq. 2.23. Such second order differential equations are commonly integrated and solved numerically by finding expressions able to define the positions $\mathbf{r}_i(t + \Delta t)$ at time $(t + \Delta t)$ in terms of the already known positions at time t . In particular, due to its simple and stable form, the Verlet algorithm [50] is heavily used in MD simulation software. The basic formula for this algorithm can be derived from the Taylor expansion of the particle position around time t . Once Δt is defined as a sufficiently small integration time step, $\mathbf{r}_i(t + \Delta t)$ can be defined as:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\mathbf{f}_i(t)}{2m_i}\Delta t^2 + \frac{\Delta t^3}{3!}\ddot{\mathbf{r}}_i + \mathcal{O}(\Delta t^4) \dots \quad (2.24)$$

and similarly, $\mathbf{r}_i(t - \Delta t)$ can be defined as:

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \mathbf{v}_i(t)\Delta t + \frac{\mathbf{f}_i(t)}{2m_i}\Delta t^2 - \frac{\Delta t^3}{3!}\ddot{\mathbf{r}}_i + \mathcal{O}(\Delta t^4) \dots \quad (2.25)$$

By adding eq. 2.24 and eq. 2.25 we obtain:

$$\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = 2\mathbf{r}_i(t) + \frac{\mathbf{f}_i(t)}{m_i}\Delta t^2 + \mathcal{O}(\Delta t^4) \quad (2.26)$$

$$\mathbf{r}_i(t + \Delta t) \approx 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{f}_i(t)}{m_i}\Delta t^2. \quad (2.27)$$

Eq. 2.27 is the Verlet algorithm to compute the new position of each particle of the system at time $(t + \Delta t)$, with Δt being the integration time step [50]. This expression is accurate up to Δt^4 and does not use the velocity to evaluate new positions. Velocities can be derived from the positions, as shown in eq. 2.28 and 2.29:

$$\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t) = 2\mathbf{v}_i(t)\Delta t + \mathcal{O}(\Delta t^2) \quad (2.28)$$

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2). \quad (2.29)$$

The expression in eq. 2.29 is only accurate up to the Δt^2 . However, it is possible to compute velocities, and therefore kinetic and total energy, directly with alternative schemes like the Velocity-Verlet [51] or the leapfrog algorithms [52].

By directly integrating Newton's equations of motion a simulation in the NVE (microcanonical) ensemble is produced, in which the number of particles N , the volume V and the total energy E are kept constant. However, experiments are often conducted at constant pressure and/or constant temperature. For this reason, several thermostat and barostat algorithms have been

developed in order to perform MD simulations in the canonical NVT ensemble (in which the total energy is allowed to change in order to maintain volume and temperature constant) or in the isothermal-isobaric NPT ensemble (in which temperature and pressure are simultaneously kept constant, thus allowing for changes in volume). Examples of thermostats currently used in MD simulations are the Nosé-Hoover thermostat, the Berendsen weak-coupling scheme and its improved version the Parrinello-Bussi thermostats, Langevin dynamics, and Andersen thermostat. Exhaustive details regarding the theoretical background and the implementation of the mentioned thermostats are given in references [53–58], [59–61], [62], [63] respectively. Among the barostats used in MD simulations, it is possible to list instead the Andersen barostat, the Nosé-Hoover one (which is an extension of Andersen's approach) and the Parrinello-Rahman barostat. Once again the reader can refer to the following literature for further details [63], [64], [65, 66].

When setting up an MD simulation it is necessary to choose suitable boundary conditions. If these are not specified (open system) particles can indeed freely move without limits in space, thus making the simulation unfeasible. On the other side, if the computational system is closed inside a limited size box (closed system) a substantial portion of particles will be placed close to the box surfaces, thus leading to unphysical boundary effects on the dynamical behaviour of particles. The most commonly used strategy to overcome such artifacts is to impose periodic boundary conditions (PBC). When PBC are applied, the actual simulation box (central box) is replicated infinitely in order to form an infinite periodic lattice filling the whole space, as shown in Fig. 2.2. Periodic images of particles move in each of the replica boxes exactly as the original particle moves in the central one. In this way, when a particle leaves the central box from one of its faces one of its replica will enter the central box from the opposite face, thus ensuring that the number of particles of the system remains constant all over the simulation. It is noteworthy to underline that the box surfaces are only fictitious and, therefore, no particle lays on a surface [67–69]. Because of their simple geometries cubic and rectangular boxes are the most widely used. However other space-filling geometries that can reduce the system size can be used. As an example, rhombic dodecahedron or truncated octahedron boxes are more suitable to be used in the simulation of nearly spherical systems, since the number of solvent molecules required to fill the box is lower [34].

The very high frequencies of bond vibrations, in particular the ones involving light hydrogen atoms, would require the integration time step to be extremely short. However, since these bond vibrations are usually not particularly interesting when exploring configurational space, molecules are typically treated as rigid or semirigid bodies characterized by fixed bond lengths (and also fixed bond angles sometimes). For this reason several constraint algorithms have been developed to remove fast degrees of freedom and resulting in remarkable speed-ups of all-atom classical MD simulations. One of the most famous algorithm to impose holonomic constraints is the SHAKE algorithm [70]. It iteratively adjusts the atomic coordinates until each of the imposed constraints is satisfied within a given tolerance threshold. RATTLE is a revised version of SHAKE designed for the velocity Verlet algorithm [71]. SETTLE is a constraint algorithm for imposing rigid water where constraints are solved analytically [72]. Another

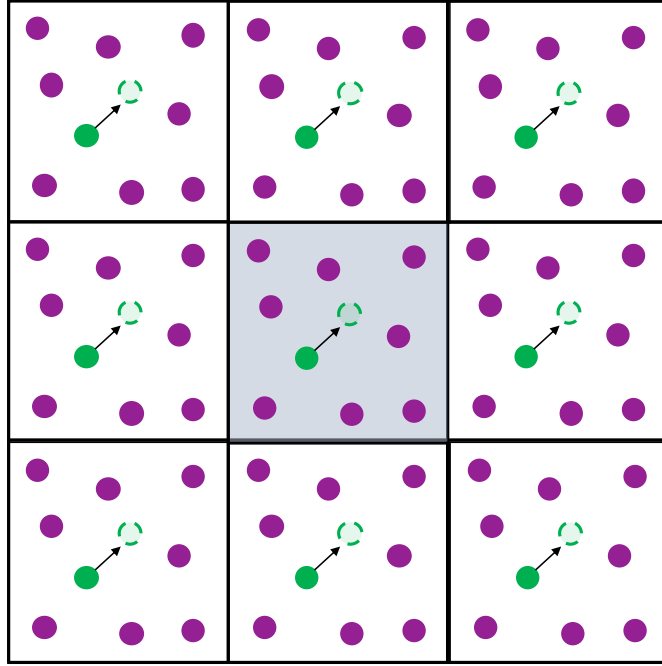


Figure 2.2: Two-dimensional schematic illustration of the Periodic Boundary Condition (PBC) scheme. The purple and green spheres represent the static and moving particles, respectively. The green dashed sphere represents the arrival point of the green sphere and the black arrow represents the movement direction. As it is possible to see, the same system in the central gray-shaded box is replicated in each of the mirror boxes.

constraints algorithm used in biomolecular simulation is LINCS (LINEar Constraint Solver). Its major feature is that it is non-iterative and bonds are reset to their imposed values after each unconstrained propagation step without the need of calculating the actual constraint forces making the algorithm more stable and faster than SHAKE. However, its use is limited to bond constraints and isolated angle constraints [73–75].

2.3 *Ab Initio* Molecular Dynamics (AIMD)

2.3.1 From the full quantum problem to AIMD

A complete description of physical phenomena implies that the quantum nature of both electrons and nuclei has to be taken into account. The dynamic evolution of a quantum system composed of M nuclei and N electrons, with the nuclear and electronic degrees of freedom represented as $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_M)$ and $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N)$, respectively, is described by the time-dependent Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \Phi(\mathbf{r}, \mathbf{R}, t) = \mathcal{H} \Phi(\mathbf{r}, \mathbf{R}, t) \quad (2.30)$$

where $\Phi(\mathbf{r}, \mathbf{R}, t)$ is a many-body wavefunction depending on $3(M + N)$ geometric coordinates and time:

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \Phi(\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N\}, \{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_M\}, t) \quad (2.31)$$

and \mathcal{H} represents the electronic-nuclear Hamiltonian, given by:

$$\begin{aligned} \mathcal{H} = & - \sum_{I=1}^M \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 + \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ & + \frac{1}{4\pi\epsilon_0} \sum_{I=1}^{M-1} \sum_{J=I+1}^M \frac{e^2 Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \frac{1}{4\pi\epsilon_0} \sum_{I=1}^M \sum_{i=1}^N \frac{e Z_I}{|\mathbf{R}_I - \mathbf{r}_i|} \end{aligned} \quad (2.32)$$

where m_e and $-e$ are the mass and charge of the electron and M_I and Z_I are the mass and charge of the I -th nucleus while ϵ_0 is the vacuum permittivity. In the Hamiltonian in eq. 2.32 the first two terms represent the kinetic energy of nuclei and electrons while the third, fourth and fifth term represent the potential energy associated with electron-electron, nucleus-nucleus and electron-nucleus interactions, respectively.

If we consider an infinitesimally small instant of time in which the nuclei are frozen (clamped nuclei approximation), the time-independent electronic Schrödinger equation can be solved since the wavefunction results to depend only on electronic degrees of freedom and only parametrically on the nuclear degrees of freedom:

$$\mathcal{H}_e(\mathbf{r}; \mathbf{R}) \psi_I(\mathbf{r}; \mathbf{R}) = E_I(\mathbf{R}) \psi_I(\mathbf{r}; \mathbf{R}) \quad (2.33)$$

The repetition of the same procedure for each instant in time and for each corresponding set of nuclear coordinates allows for the expansion of the time-dependent coupled electronic-nuclear wavefunction in eq. 2.31 in a complete set of electronic wavefunctions $\{\psi_I\}$ in which the time-dependent nuclear wavefunctions are included in the form of expansion coefficients [76, 77]:

$$\Phi(\mathbf{r}, \mathbf{R}, t) = \sum_{I=0}^{\infty} \psi_I(\mathbf{r}; \mathbf{R}) \chi_I(\mathbf{R}, t) \quad (2.34)$$

By inserting eq. 2.34 into eq. 2.30, we obtain:

$$i\hbar \frac{\partial}{\partial t} \sum_{l=0}^{\infty} \psi_l(\mathbf{r}; \mathbf{R}) \chi_l(\mathbf{R}, t) = \mathcal{H} \sum_{l=0}^{\infty} \psi_l(\mathbf{r}; \mathbf{R}) \chi_l(\mathbf{R}, t) \quad (2.35)$$

The multiplication of eq. 2.35 from the left by $\psi_k^*(\mathbf{r}; \mathbf{R})$ and integration over the electronic coordinates \mathbf{r} leads to a set of coupled differential equations for the time-dependent nuclear wavefunctions $\chi_k(\mathbf{R}, t)$:

$$\left[-\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + E_k(\mathbf{R}) \right] \chi_k + \sum_l C_{kl} \chi_l = i\hbar \frac{\partial}{\partial t} \chi_k \quad (2.36)$$

where:

$$C_{kl} = \int \psi_k^* \left[-\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 \right] \psi_l d\mathbf{r} + \frac{1}{M_I} \sum_I \left\{ \int \psi_k^* [-i\hbar \nabla_I] \psi_l d\mathbf{r} \right\} [-i\hbar \nabla_I] \quad (2.37)$$

is the nonadiabatic coupling operator whose first term represents a matrix element of the nuclear kinetic energy operator while the second term depends on the nuclear momenta. The so called "adiabatic approximation" to the nonadiabatic problem presented in eq. 2.36 consists in considering only the diagonal terms of the coupling operator since they depend only on a single adiabatic wavefunction ψ_k :

$$C_{kk} = -\sum_I \frac{\hbar^2}{2M_I} \int \psi_k^* \nabla_I^2 \psi_k d\mathbf{r} \quad (2.38)$$

while the second term of eq. 2.37 is equal to zero when the electronic wavefunction is real. This leads to a complete decoupling of the set of coupled differential equations in eq. 2.36, as shown below:

$$\left[-\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + E_k(\mathbf{R}) + C_{kk}(\mathbf{R}) \right] \chi_k = i\hbar \frac{\partial}{\partial t} \chi_k \quad (2.39)$$

The above eq. 2.39 indicates that the dynamical evolution of the nuclei proceeds without a change in the electronic quantum state (k-th in this case) and, thus, the coupled electronic-nuclear wavefunction (eq. 2.31) can be decoupled as the direct product of a nuclear and an

electronic wavefunction:

$$\Phi(\mathbf{r}, \mathbf{R}, t) \approx \psi_k(\mathbf{r}; \mathbf{R}) \chi_k(\mathbf{R}, t) \quad (2.40)$$

It is important also to notice that this means considering a single term of the expansion defined in eq. 2.34.

The last simplification of the problem consists in further neglecting the diagonal terms of the nonadiabatic coupling matrix. This is justified by the fact that these terms (eq. 2.38) represent just a constant shift to the adiabatic eigenvalues of the electronic Schrödinger equation for a given state. From this approximation we obtain:

$$\left[-\sum_I \frac{\hbar^2}{2M_I} \nabla_I^2 + E_K(\mathbf{R}) \right] \chi_k = i\hbar \frac{\partial}{\partial t} \chi_k \quad (2.41)$$

which represents the exact definition of the already mentioned "Born-Oppenheimer approximation" [78]. Assuming the Born-Oppenheimer approximation (which is the case for a wide range of physical systems), is the first step towards a treatment of the nuclear dynamics in a framework of classical point particles. In the following, the motivation for this will be illustrated.

We can start by rewriting the nuclear wavefunction in terms of the amplitude factor A_k and the phase S_k , both of them real and with $A_k > 0$ [79–81]:

$$\chi_k(\mathbf{R}, t) = A_k(\mathbf{R}, t) \exp\left(\frac{iS_k(\mathbf{R}, t)}{\hbar}\right) \quad (2.42)$$

By inserting eq. 2.42 into eq. 2.41 and separating real and imaginary parts, we obtain the following two coupled equations:

$$\frac{\partial S_k}{\partial t} + \sum_I \frac{1}{2M_I} (\nabla_I S_k)^2 + E_k = \hbar^2 \sum_I \frac{1}{2M_I} \frac{\nabla_I^2 A_k}{A_k} \quad (2.43)$$

$$\frac{\partial A_k}{\partial t} + \sum_I \frac{1}{M_I} (\nabla_I A_k)(\nabla_I S_k) + \sum_I \frac{1}{2M_I} A_k (\nabla_I^2 S_k) = 0 \quad (2.44)$$

By multiplying the amplitude equation (eq. 2.44) from the left by $2A_k$ we can rewrite it as follows:

$$\frac{\partial A_k^2}{\partial t} + \sum_I \frac{1}{M_I} \nabla_I (A_k^2 \nabla_I S_k) = 0 \quad (2.45)$$

or:

$$\frac{\partial \rho_k}{\partial t} + \sum_I \nabla_I \mathbf{J}_{k,I} = 0 \quad (2.46)$$

where $\rho_k = |\chi_k|^2 \equiv A_k^2$ is the nuclear probability density and $\mathbf{J}_{k,I} = A_k^2 (\nabla_I S_k) / M_I$ is the related current density. This equation is independent of \hbar and ensures the conservation of the nuclear probability density $|\chi_k|^2$ in the presence of a flux.

Some considerations need to be made also regarding the phase equation (eq. 2.43). If we apply the classical limit $\hbar \rightarrow 0$, the term of the equation directly depending on \hbar goes to zero and we obtain:

$$\frac{\partial S_k}{\partial t} + \sum_I \frac{1}{2M_I} (\nabla_I S_k)^2 + E_k = 0 \quad (2.47)$$

The above equation is isomorphous to the Hamilton-Jacobi formulation of classical mechanics [82] for a given conserved value of the total energy $dE_k^{tot}/dt = 0$:

$$\frac{\partial S_k}{\partial t} + H_k(\mathbf{R}, \nabla_I S_k) = 0 \quad (2.48)$$

where the classical Hamilton function is given by:

$$H_k(\mathbf{R}, \mathbf{P}) = T(\mathbf{P}) + V_k(\mathbf{R}) \quad (2.49)$$

Hence:

$$\frac{\partial S_k}{\partial t} = -(T + E_k) = -E_k^{tot} = constant \quad (2.50)$$

Now, by considering that:

$$\mathbf{P} \equiv \nabla_I S_k \quad (2.51)$$

the equations of motion can be rewritten in their Newtonian formulation:

$$\dot{\mathbf{P}}_I = -\nabla_I V_k(\mathbf{R}) \quad (2.52)$$

can be written as

$$\frac{d\mathbf{P}_I}{dt} = -\nabla_I E_k \quad (2.53)$$

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I V_k^{BO}(\mathbf{R}(t)) \quad (2.54)$$

2.3.2 Born-Oppenheimer Molecular Dynamics

As just shown above, in the classical limit $\hbar \rightarrow 0$, the nuclei can be treated as point particles that move according to classical mechanics over an effective potential V_k^{BO} given by the Born-Oppenheimer potential energy surface $E_k(\mathbf{R})$ for the k-th electronic state for every set of nuclear coordinates \mathbf{R} , as described in eq. 2.54. An efficient approach to simulate this dynamics consists in calculating the potential energy surface $E_k(\mathbf{R})$ *on-the-fly* by solving the time-independent Schrödinger equation (TISE) for each nuclear configuration $\mathbf{R}(t)$ visited during the dynamics. The TISE is firstly solved for an initial set of nuclear coordinates $\mathbf{R}(t=0)$ and the forces acting on the nuclei are calculated and used to propagate the system to a new configuration $\mathbf{R}(t + \Delta t)$ according to classical mechanics. The iterative repetition of this procedure over time leads to what is commonly called Born-Oppenheimer Molecular Dynamics (BOMD) [83]. Born-Oppenheimer MD can be defined as:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \min_{\psi_0} \{ \langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle \} \quad (2.55)$$

$$E_0 \psi_0 = \mathcal{H}_e \psi_0 \quad (2.56)$$

for the electronic ground state. In order to solve the nuclear equations of motion a minimum for the electronic ground state has to be evaluated by diagonalising the Hamiltonian in eq. 2.56 at each time propagation step. It is important to notice the coupled dependence between electrons and nuclei. At each MD step the nuclear propagation depends indeed on the electronic state determining the potential energy to which they are subject and, at the same time, the electronic state parametrically depends on the nuclear configuration.

Such a Born-Oppenheimer MD scheme was initially performed using semiempirical approximations to the electronic structure problem [84, 85]. Only few years later the *ab initio* Hartree-Fock (HF) theory was implemented within BOMD [86]. In the HF method a single Slater determinant $\psi_0 = \frac{1}{\sqrt{N!}} \det\{\phi_i\}$ is used to represent the electronic wavefunction with the condition that the one-electron wavefunctions (i.e. the orbitals) ϕ_i are orthonormal: $\langle \phi_i | \phi_j \rangle = \delta_{ij}$ [87]. The variational minimum of the expectation value of the total energy with respect to the orbitals (HF approximation) is:

$$\min_{\{\phi_i\}} \{ \langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle \} \Big|_{\{ \langle \phi_i | \phi_j \rangle = \delta_{ij} \}} \quad (2.57)$$

Eq. 2.57 can be rewritten with the Lagrangian formalism:

$$\mathcal{L} = -\langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle + \sum_{i,j} \Lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \quad (2.58)$$

where the associated Lagrange multipliers Λ_{ij} are needed to impose the orthonormality condition. From the Hartree-Fock equations:

$$H_e^{HF} \phi_i = \sum_j \Lambda_{ij} \phi_j \quad (2.59)$$

we obtain for the equations of motion in the HF-BOMD scheme:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \min_{\{\phi_i\}} \{\langle \psi_0 | \mathcal{H}_e^{HF} | \psi_0 \rangle\} \quad (2.60)$$

$$0 = -\mathcal{H}_e^{HF} \phi_i + \sum_j \Lambda_{ij} \phi_j \quad (2.61)$$

However, towards the late 1980s and the early 1990s the combination of BOMD with DFT [5, 88] started to gain popularity and it is currently the most widely used among the BOMD schemes [89]. A set of equations similar to eq. 2.60 and 2.61 for the BOMD-DFT scheme is available:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \min_{\{\phi_i\}} \{\langle \psi_0 | \mathcal{H}_e^{KS} | \psi_0 \rangle\} \quad (2.62)$$

$$0 = -\mathcal{H}_e^{KS} \phi_i + \sum_j \Lambda_{ij} \phi_j \quad (2.63)$$

being \mathcal{H}_e^{KS} the Kohn-Sham effective one-particle hamiltonian. For more details regarding the Density Functional Theory (DFT) please refer to paragraph 2.1.

2.3.3 Car-Parrinello Molecular Dynamics

An alternative and elegant approach to *ab initio* MD, which represented a real breakthrough in the field of first-principles molecular dynamics simulations, was introduced in 1985 by Car and Parrinello [83, 90]. The basic idea behind the Car-Parrinello method was to take maximum advantage of the adiabatic time scale separation between the fast electronic and the slow nuclear motions in order to integrate the equations of motion on the long time scales set by nuclear dynamics and, simultaneously, profit by the smooth time evolution of the electronic degrees of freedom. They achieved this goal by transforming the adiabatic time scale separation in an adiabatic energy scale separation. The coupled quantum-classical problem was indeed converted into a purely classical problem with two separate energy scales at the expense of losing the physical time information of the quantum subsystem dynamics.

In Car-Parrinello Molecular Dynamics (CPMD), electrons are included as active degrees of freedom by performing a fictitious classical dynamics to propagate them along the nuclear

motion. The method is based on the extended Lagrangian \mathcal{L}_{CP} describing the temporal evolution of a fictitious classical system in which electronic degrees of freedom as well as the nuclear ones are treated as dynamic variables:

$$\begin{aligned} \mathcal{L}_{CP} = & T_N + T_e - E_{pot} + constraints \\ = & \underbrace{\sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2}_{T_N} + \underbrace{\sum_i \frac{1}{2} \mu |\dot{\phi}_i|^2}_{T_e} - \underbrace{\langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle}_{E_{pot}} + \underbrace{\sum_{ij} \Lambda_{ij} \left[\left\{ \int \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} \right\} - \delta_{ij} \right]}_{constraints} \end{aligned} \quad (2.64)$$

where T_N and T_e represent the kinetic energy terms for the nuclear and electronic degrees of freedom respectively, E_{pot} is the potential energy, the Lagrange multipliers Λ_{ij} ensure the orthonormality of one-electron orbitals ϕ_i . μ is the fictitious mass (also referred to as "inertia parameter") assigned to the one-electron orbitals representing the electrons. By using the Euler-Lagrange equations it is possible to derive the Newtonian equations of motions:

$$\frac{d}{dt} \left[\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{R}}_I} \right] = \frac{\partial \mathcal{L}}{\partial \mathbf{R}_I} \quad (2.65)$$

$$\frac{d}{dt} \left[\frac{\delta \mathcal{L}}{\delta \dot{\phi}_i^*} \right] = \frac{\delta \mathcal{L}}{\delta \phi_i^*} \quad (2.66)$$

In eq. 2.66 the derivation of the Lagrangian has to be done with respect to the one-electron orbitals $\phi_i(\mathbf{r})$, representing the electronic coordinates.

From eq. 2.65 and 2.66 we obtain the Car-Parrinello equations of motion:

$$M_I \ddot{\mathbf{R}}_I(t) = - \frac{\partial}{\partial \mathbf{R}_I} \langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle + \frac{\partial}{\partial \mathbf{R}_I} \{constraints\} \quad (2.67)$$

$$\mu \ddot{\phi}_i(t) = - \frac{\delta}{\delta \phi_i^*} \langle \psi_0 | \mathcal{H}_e | \psi_0 \rangle + \frac{\delta}{\delta \phi_i^*} \{constraints\} \quad (2.68)$$

When one-particle Hamiltonians, such as those from DFT theory in the Kohn-SHam formulation are applied to the CPMD scheme together with position-independent constraints, the

\mathcal{L}_{CP} becomes:

$$\mathcal{L}_{CP} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \mu \langle \dot{\phi}_i | \dot{\phi}_i \rangle - \langle \psi_0 | \mathcal{H}_e^{KS} | \psi_0 \rangle + \sum_{i,j} \Lambda_{ij} (\langle \phi_i | \phi_j \rangle - \delta_{ij}) \quad (2.69)$$

where the orthonormality of KS one-electron orbitals $\langle \phi_i | \phi_j \rangle = \delta_{ij}$ is imposed by Lagrangian multipliers Λ_{ij} . By inserting this Lagrangian into eq. 2.65 and 2.66 we obtain the corresponding Car-Parrinello equations of motions:

$$M_I \ddot{\mathbf{R}}_I(t) = -\nabla_I \langle \psi_0 | \mathcal{H}_e^{KS} | \psi_0 \rangle \quad (2.70)$$

$$\mu \ddot{\phi}_i(t) = -\mathcal{H}_e^{KS} \phi_i + \sum_j \Lambda_{ij} \phi_j \quad (2.71)$$

2.4 Hybrid Quantum Mechanics/Molecular Mechanics Simulations

The size and complexity of typical biological systems as well as the long time scales that need to be spanned require the use of classical molecular dynamics. However, as already mentioned above, electrons are not explicitly treated in molecular mechanics force fields and their effects are only included by empirical parameters. As a consequence, all physical processes involving electronic rearrangements, such as chemical reactions with bonds breaking and forming or photochemical processes, or systems in which electronic effects play a major role, such as those involving (transition)metal ion complexes, are not properly described by classical MD. These kind of processes and systems would rather require a quantum mechanics description able to take the electronic degrees of freedom explicitly into account.

In order to overcome the limitations imposed by both full quantum mechanics (QM) and molecular mechanics (MM) treatments, multiscale hybrid approaches have been developed in which only a small portion of the system is treated at the QM level while the rest of the system is treated at a lower level method, usually a molecular mechanics method based on empirical force fields (MM) [91–96]. This mixed Quantum Mechanical/Molecular Mechanical (QM/MM) strategy was introduced by Warshel and Levitt [97] (developing former ideas by Warshel and Karplus [98] and Birge et al. [99]) with the publication of a seminal paper in which a semiempirical QM/MM approach was used for the first time to describe the enzymatic reaction of lysozyme. The partitioning of the system into regions treated at different levels of theory is based on the fact that electronic effects have, in most cases, a local character [100]. As an example, reactions in solution involve in most cases only the reactants and

the first few solvation shells while the rest of the bulk solvent is not affected by the reaction (even if it can influence it through long-range interactions). By using a QM/MM approach the computational resources can be concentrated on the portion of the system in which the electronic rearrangements take place (i.e. the "reaction centre") while the rest of the system that does not explicitly participate to the reaction is treated at a computationally cheaper level.

The QM/MM total energy contains three interaction terms: (i) interactions in the QM region, (ii) interaction in the MM region and (iii) interactions between QM and MM regions. While QM-QM and MM-MM interactions are quite straightforward to describe, that is at the QM and MM levels, respectively, interactions between the two subsystem (QM-MM) are more difficult to treat. Most of the current QM/MM methods use either a subtractive or an additive coupling scheme.

In the subtractive scheme the QM energy is calculated for the isolated QM subsystem whereas the energy of the MM subsystem is evaluated at the MM level of theory as the difference between the MM energy of the entire system (QM+MM) and the MM energy of the QM subsystem. The total energy of the embedded system is therefore:

$$E_{QM/MM} = E^{QM}(QM) + E^{MM}(QM + MM) - E^{MM}(QM) \quad (2.72)$$

where E^{QM} and E^{MM} represent the total energies evaluated at the quantum mechanics and molecular mechanics levels of theory, respectively. It is important to highlight that in eq. 2.72 it is assumed that the change in energy from the MM to the QM level of theory is fully transferable. The forces acting on the nuclei are derived by direct differentiation with respect to the nuclear coordinates:

$$\mathbf{F}_I = -\frac{\partial E_{QM/MM}}{\partial \mathbf{R}_I} = -\frac{\partial E^{QM}(QM)}{\partial \mathbf{R}_I} - \frac{\partial E^{MM}(QM + MM)}{\partial \mathbf{R}_I} + \frac{\partial E^{MM}(QM)}{\partial \mathbf{R}_I} \quad (2.73)$$

where \mathbf{F}_I is the force acting on the i-th nucleus at position \mathbf{R}_I .

In a standard subtractive scheme the so called "mechanical coupling" between the QM and MM regions is typically used [101]. In this scheme the electrostatic interactions between QM and MM atoms are described fully at the lower (MM) level as Coulomb interactions between classical point charges and, therefore, the electrons of the QM subsystem do not feel the electrostatic field generated by the MM environment. This represents a good approximation only if the difference between the lower and higher level calculations is sufficiently small. One of the main advantages of the QM/MM subtractive scheme is that the implementation results to be quite simple since no communication is required between the QM and MM routines. Furthermore, it is straightforward extendable to multiple regions treated with as many levels

of theory as in the ONIOM approach introduced by Morokuma and colleagues [102, 103]. On the other hand, as already mentioned, the absence of polarization of the QM electron density by the MM environment represents a major drawback. This is the case of enzymatic reactions in which the active site is treated at the QM level and the rest of the protein at the MM level but the electrostatic background created by the latter is however capable to positively or negatively affect the reaction. A realistic description of such kind of processes requires a more advanced treatment of the interactions between QM atoms and the surrounding environment.

In the additive scheme, at the basis of most of the QM/MM approaches currently used, the system is described by a total Hamiltonian:

$$\mathcal{H} = \mathcal{H}_{QM} + \mathcal{H}_{MM} + \mathcal{H}_{QM/MM} \quad (2.74)$$

where H_{QM} and H_{MM} are the Hamiltonians describing the QM and MM regions respectively, while $H_{QM/MM}$ is the Hamiltonian describing the coupling between the two regions. As a consequence, the total (ground state) energy of the hybrid QM/MM system is the lowest eigenvalue of the Hamiltonian in eq. 2.74:

$$E = E_{QM} + E_{MM} + E_{QM/MM} \quad (2.75)$$

and results to be just the sum of the QM, MM and QM/MM coupling energy terms. In contrast to the subtractive scheme, here the QM calculation is performed in presence of the MM environment. The QM/MM electrostatic coupling can be described at different levels of sophistication.

In the most basic approach, referred to as "mechanical embedding", similarly to what has already been discussed for the subtractive scheme, the electrostatic interactions between QM and MM atoms are treated as classical point charge Coulombic interactions [101]. The electronic wavefunction is evaluated for the isolated QM subsystem and, therefore, the MM environment is not able to induce polarization of the electron density. An improvement is represented by the so called "electrostatic embedding" scheme, in which QM atoms feel polarization effects induced by the MM environment [101]. Here, the electrostatic interactions between the two subsystem are indeed handled at the level of the QM wavefunction computation by including the MM point charges into the QM Hamiltonian as an additional part of the external potential. A further improvement in the treatment of the QM/MM electrostatic coupling consists in the use of a polarizable MM environment and the two subsystem are able to mutually polarize each other ("polarizable embedding") [101]. Most of the additive schemes currently used are based on the "electrostatic embedding" scheme.

The main advantage of the additive formalism is the high flexibility in the use of different levels of theory for the description of the QM, MM and specially the QM/MM interface regions. The main drawback consists instead in the creation of an abrupt boundary between the QM and MM regions. This leads to a discontinuity in the description of the electrons ceasing to exist when crossing the boundaries that need to be dealt with in an appropriate way.

2.4.1 Theoretical background of QM/MM

In the following, the theoretical framework of an additive QM/MM scheme where the QM region is described at the DFT level is illustrated [104]. We start by considering the whole system (QM+MM) as described at the QM DFT level. The first approximation consists in dividing the total system in two parts, A and B, represented by the densities ρ_A and ρ_B , respectively, such that the total density ρ is the sum of the two:

$$\rho(\mathbf{r}) = \rho_A(\mathbf{r}) + \rho_B(\mathbf{r}) \quad (2.76)$$

Similarly to eq. 2.75 the total energy of the system is given by:

$$E = E_A + E_B + E_{A-B} \quad (2.77)$$

or, using the energy functional in eq.(2.6):

$$\begin{aligned} E = & T[\rho_A] + T[\rho_B] + T^{NL} + \int_{\Omega} V^{ex}(\mathbf{r})\rho_A(\mathbf{r})d\mathbf{r} + \int_{\Omega} V^{ex}(\mathbf{r})\rho_B(\mathbf{r})d\mathbf{r} \\ & + \frac{1}{2} \iint \frac{\rho_A(\mathbf{r})\rho_A(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{1}{2} \iint \frac{\rho_B(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ & + \frac{1}{2} \iint \frac{\rho_A(\mathbf{r})\rho_B(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[\rho_A] + E_{xc}[\rho_B] + E_{xc}^{NL} + \\ & + \frac{1}{2} \sum_I \sum_J \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \quad (2.78)$$

where the terms T^{NL} and E_{xc}^{NL} account for the nonlinearity of kinetic energy and exchange-correlation functionals:

$$T^{NL} = T[\rho_A + \rho_B] - (T[\rho_A] + T[\rho_B]) \quad (2.79)$$

$$E_{xc}^{NL} = E_{xc}[\rho_A + \rho_B] - (E_{xc}[\rho_A] + E_{xc}[\rho_B]) \quad (2.80)$$

These terms become negligible (or zero) only when ρ_A and ρ_B are spatially far from each other and non overlapping.

Since we want to describe A and B at different levels of theory, it is convenient to separate the external potential V^{ex} as the sum of two contributions coming from the nuclei located on the A and B parts, namely V_A^{ex} and V_B^{ex} respectively:

$$V^{ex}(\mathbf{r}) = V_A^{ex}(\mathbf{r}) + V_B^{ex}(\mathbf{r}) \quad (2.81)$$

The terms E_A and E_B of eq. 2.77 can be rewritten as:

$$\begin{aligned} E_\kappa = & T[\rho_\kappa] + \int_\Omega V_\kappa^{ex}(\mathbf{r})\rho_\kappa(\mathbf{r})d\mathbf{r} + \frac{1}{2} \iint \frac{\rho_\kappa(\mathbf{r})\rho_\kappa(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \\ & + E_{xc}[\rho_\kappa] + \frac{1}{2} \sum_{I \in \kappa} \sum_{J \in \kappa} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \end{aligned} \quad (2.82)$$

where $\kappa = A, B$.

By expanding the nuclear charges into Gaussian charge distributions of the form:

$$Z_I = \int \rho_I^{nuclear}(\mathbf{r} - \mathbf{R}_I) d\mathbf{r} = \int_\Omega \frac{Z_I}{R_c^3} \pi^{-2/3} \exp \left[-\frac{|\mathbf{r} - \mathbf{R}_I|^2}{R_c^2} \right] \quad (2.83)$$

where R_c describes the width of the Gaussian, the three electrostatic terms of eq. 2.82 can be summarized into one term depending on the total (electronic and nuclear) charge density $\rho^{el+nuc} = \rho^{el} + \rho^{nuc}$:

$$\begin{aligned} & \int_\Omega V_\kappa^{ex}(\mathbf{r})\rho_\kappa(\mathbf{r})d\mathbf{r} + \frac{1}{2} \iint \frac{\rho_\kappa(\mathbf{r})\rho_\kappa(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{1}{2} \sum_{I \in \kappa} \sum_{J \in \kappa} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \\ & = \frac{1}{2} \iint \frac{\rho_\kappa^{el+nuc}(\mathbf{r})\rho_\kappa^{el+nuc}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' \end{aligned} \quad (2.84)$$

The term E_{A-B} of eq. 2.77, accounting for the interactions between the two regions A and B, is described by the remaining terms of eq. 2.78:

$$\begin{aligned}
 E_{A-B} = & T^{NL} + \int_{\Omega} v_B^{ex}(\mathbf{r}) \rho_A(\mathbf{r}) d\mathbf{r} + \int_{\Omega} v_A^{ex}(\mathbf{r}) \rho_B(\mathbf{r}) d\mathbf{r} \\
 & + \frac{1}{2} \iint \frac{\rho_A(\mathbf{r}) \rho_B(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}^{NL} \\
 & + \frac{1}{2} \sum_{I \in A} \sum_{J \in B} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}
 \end{aligned} \tag{2.85}$$

We are now considering the particular case in which A is treated at the QM level while B at the MM level, so that in:

$$E_{\kappa} = T[\rho_{\kappa}] + \frac{1}{2} \iint \frac{\rho_{\kappa}^{el+nuc}(\mathbf{r}) \rho_{\kappa}^{el+nuc}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + E_{xc}[\rho_{\kappa}] \tag{2.86}$$

if $\kappa=A$, then $E_{\kappa} = E_A = E_{QM}$ while if $\kappa=B$, then $E_{\kappa} = E_B = E_{MM}$.

Despite the fact that the terms of eq. 2.86 do not directly match those contained in common force fields, the latter are designed in order to mimic most of the physical effects described in the equation and, therefore, several approximations can be done. Primarily, in force fields, electronic effects are described only implicitly. This corresponds, from an electronic point of view, to a frozen density decomposition, and, therefore, the kinetic energy term $T[\rho_{MM}]$ can be considered as a (composition-dependent) additive constant and not be directly taken into account. The exchange-correlation term $E_{xc}[\rho_{MM}]$ is approximated by a pair-additive van der Waals term, in the form of a 12-6 Lennard-Jones potential:

$$E_{xc}[\rho_{MM}] \approx E_{vdW} = \frac{1}{2} \sum_{I \in MM} \sum_{J \in MM} 4\epsilon_{IJ} \left(\left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^6 \right) \tag{2.87}$$

where $R_{IJ} = |\mathbf{R}_I - \mathbf{R}_J|$. The second term of eq. 2.86, describing the electrostatic interaction energy arising from the combined electronic and nuclear charge distribution is approximated in most commonly used force fields by effective fixed point charges:

$$\frac{1}{2} \iint \frac{\rho_{MM}^{el+nuc}(\mathbf{r}) \rho_{MM}^{el+nuc}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \approx \frac{1}{2} \sum_{I \in MM} \sum_{J \in MM} \frac{q_I q_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{2.88}$$

While the approximations made so far are valid for the treatment of non-bonded interactions between the atoms contained in the MM region, they are obviously not able to reproduce the highly directional and overlapping electron density distribution characterizing covalent chemical bonds. Most biomolecular force fields handle interactions between covalently bound atoms through explicit bond, angle and torsional terms, as in the equation below:

$$E_{MM}^{bonded} = \sum_d \frac{1}{2} k_d (d_{IJ} - d_0)^2 + \sum_\theta \frac{1}{2} k_\theta (\theta_{IJK} - \theta_0)^2 + \sum_\phi \sum_n k_n [1 + \cos(n\phi_{IJKL} + \phi_0)] \quad (2.89)$$

where the first term, describing bond vibrations, runs over all bonds d , with harmonic force constant k_d and equilibrium bond length d_0 ; the second term, describing angle bendings, runs over all bonding angles θ , with harmonic force constant k_θ and equilibrium angle θ_0 . The last term, describing dihedral torsions, includes all dihedral angles ϕ , with multiplicity n , force constants k_n and phases ϕ_0 . The total E_{MM} energy is therefore given by the sum of the three terms given in eq. 2.87, 2.88 and 2.89.

The interaction energy between QM and MM regions E_{QM-MM} is therefore given by:

$$E_{QM-MM} = \sum_I \int \frac{q_I}{|\mathbf{R}_I - \mathbf{r}|^2} \rho_{QM}^{el+nuc}(\mathbf{r}) d\mathbf{r} + \sum_I \sum_{I'} 4\epsilon_{I'I} \left(\left(\frac{\sigma_{I'I}}{R_{I'I}} \right)^{12} - \left(\frac{\sigma_{I'I}}{R_{I'I}} \right)^6 \right) + \sum_d \frac{1}{2} k_d (d_{I'I} - d_0)^2 + \sum_\theta \frac{1}{2} k_\theta (\theta_{I''J''K''} - \theta_0)^2 + \sum_\phi \sum_n k_n [1 + \cos(n\phi_{I''J''K''L''} + \phi_0)] \quad (2.90)$$

Here, I runs over all MM atoms and I' over all QM atoms, while (I'', J'', K'') and (I'', J'', K'', L'') represent the triple or quadruple bonded set of atoms in which at least one atom belongs to the QM region while the other to the MM region.

2.4.2 Pitfalls of the QM/MM method

QM/MM boundaries

Most times, especially when dealing with biological macromolecules, the QM/MM partitioning is defined in such a way that it intersects covalent bonds existing between QM and MM atoms, thus creating in the QM region one or more unpaired electrons. These are, in nature,

chemically reactive and would therefore cause severe artifacts in the QM/MM simulation. Several approaches have been developed in order to properly treat these "dangling bonds" and resolve the problem [91–96].

One method consists in introducing a monovalent link atom at a proper position along the bond vector in order to saturate the cut covalent bond. In principle, each link atom introduces three additional degrees of freedom to the system. However, in practice the link atom is placed at a fixed position so that these additional degrees of freedom are removed again. These "dummy atoms" are generally invisible to the MM atoms and interact via the force field only with the border QM atoms. However, in some cases they need to be taken into account also from the MM region, specially in case of non-negligible polarization effects. In most cases hydrogen-like link-atom are used to cap the QM subsystem, but fluorine atoms or even complete fragments, like methyl groups, are also possible. It is noteworthy that link-atoms need to be placed far from each other in order to avoid spurious interactions [105–107].

An alternative consists in replacing the broken chemical bonds by frontier orbitals ϕ_{FO} . In the localized self-consistent field (LSCF) method, the added atomic orbitals are localized at the QM atom of the broken bond. The hybrid orbital pointing towards the MM atom is occupied by two electrons while the other orbitals are each occupied by a single electron and, during the SCF optimization of the QM wave function, the double-occupied orbital is kept frozen, while the other hybrid orbitals are optimized together with all orbitals of the QM region [108]. In the alternative generalized hybrid orbital approach (GHO), the atomic hybrid orbitals are instead localized at the MM atom of the broken bond and, contrarily to the LSCF method, the orbital pointing towards the QM atom is optimized while the others are kept doubly-occupied and frozen [109].

Another way to deal with the problem is represented by the use of boundary pseudopotentials located at the first covalently bound MM atom at the QM/MM interface. The method is based on Effective Core Pseudopotentials (ECP), written as the sum of a local and a non-local term:

$$V_I^{ECP}(\mathbf{r}, \mathbf{r}') = V^{loc}(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}') + \sum_l V_l^{nl}(\mathbf{r}, \mathbf{r}') \quad (2.91)$$

$$V^{loc}(\mathbf{r}) = -\frac{Z_{ion}}{r} \operatorname{erf}\left[\frac{r}{r_{loc}\sqrt{2}}\right] + \exp\left(-\frac{r^2}{2r_{loc}^2}\right) \times \left[c_1 + c_2\left(\frac{r}{r_{loc}}\right)^2 + c_3\left(\frac{r}{r_{loc}}\right)^4 + c_4\left(\frac{r}{r_{loc}}\right)^6 \right] \quad (2.92)$$

$$V_l^{nl}(\mathbf{r}, \mathbf{r}') = \sum_{m=-l}^{+l} Y_{lm}(\hat{\mathbf{r}}) \sum_{j,h=1}^3 p_{lh}(r) h_{lhj} p_{lj}(\mathbf{r}') Y_{lm}^*(\hat{\mathbf{r}}') \quad (2.93)$$

where Y_{lm} are spherical harmonics, $p_{lh}(r) \propto r^{l+2(h-1)} \exp(-r^2/(2r_l^2))$, $r = |\mathbf{r} - \mathbf{R}_I|$ with \mathbf{R}_I being the position of the I th nucleus and $\hat{\mathbf{r}}$ is the unit vector in the direction of \mathbf{r} . The set of parameters $\{r_{loc}, c_1, c_2, c_3, c_4, h_{lhj}, r_l\}$ are optimized by iteratively minimizing the difference between the electron density ρ_{QM} of the QM subsystem and a reference density ρ_{ref} obtained performing a preliminary QM/MM simulation with a larger QM region. From the optimization process the set of parameters and, consequently, the Optimized Effective Core Pseudopotential (OECP) are derived [110].

Electron Spill-Out

The so called "electron spill-out" is an artifact that represents a common pitfall in QM/MM simulations, especially when fully delocalized basis sets (such as plane waves) are used. This phenomenon arises from the fact that, due to the absence of electrons in the MM region, the electrons of the QM region do not experience any Pauli repulsion effect from the MM atoms. This may lead to a localization of electron density on nearby positively charged classical point charges, that is effectively a "spill-out" of electron density into the MM region. This artificial effect can be avoided by using smeared-out charges instead of the traditional point charges. A convenient way for smearing the charges is to use a Gaussian distribution centered at the MM atom:

$$\Omega_I^{MM}(\mathbf{r}) = \sqrt{\frac{q_I^{MM}}{\pi\alpha^3}} \exp\left[-\frac{|\mathbf{r} - \mathbf{R}_I|^2}{2\alpha^2}\right] \quad (2.94)$$

where $|\Omega_I^{MM}(\mathbf{r})|^2$ represents the charge density at the point \mathbf{r} originating from the MM atom I located at position \mathbf{R}_I and having a charge q_I while α is a parameter controlling the width of the distribution.

Alternatively, it is possible to screen the electrostatic interactions at short-range [111]:

$$E_{QM/MM}^{ele} = \sum_{I \in MM} q_I \int d\mathbf{r} \rho(\mathbf{r}) v_I(|\mathbf{r} - \mathbf{R}_I|)$$

$$v_I(|\mathbf{r} - \mathbf{R}_I|) = v_I(r_I) \frac{r_c^4 - r_I^4}{r_c^5 - r_I^5} \quad (2.95)$$

where q_I is the classical point charge at position \mathbf{R}_I while r_c is the covalent radius of atom I.

References

- [1] R. G. Parr. “Density Functional Theory”. In: *Annu. Rev. Phys. Chem.* 34.1 (1983), pp. 631–656. DOI: 10.1146/annurev.pc.34.100183.003215.
- [2] T. Ziegler. “Approximate density functional theory as a practical tool in molecular energetics and dynamics”. In: *Chem. Rev.* 91.5 (1991), pp. 651–667. DOI: 10.1021/cr00005a001.
- [3] A. D. Becke. “Perspective: Fifty years of density-functional theory in chemical physics”. In: *J. Chem. Phys.* 140.18 (2014), 18A301. DOI: 10.1063/1.4869598.
- [4] R. O. Jones. “Density functional theory: Its origins, rise to prominence, and future”. In: *Rev. Mod. Phys.* 87 (3 2015), pp. 897–923. DOI: 10.1103/RevModPhys.87.897.
- [5] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford: Oxford University Press, 1989. DOI: 10.1093/oso/9780195092769.001.0001.
- [6] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Phys. Rev.* 136 (3B 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.
- [7] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Phys. Rev.* 140 (4A 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.
- [8] L. H. Thomas. “The calculation of atomic fields”. In: *Math. Proc. Camb. Phil. Soc.* 23.5 (1927), pp. 542–548. DOI: 10.1017/S0305004100011683.
- [9] E. Fermi. In: *Rend. Accad. Lincei* 6 (1927), pp. 602–607.
- [10] J. P. Perdew and K. Schmidt. “Jacob’s ladder of density functional approximations for the exchange-correlation energy”. In: *AIP Conference Proceedings* 577.1 (2001), pp. 1–20. DOI: 10.1063/1.1390175.
- [11] N. Mardirossian and M. Head-Gordon. “Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals”. In: *Mol. Phys.* 115.19 (2017), pp. 2315–2372. DOI: 10.1080/00268976.2017.1333644.
- [12] M. Gell-Mann and K. A. Brueckner. “Correlation Energy of an Electron Gas at High Density”. In: *Phys. Rev.* 106 (2 1957), pp. 364–368. DOI: 10.1103/PhysRev.106.364.
- [13] D. M. Ceperley and B. J. Alder. “Ground State of the Electron Gas by a Stochastic Method”. In: *Phys. Rev. Lett.* 45 (7 1980), pp. 566–569. DOI: 10.1103/PhysRevLett.45.566.
- [14] O. Gunnarsson and B. I. Lundqvist. “Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism”. In: *Phys. Rev. B* 13 (10 1976), pp. 4274–4298. DOI: 10.1103/PhysRevB.13.4274.
- [15] O. Gunnarsson, M. Jonson, and B. I. Lundqvist. “Exchange and correlation in inhomogeneous electron systems”. In: *Sol. Stat. Comm.* 24.11 (1977), pp. 765–768. DOI: [https://doi.org/10.1016/0038-1098\(77\)91185](https://doi.org/10.1016/0038-1098(77)91185).

- [16] T. Ziegler, A. Rauk, and E. J. Baerends. “On the calculation of multiplet energies by the hartree-fock-slater method”. In: *Theoret. Chim. acta* 43 (1977), pp. 261–271. DOI: <https://doi.org/10.1007/BF00551551>.
- [17] J. P. Perdew and A. Zunger. “Self-interaction correction to density-functional approximations for many-electron systems”. In: *Phys. Rev. B* 23 (10 1981), pp. 5048–5079. DOI: 10.1103/PhysRevB.23.5048.
- [18] J. P. Perdew and Y. Wang. “Accurate and simple analytic representation of the electron-gas correlation energy”. In: *Phys. Rev. B* 45 (23 1992), pp. 13244–13249. DOI: 10.1103/PhysRevB.45.13244.
- [19] L. A. Cole and J. P. Perdew. “Calculated electron affinities of the elements”. In: *Phys. Rev. A* 25 (3 1982), pp. 1265–1271. DOI: 10.1103/PhysRevA.25.1265.
- [20] S. H. Vosko, L. Wilk, and M. Nusair. “Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis”. In: *Can. J. Phys.* 58.8 (1980), pp. 1200–1211. DOI: <https://doi.org/10.1139/p80-159>.
- [21] J. P. Perdew, K. Burke, and Y. Wang. “Generalized gradient approximation for the exchange-correlation hole of a many-electron system”. In: *Phys. Rev. B* 54 (23 1996), pp. 16533–16539. DOI: 10.1103/PhysRevB.54.16533.
- [22] J. P. Perdew, K. Burke, and Y. Wang. “Erratum: Generalized gradient approximation for the exchange-correlation hole of a many-electron system [Phys. Rev. B 54, 16 533 (1996)]”. In: *Phys. Rev. B* 57 (23 1998), pp. 14999–14999. DOI: 10.1103/PhysRevB.57.14999.
- [23] K. Burke, J. P. Perdew, and Y. Wang. *Derivation of a Generalized Gradient Approximation: The PW91 Density Functional*. Boston: Springer, 1998. DOI: https://doi.org/10.1007/978-1-4899-0316-7_7.
- [24] J. P. Perdew, K. Burke, and M. Ernzerhof. “Generalized Gradient Approximation Made Simple”. In: *Phys. Rev. Lett.* 77 (18 1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.
- [25] J. P. Perdew, K. Burke, and M. Ernzerhof. “Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]”. In: *Phys. Rev. Lett.* 78 (7 1997), pp. 1396–1396. DOI: 10.1103/PhysRevLett.78.1396.
- [26] Y. Zhang and W. Yang. “Comment on “Generalized Gradient Approximation Made Simple””. In: *Phys. Rev. Lett.* 80 (4 1998), pp. 890–890. DOI: 10.1103/PhysRevLett.80.890.
- [27] B. Hammer, L. B. Hansen, and J. K. Norskov. “Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals”. In: *Phys. Rev. B* 59 (11 1999), pp. 7413–7421. DOI: 10.1103/PhysRevB.59.7413.
- [28] Z. Wu and R. E. Cohen. “More accurate generalized gradient approximation for solids”. In: *Phys. Rev. B* 73 (23 2006), p. 235116. DOI: 10.1103/PhysRevB.73.235116.

- [29] J. P. Perdew et al. "Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces". In: *Phys. Rev. Lett.* 100 (13 2008), p. 136406. DOI: 10.1103/PhysRevLett.100.136406.
- [30] J. Tao et al. "Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids". In: *Phys. Rev. Lett.* 91 (14 2003), p. 146401. DOI: 10.1103/PhysRevLett.91.146401.
- [31] C. Adamo and V. Barone. "Toward reliable density functional methods without adjustable parameters: The PBE0 model". In: *J. Chem. Phys.* 110.13 (1999), pp. 6158–6170. DOI: 10.1063/1.478522.
- [32] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". In: *J. Chem. Phys.* 98.7 (1993), pp. 5648–5652. DOI: 10.1063/1.464913.
- [33] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Phys. Rev. B* 37 (2 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.
- [34] A. R. Leach. *Molecular Modelling: Principles and Applications*. second edition. Pearson Education Limited, 2001.
- [35] D. van der Spoel. "Systematic design of biomolecular force fields". In: *Curr. Opin. Struct. Bio.* 67 (2021), pp. 18–24. DOI: <https://doi.org/10.1016/j.sbi.2020.08.006>.
- [36] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [37] *VMD Visual Molecular Dynamics*. URL: <https://www.ks.uiuc.edu/Research/vmd/>.
- [38] O. Guvench and A. D. MacKerell Jr. "Comparison of Protein Force Fields for Molecular Dynamics Simulations". In: *Molecular Modeling of Proteins*. Humana Press, 2008, pp. 63–88. ISBN: 978-1-59745-177-2. DOI: 10.1007/978-1-59745-177-2_4.
- [39] B. R. Brooks et al. "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations". In: *J. Comput. Chem.* 4.2 (1983), pp. 187–217. DOI: <https://doi.org/10.1002/jcc.540040211>.
- [40] L. Nilsson and M. Karplus. "Empirical energy functions for energy minimization and dynamics of nucleic acids". In: *J. Comput. Chem.* 7.5 (1986), pp. 591–616. DOI: <https://doi.org/10.1002/jcc.540070502>.
- [41] S. J. Weiner et al. "A new force field for molecular mechanical simulation of nucleic acids and proteins". In: *J. Am. Chem. Soc.* 106.3 (1984), pp. 765–784. DOI: 10.1021/ja00315a051.
- [42] W. D. Cornell et al. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules". In: *J. Am. Chem. Soc.* 117.19 (1995), pp. 5179–5197. DOI: 10.1021/ja00124a002.

- [43] W. L. Jorgensen and J. Tirado-Rives. "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin". In: *J. Am. Chem. Soc.* 110.6 (1988), pp. 1657–1666. DOI: 10.1021/ja00214a001.
- [44] G. A. Kaminski et al. "Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides". In: *J. Phys. Chem. B* 105.28 (2001), pp. 6474–6487. DOI: 10.1021/jp003919d.
- [45] X. Daura, A. E. Mark, and W. F. van Gunsteren. "Parametrization of aliphatic CH_n united atoms of GROMOS96 force field". In: *J. Comput. Chem.* 19.5 (1998), pp. 535–547. DOI: 10.1002/(SICI)1096-987X(19980415)19:5<535::AID-JCC6>3.0.CO;2-N.
- [46] L. D. Schuler, X. Daura, and W. F. van Gunsteren. "An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase". In: *J. Comput. Chem.* 22.11 (2001), pp. 1205–1218. DOI: <https://doi.org/10.1002/jcc.1078>.
- [47] P. Ren and J. W. Ponder. "Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations". In: *J. Comput. Chem.* 23.16 (2002), pp. 1497–1506. DOI: <https://doi.org/10.1002/jcc.10127>.
- [48] P. Ren and J. W. Ponder. "Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation". In: *J. Phys. Chem. B* 107.24 (2003), pp. 5933–5947. DOI: 10.1021/jp027815+.
- [49] Z. Jing et al. "Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications". In: *Ann. Rev. Biophys.* 48.1 (2019), pp. 371–394. DOI: 10.1146/annurev-biophys-070317-033349.
- [50] L. Verlet. "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules". In: *Phys. Rev.* 159 (1 1967), pp. 98–103. DOI: 10.1103/PhysRev.159.98.
- [51] W. C. Swope et al. "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". In: *J. Chem. Phys.* 76.1 (1982), pp. 637–649. DOI: 10.1063/1.442716.
- [52] R. W. Hockney, S. P. Goel, and J. W. Eastwood. "Quiet high-resolution computer models of a plasma". In: *J. Comput. Phys.* 14.2 (1974), pp. 148–158. DOI: [https://doi.org/10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2).
- [53] S. Nosé. "A molecular dynamics method for simulations in the canonical ensemble". In: *Mol. Phys.* 52.2 (1984), pp. 255–268. DOI: 10.1080/00268978400101201.
- [54] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". In: *Phys. Rev. A* 31 (3 1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695.
- [55] S. Nosé. "An extension of the canonical ensemble molecular dynamics method". In: *Mol. Phys.* 57.1 (1986), pp. 187–191. DOI: 10.1080/00268978600100141.

- [56] G. Ciccotti and G. Kalibaeva. “Molecular Dynamics of Complex Systems: Non-Hamiltonian, Constrained, Quantum-Classical”. In: *Novel Methods in Soft Matter Simulations*. Ed. by M. Karttunen, A. Lukkarinen, and I. Vattulainen. Springer Berlin Heidelberg, 2004, pp. 150–189. ISBN: 978-3-540-39895-0. DOI: 10.1007/978-3-540-39895-0_5.
- [57] G. J. Martyna, M. L. Klein, and M. Tuckerman. “Nosé–Hoover chains: The canonical ensemble via continuous dynamics”. In: *J. Chem. Phys.* 97.4 (1992), pp. 2635–2643. DOI: 10.1063/1.463940.
- [58] M. E. Tuckerman et al. “Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals”. In: *J. Chem. Phys.* 99.4 (1993), pp. 2796–2808. DOI: 10.1063/1.465188.
- [59] H. J. C. Berendsen et al. “Molecular dynamics with coupling to an external bath”. In: *J. Chem. Phys.* 81.8 (1984), pp. 3684–3690. DOI: 10.1063/1.448118.
- [60] S. C. Harvey, R. K. Z. Tan, and T. E. Cheatham III. “The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition”. In: *J. Comput. Chem.* 19.7 (1998), pp. 726–740. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199805\)19:7<726::AID-JCC4>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-987X(199805)19:7<726::AID-JCC4>3.0.CO;2-S).
- [61] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *J. Chem. Phys.* 126.1 (2007), p. 014101. DOI: 10.1063/1.2408420.
- [62] T. Schneider and E. Stoll. “Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions”. In: *Phys. Rev. B* 17 (3 1978), pp. 1302–1322. DOI: 10.1103/PhysRevB.17.1302.
- [63] H. C. Andersen. “Molecular dynamics simulations at constant pressure and/or temperature”. In: *J. Chem. Phys.* 72.4 (1980), pp. 2384–2393. DOI: 10.1063/1.439486.
- [64] S. Nosé and M. L. Klein. “Constant pressure molecular dynamics for molecular systems”. In: *Mol. Phys.* 50.5 (1983), pp. 1055–1076. DOI: 10.1080/00268978300102851.
- [65] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. DOI: 10.1063/1.328693.
- [66] M. Parrinello, A. Rahman, and P. Vashishta. “Structural Transitions in Superionic Conductors”. In: *Phys. Rev. Lett.* 50 (14 1983), pp. 1073–1076. DOI: 10.1103/PhysRevLett.50.1073.
- [67] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. 2nd. Cambridge University Press, 2004. ISBN: 0521825687.
- [68] D. Frenkel and B. Smit. *Understanding Molecular Simulations (Second Edition)*. second edition. Academic Press, 2002. ISBN: 978-0-12-267351-1.
- [69] M. P. Allen and D. J. Tildesley. *Computer Simulations of Liquids: Second Edition*. second edition. Oxford University Press, 2017. ISBN: 9780198803195.

- [70] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". In: *J. Comput. Phys.* 23.3 (1977), pp. 327–341. DOI: [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5).
- [71] H. C. Andersen. "Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations". In: *J. Comput. Phys.* 52.1 (1983), pp. 24–34. DOI: [https://doi.org/10.1016/0021-9991\(83\)90014-1](https://doi.org/10.1016/0021-9991(83)90014-1).
- [72] S. Miyamoto and P. A. Kollman. "Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models". In: *J. Comput. Chem.* 13.8 (1992), pp. 952–962. DOI: <https://doi.org/10.1002/jcc.540130805>.
- [73] B. Hess et al. "LINCS: A linear constraint solver for molecular simulations". In: *J. Comput. Chem.* 18.12 (1997), pp. 1463–1472. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [74] R. Edberg, D. J. Evans, and G. P. Morriss. "Constrained molecular dynamics: Simulations of liquid alkanes with a new algorithm". In: *J. Chem. Phys.* 84.12 (1986), pp. 6933–6939. DOI: 10.1063/1.450613.
- [75] A. Baranyai and D. J. Evans. "New algorithm for constrained molecular-dynamics simulation of liquid benzene and naphthalene". In: *Mol. Phys.* 70.1 (1990), pp. 53–63. DOI: 10.1080/00268979000100841.
- [76] M. Born and K. Huang. *Dynamical Theory of Crystal Lattices*. Ed. by Clarendon Press. 1954.
- [77] W. Kutzelnigg. "The adiabatic approximation I. The physical background of the Born-Handy ansatz". In: *Mol. Phys.* 90.6 (1997), pp. 909–916. DOI: 10.1080/002689797171904.
- [78] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln". In: *Annalen der Physik* 84.20 (1927), pp. 457–484.
- [79] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Ed. by Oxford University Press. 3rd edition. Oxford, 1947.
- [80] A. Messiah. *Quantum Mechanics*. Ed. by North Holland. Amsterdam, 1964.
- [81] J. J. Sakurai. *Modern Quantum Mechanics*. Ed. by Addison-Wesley. Redwood City, 1985.
- [82] H. Goldstein, C. P. Poole, and J. L. Safko. *Classical Mechanics*. Ed. by Addison-Wesley. 3rd edition. San Francisco, 2002.
- [83] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge: Cambridge University Press, 2009.
- [84] I. S. Y. Wang and M. Karplus. "Dynamics of organic reactions". In: *J. Am. Chem. Soc.* 95.24 (1973), pp. 8160–8164. DOI: 10.1021/ja00805a033.
- [85] A. Warshel and M. Karplus. "Semiclassical trajectory approach to photoisomerization". In: *Chem. Phys. Lett.* 32.1 (1975), pp. 11–17. DOI: [https://doi.org/10.1016/0009-2614\(75\)85157-8](https://doi.org/10.1016/0009-2614(75)85157-8).

- [86] C. Leforestier. "Classical trajectories using the full ab initio potential energy surface $\text{H}+\text{CH}_4\rightarrow\text{CH}_4+\text{H}$ ". In: *J. Chem. Phys.* 68.10 (1978), pp. 4406–4410. DOI: 10.1063/1.435520.
- [87] T. Helgaker, P. Jørgensen, and J. Olsen. *Molecular Electronic-Structure Theory*. John Wiley & Sons, Ltd, 2000. ISBN: 9781119019572. DOI: <https://doi.org/10.1002/9781119019572.ch11>.
- [88] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Berlin: Springer Science & Business Media, 1990. ISBN: 978-3-540-51993-5. DOI: <https://doi.org/10.1007/978-3-642-86105-5>.
- [89] M. C. Payne et al. "Molecular Dynamics and ab initio Total Energy Calculations". In: *Phys. Rev. Lett.* 56 (24 1986). DOI: 10.1103/PhysRevLett.56.2656.
- [90] R. Car and M. Parrinello. "Unified Approach for Molecular Dynamics and Density-Functional Theory". In: *Phys. Rev. Lett.* 55 (22 1985), pp. 2471–2474. DOI: 10.1103/PhysRevLett.55.2471.
- [91] A. Warshel. "Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights". In: *Annu. Rev. Biophys. Biomol. Struct.* 32.1 (2003), pp. 425–443. DOI: 10.1146/annurev.biophys.32.110601.141807.
- [92] H. Lin and D. G. Truhlar. "QM/MM: what have we learned, where are we, and where do we go from here?" In: *Theor. Chem. Acc.* 117 (2007), pp. 185–199. DOI: <https://doi.org/10.1007/s00214-006-0143-z>.
- [93] H. M. Senn and W. Thiel. "QM/MM Methods for Biomolecular Systems". In: *Angew. Chem. Int. Ed.* 48.7 (2009), pp. 1198–1229. DOI: <https://doi.org/10.1002/anie.200802019>.
- [94] P. Vidossich and A. Magistrato. "QM/MM Molecular Dynamics Studies of Metal Binding Proteins". In: *Biomolecules* 4.3 (2014), pp. 616–645. DOI: 10.3390/biom4030616.
- [95] M. W. van der Kamp and A. J. Mulholland. "Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology". In: *Biochem.* 52.16 (2013), pp. 2708–2728. DOI: 10.1021/bi400215w.
- [96] R. A. Friesner and V. Guallar. "Ab Initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis". In: *Annu. Rev. Phys. Chem.* 56.1 (2005), pp. 389–427. DOI: 10.1146/annurev.physchem.55.091602.094410.
- [97] A. Warshel and M. Levitt. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *J. Mol. Biol.* 103.2 (1976), pp. 227–249. DOI: [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9).
- [98] A. Warshel and M. Karplus. "Calculation of ground and excited state potential surfaces of conjugated molecules. I. Formulation and parametrization". In: *J. Am. Chem. Soc.* 94.16 (1972), pp. 5612–5625. DOI: 10.1021/ja00771a014.

- [99] R. R. Birge, M. J. Sullivan, and B. E. Kohler. "The Effect of Temperature and Solvent Environment on the Conformational Stability of 11-Cis-Retinal". In: *J. Am. Chem. Soc.* 98 (1976), pp. 358–367.
- [100] W. Kohn. "Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms". In: *Phys. Rev. Lett.* 76 (17 1996), pp. 3168–3171. DOI: 10.1103/PhysRevLett.76.3168.
- [101] D. Bakowies and W. Thiel. "Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches". In: *J. Phy. Chem.* 100.25 (1996), pp. 10580–10594. DOI: 10.1021/jp9536514.
- [102] M. Svensson et al. "ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for DielsAlder Reactions and Pt(P(t-Bu)₃)₂ + H₂ Oxidative Addition". In: *J. Phys. Chem.* 100.50 (1996), pp. 19357–19363. DOI: 10.1021/jp962071j.
- [103] F. Maseras and K. Morokuma. "IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states". In: *J. Comput. Chem.* 16.9 (1995), pp. 1170–1179. DOI: <https://doi.org/10.1002/jcc.540160911>.
- [104] E. Brunk and U. Rothlisberger. "Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States". In: *Chem. Rev.* 115.12 (2015), pp. 6217–6263. DOI: <https://doi.org/10.1021/cr500628b>.
- [105] U. C. Singh and P. A. Kollman. "A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH₃Cl + Cl exchange reaction and gas phase protonation of polyethers". In: *J. Comp. Chem.* 7.6 (1986), pp. 718–730. DOI: <https://doi.org/10.1002/jcc.540070604>.
- [106] I. Antes and W. Thiel. "Adjusted Connection Atoms for Combined Quantum Mechanical and Molecular Mechanical Methods". In: *J. Phys. Chem. A* 103.46 (1999), pp. 9290–9295. DOI: 10.1021/jp991771w.
- [107] N. Reuter et al. "Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches". In: *J. Phys. Chem. A* 104.8 (2000), pp. 1720–1735. DOI: 10.1021/jp9924124.
- [108] X. Assfeld and J. L. Rivail. "Quantum chemical computations on parts of large molecules: the ab initio local self consistent field method". In: *Chem. Phys. Lett.* 263.1 (1996), pp. 100–106. DOI: [https://doi.org/10.1016/S0009-2614\(96\)01165-7](https://doi.org/10.1016/S0009-2614(96)01165-7).
- [109] J. Gao et al. "A Generalized Hybrid Orbital (GHO) Method for the Treatment of Boundary Atoms in Combined QM/MM Calculations". In: *J. Phys. Chem. A* 102.24 (1998), pp. 4714–4721. DOI: 10.1021/jp9809890.

-
- [110] O. A. von Lilienfeld et al. "Variational optimization of effective atom centered potentials for molecular properties". In: *J. Chem. Phys.* 122.1 (2005), p. 014113. DOI: 10.1063/1.1829051.
- [111] A. Laio, J. VandeVondele, and U. Röthlisberger. "A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations". In: *J. Chem. Phys.* 116.16 (2002), pp. 6941–6947. DOI: 10.1063/1.1462041.

3 Coordination of Mn(II) in SLC11/NRAMP transporters uncovered by QM/MM simulations

Chapter 3 is a pre-submission version of the article: "Coordination of Mn(II) in SLC11/NRAMP transporters uncovered by QM/MM simulations" by **Maria Letizia Merlini**, Alessandra Magistrato, Paolo Carloni and Ursula Röthlisberger. My contribution: computational studies, results elaboration and discussion, graphics and pictures production, draft writing.

3.1 Abstract

The SLC11/NRAMP transporters are responsible for the proton-coupled transport of divalent transition-metal ions across cellular membranes. They provide micronutrients fundamental for cell functionalities and contribute in preserving cellular homeostasis. Recent studies have demonstrated that malfunction of SLC11/NRAMP transporters can lead to an overload of metal ions in nervous system cells in turn leading to a severe impairment of cellular metabolic functions and eventually to the onset of neurodegenerative diseases. Despite recent advances in the comprehension of the structural basis of transport, most mechanistic details remain obscure. A crucial information needed to understand the transport mechanism is the identification of the complete coordination sphere of the divalent transition-metal ion in the protein binding site. Unfortunately, in the only available crystal structure of the substrate-bound inward facing conformation, the metal binding site is only partially resolved presenting a 4-fold coordinated Mn^{2+} ion in a highly distorted geometry, largely different from the one found for the outward facing conformation. In this work, QM/MM molecular dynamics simulations have been used as a general tool to assist the refinement of partially resolved metal binding sites in the case of the prokaryotic member of the family ScaDMT in complex with a Mn^{2+} ion in the inward facing conformation. This multiscale procedure, using an accurate quantum treatment of the active site and a classical treatment of the surroundings, allowed for state-of-the-art calculations of the entire system. Our findings suggest a nearly perfect octahedral coordination geometry with two additional water molecules filling axial coordination sites unresolved by the crystallographic experiments.

3.2 Introduction

Metal ions transporters are trans-membrane proteins able to transfer metal ions in and out of the cell and are therefore fundamental to maintain cellular homeostasis [1, 2]. They recently gained special attention because of their possible implication in the onset of neurodegenerative diseases such as Alzheimer Disease (AD), Parkinson Disease (PD), manganism and others, as indicated during the last decades by several medical studies [3–11]. Their malfunctioning can, in fact, lead to an abnormal concentration of metal ions in nervous system cells, which in turn leads to a severe impairment of cellular metabolic functions.

The influx of divalent transition-metal ions (Fe^{2+} , Mn^{2+} , Cu^{2+} , Zn^{2+} , etc.) across cellular membranes is in particular regulated by members of the “solute carrier 11” (SLC11) or “natural resistance-associated macrophage proteins” (NRAMPs) family of transporters [12, 13]. These proteins are widely expressed in both prokaryotic and eukaryotic living organisms with a high degree of sequence conservation [14] (with sequence identities/ similarities with respect to the mammalian homologs of 28%/46%, 40%/58% and 55%/73% for yeasts, plants and fly proteins, respectively) [15]. All the characterized members of the family have been found to carry out coupled divalent transition-metal ion/ H^+ transport [16]. Furthermore, they display remarkable selectivity towards Mn^{2+} , Fe^{2+} , Co^{2+} and Cd^{2+} while Ni^{2+} , Cu^{2+} and Zn^{2+} are

weaker substrates and divalent alkaline-earth metal ions (Ca^{2+} , Sr^{2+} , Ba^{2+} , etc.) are selectively excluded from transport [17].

The human genome, in particular, encodes for two SLC11 transporters: SLC11A1 (NRAMP1) [18–21] and SLC11A2 (NRAMP2) [22–25]. The latter, better known as DMT1 [26–29] (Divalent Metal-Ion Transporter 1) is the best characterized member of the family because of its expression in a wide range of human tissues such as brain, kidney, testis and placenta [30]. Point mutations in DMT1 have been found to cause anemia in humans and rodents [31–41].

NRAMP transporters work via an “alternate access mechanism”, in which the protein is alternately open towards the extracellular matrix, from where the metal ion approaches and binds to the active site (“outward facing conformation”, OFC) or towards the cellular cytoplasm where the ion is finally released (“inward facing conformation”, IFC) (Fig. 3.1) [42–45].

Several crystal structures of SLC11/NRAMP transporters are nowadays available. All of them are derived from prokaryotic members of the family: *Staphylococcus Capitis* DMT (ScaDMT), *Deinococcus Radiodurans* DMT (DraDMT) and *Eremococcus Coleocola* DMT (EcoDMT). In particular, Dutzler and co-workers crystallized ScaDMT-IFC in both its complex with a Mn^{2+} ion as substrate and in a substrate-free (apo) form (PDB IDs: 4WGW/5M95 and 4WGV/5M94 respectively) (2014) [46] while EcoDMT-OFC was crystallized in an apo conformation (PDB ID: 5M87) (2017) [47]. Bozzi et al. derived instead crystals of DraDMT in four different conformations based on the apo IFC (PDB ID: 6D9W) (equivalent to the apo form of ScaDMT) crystallized in 2016 [48]. In 2019, a G223W mutant of DraDMT-OFC was resolved in the apo form (PDB ID: 6D91) (equivalent to EcoDMT), and in a Mn^{2+} -bound state (PDB ID: 6BU5) (the first substrate-bound outward-facing conformation). Moreover, a novel inward-facing apo structure in which the active site is not in direct contact with the aqueous cytoplasmic vestibule, referred to by the authors as “inward-occluded” form (IOC) was crystallized for the Dra-G45R mutant (PDB ID: 6C3I) [49].

The latter is reputed to be an intermediate conformation between the apo IFC (structure 4 Fig. 3.1) and the apo OFC (structure 6 Fig. 3.1) conformations. The five distinct conformations observed for SLC11/NRAMP transporters thus outline an alternate access metal-ion transport cycle likely consisting of at least six conformational states, as schematically depicted in Fig. 3.1. The sixth conformational state of the supposed cycle, that has not yet been identified, is likely a substrate-bound intermediate form between the substrate-bound OFC and the substrate-bound IFC (structure 2 Fig. 3.1).

The structural information obtained from these studies revealed the general architecture of SLC11/NRAMP transporters and provided a framework for the conformational changes underlying the metal ion transport via the alternate access mechanism. Despite of this, many mechanistic details about the proton-coupled transport of divalent transition-metal ions remain obscure. Such knowledge would be of crucial importance in order to develop therapeutic strategies for the regulation of metal-ions homeostasis in cells with many potential applications in medicine.

In particular, in this work we focus on the metal-ion binding site of ScaDMT in the inward-

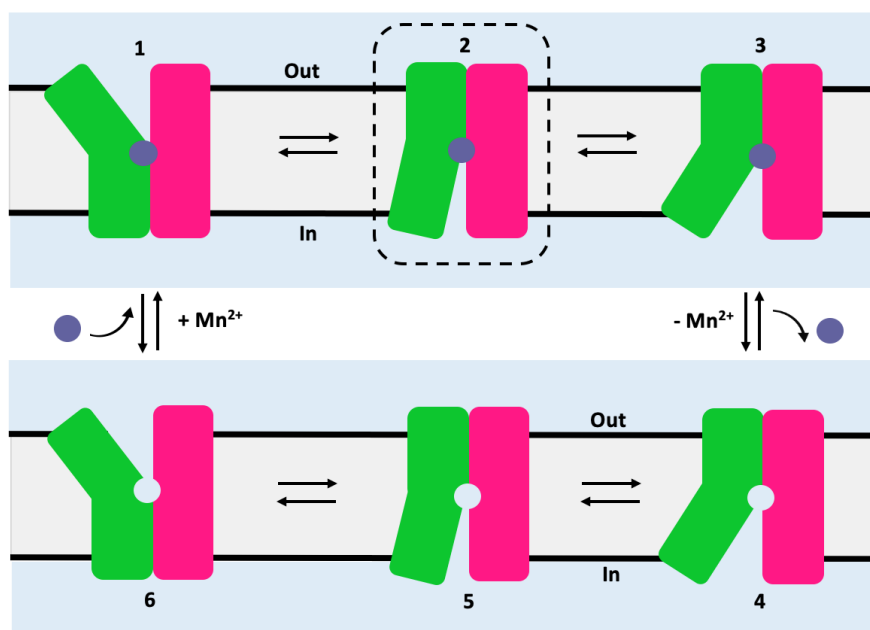


Figure 3.1: Schematic representation of the putative metal ion transport alternate access mechanism [49]. It involves at least six distinct conformational states, five of which have been identified and (at least partially) crystallized for different systems: (1) Mn²⁺-bound outward-facing conformation (for DraDMT with a G223W mutant) [49]; (2) Mn²⁺-bound outward-occluded conformation (no crystal structure available) ; (3) Mn²⁺-bound inward-facing conformation (ScaDMT) [46]; (4) apo inward-facing conformation (apo ScaDMT [46] and apo DraDMT [48]); (5) apo inward-occluded conformation (DraDMT as G45R mutant) [49]; and (6) apo outward-facing conformation (apo DraDMT as G223W mutant [49] and EcoDMT [47]). The conformational state (2) is still missing. The two symmetrically related halves of the protein (α -helices 1-5 and 6-10) are colored in green and magenta, respectively while the Mn²⁺ ion is represented as a blue sphere. The cellular membrane is represented in light gray whereas the extracellular matrix and the cytoplasm are represented in light blue.

facing conformation. As already mentioned above, the X-ray structure of ScaDMT-IFC was determined in 2014 by Dutzler and co-workers, with a resolution of 3.40 Å in a substrate (Mn²⁺)-bound state [46]. The observed partially resolved active site and the cytoplasmatic vestibule in contact with it are shown in Fig. 3.3.

The protein structure, containing 11 trans-membrane α -helices, some of which bent or interrupted by short loops, revealed a pseudo-symmetric relationship between two structurally analogous sub-domains formed by α -helices 1-5 and 6-10, with the N terminus located in the cytoplasm and the C terminus on the periplasmic side (Fig. 3.2). This protein fold, conserved among all the previously mentioned members of the family [46–49], is shared with several other transporters[50–55] including the amino acid transporter LeuT [56]. Furthermore, the disease-linked mutations observed in DMT1 were all but one observed also in ScaDMT and, similarly to the human protein, they are scattered in different parts of the protein, located in trans-membrane helices remote from the ion-binding region. The Mn²⁺ ion in the ScaDMT-IFC binding site is coordinated by the carbonyl oxygen of Ala223 (Ala223:O) and the side chains of Asp49 (Asp49:O ^{δ 2}), Asn52 (Asn52:O ^{δ 1}) and Met226 (Met226:S ^{δ}) (Fig. 3.3, right panel). The coordinating atoms are thus three oxygen atoms acting as hard ligands plus a soft thioether sulfur ligand and are located in the unwound regions of α -helices 1 and 6, at the center of the

lipid bilayer. However, as shown in Table 1a, the determined Mn^{2+} -Asn52($\text{O}^{\delta 1}$) bond length appears to be unusually long (2.90 Å) [57]. Furthermore, as shown in Table 1b and in Fig. 3.3, the four ligands are arranged in a highly distorted pseudo square-planar geometry with highly dissimilar ligand-metal-ligand bond angles ranging from 72.8° (Asp49($\text{O}^{\delta 2}$)- Mn^{2+} -Asn52($\text{O}^{\delta 1}$)) to 108.5° (Met226(S^{δ})- Mn^{2+} -Asp49($\text{O}^{\delta 2}$)) and a far from planar dihedral angle formed by the four coordinating atoms Asp49($\text{O}^{\delta 2}$)-Asn52($\text{O}^{\delta 1}$)-Ala223(O)-Met226(S^{δ}) of 46.5°. The highly distorted geometry in conjunction with the limited crystallographic resolution (3.4 Å) suggest that the metal binding site might have only been partially resolved. This conjecture is strengthened by the fact that Mn^{2+} is often found in an octahedral coordination geometry [58, 59]. All this evidence seems to point to the fact that structural model of the active site is incomplete and suggests the need for further investigations for a definitive characterization of the nature of the coordination sphere of divalent transition-metal ions in the binding site of SLC11 transporters.

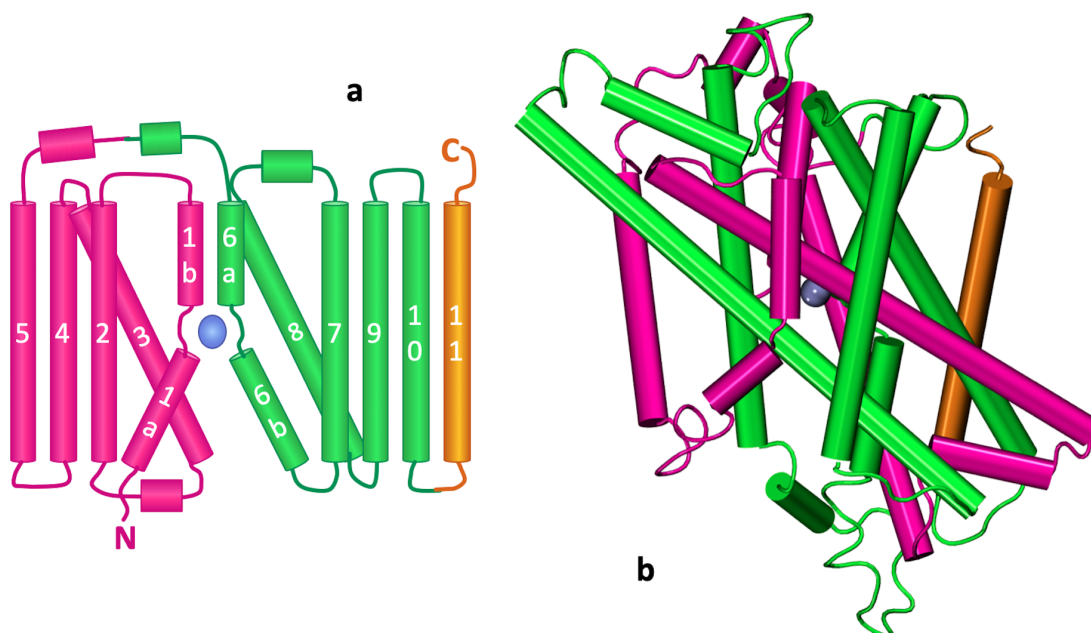


Figure 3.2: ScaDMT-Mn²⁺ topology (a) and structure (b). Protein helices are represented as cylinders while the Mn²⁺ ion is depicted as a blue sphere. The two halves of the protein related by a pseudo-symmetric relationship are colored in magenta (α -helices 1-5) and green (α -helices 6-10) respectively; the additional α -helix 11 is colored in orange. The helix numbers are labeled in (a) in order to schematically show the protein fold. The representation of the structure in (b) has been obtained using VMD [60, 61].

In this work, we performed quantum mechanics /molecular mechanics (QM/MM) molecular dynamics (MD) simulations of the ScaDMT-IFC in complex with Mn^{2+} , in order to shed light on this issue. In this multiscale approach atoms close to the Mn^{2+} binding site (as suggested by prior classical MD simulations) were treated at the quantum level, while the rest of the system was treated at a classical force field (FF) level. In this way, amino acid residues and water molecules near the active site are left free to coordinate or exchange in the metal coordination sphere. Anticipating our results, we find that after ca. 68 ps of DFT-based

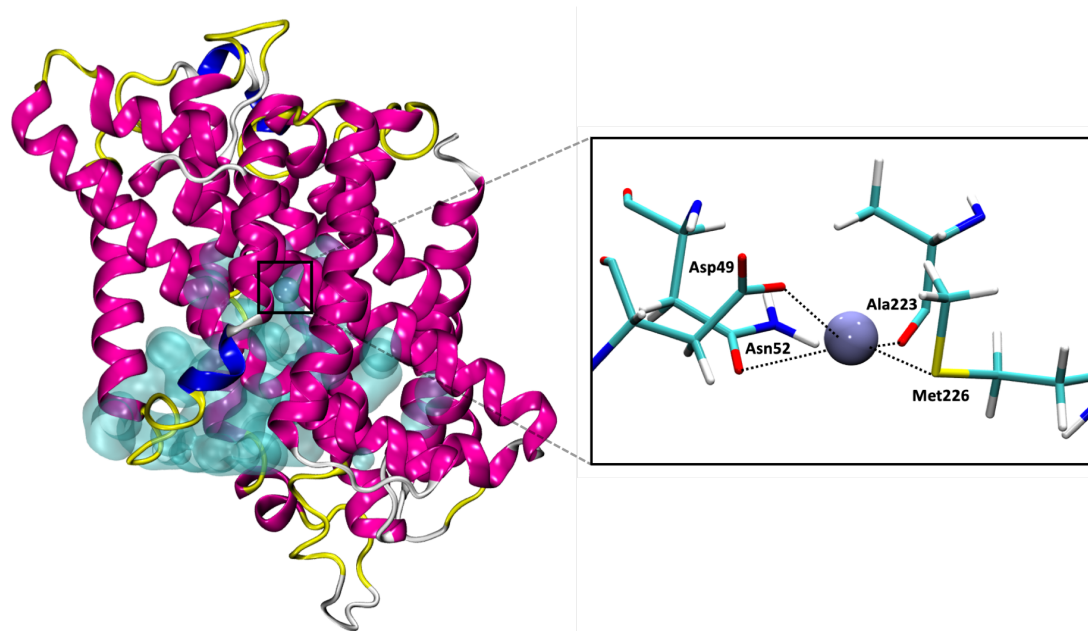


Figure 3.3: Structure of ScaDMT-Mn²⁺ in its inward-facing conformation (**left**) and detail of its metal ion binding site (**right**) as obtained from X-ray crystallography [46]. The protein α -helices are colored in magenta, Mn²⁺ is represented as a blue sphere and the four coordinating amino acid residues are represented in licorice style. C atoms are colored in cyan, O atoms in red, N atoms in blue, S in yellow and H atoms in white. The water accessible volume close to the active site is represented as a cyan mesh. Both pictures have been obtained with VMD [60, 61].

QM/MM MD, the coordination sphere of the Mn²⁺ ion reaches a stable equilibrium state in which the coordinating four amino acid residues suggested by the X-ray experiments rearrange themselves into a nearly ideal octahedral coordination geometry with two additional water molecules occupying axial positions.

3.3 Computational Details

To equilibrate the initial structure in a realistic environment, classical molecular dynamics simulations of the ScaDMT-IFC-Mn²⁺ protein-ion complex were performed in a lipid bilayer mimicking the nervous system cell membrane surrounded by a water solution with Cl⁻ and K⁺ ions in physiological concentration (96.03 and 91.35 mM, respectively). For the protein we used the crystal structure determined by Dutzler and co-workers [46] and available at the RCSB Protein Data Bank [62–64] (PDB ID: 4WGW, resolution 3.40 Å). The N- and C- termini of the protein were capped with N-acetyl (ACE) and N-methyl amide (NME) groups, respectively, in order to neutralize their charge. Protonation states of all ionizable residues of the protein were calculated at the physiological value of pH 7.4, using the H⁺⁺ web-server [65, 66].

Since ScaDMT is a transmembrane protein, we computed its initial orientation with respect to the hydrocarbon core of the lipid bilayer with the help of the OPM (Orientations of Proteins in Membranes) database [67, 68]. For this, the protein is treated as a rigid body able to freely float

in a hydrophobic slab of adjustable thickness. Its spatial arrangement is calculated through the minimization of the free energy needed for the transfer from the bulk water to the lipid bilayer. The three-component membrane containing POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine), POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) and CHL (cholesterol) was created using the MemBuilder II web-based graphical user interface [69, 70]. A symmetric structure of the membrane (i.e. same number of lipids in the two leaflets) with a square base length of 105 Å and thickness of 51 Å has been used. Initially, (i.e. before insertion of the protein) it contained a total of 450 lipid molecules: 190 POPC, 152 POPE and 108 cholesterol molecules [71, 72]. Some of the lipid molecules (24 POPC, 15 POPE and 9 CHL molecules, respectively) have been removed to incorporate the protein by using the *gmX* *membed* tool [73] of the Gromacs software [74–76]. At the end of the embedding process the membrane was thus formed by 166 POPC, 137 POPE and 99 CHL molecules (402 lipid molecules in total) with a final percentage composition of roughly 41%, 34% and 25% respectively. The system was then solvated using the *gmX* *solvate* option (with 71161 H₂O molecules) and K⁺ and Cl[−] ions (117 and 123, respectively) were added with the *gmX* *genion* option, both of them included in the Gromacs package [74–76]. The final system contained a total of 266706 atoms and is shown in Fig. 3.4.

We used the Amber99SB-ILDN force field [77] (FF) for the protein whereas the Amber lipid 14 FF [78] has been used to describe the membrane lipids. To treat the solvent, the 3-site rigid TIP3P water model has been used [79]. K⁺ and Cl[−] ions were instead described using the parameters from Joung and Cheatham [80]. Mn²⁺ force field parameters used to describe the cation have been taken from the AMBER parameter database [81, 82]. However, due to the limited accuracy of classical MD force field parameters in reproducing metal ion interactions, harmonic position restraints (force constants $f_x = f_y = f_z = 100000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) have been applied during the entire classical MD simulation for the Mn²⁺ ion and its coordination sphere (amino acidic residues Asp49, Asn52, Ala223 and Met226) in order to retain the active site as close as possible to the experimental data. In addition, harmonic position restraints (force constants $f_x = f_y = f_z = 10000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) have initially also been applied on the protein heavy atoms (i.e. protein backbone and side chains, excluding hydrogen atoms) and were subsequently slowly removed in order to thermalize the system. Classical MD simulations were performed using the Gromacs package (version 2016-4). After a long equilibration run of 600 ns, a production run of 1.0 μs was then performed using an integration time step of 2 fs with all covalent bonds involving hydrogen atoms constrained with the LINCS algorithm [83]. The Particle Mesh Ewald (PME) method [84, 85] was used to account for long-range electrostatic interactions. A switch function with a real space cutoff of 1.2 nm was used to account for van der Waals interactions. Simulations were run in the isothermal-isobaric NTP ensemble using the Parrinello-Rahman barostat [86] (1 bar) and the Nose-Hoover thermostat [87–89] (310 K).

From the end of the classical MD trajectory a single representative frame was selected to perform hybrid quantum mechanics/molecular mechanics (QM/MM) MD simulations [90]. Here we used the fully Hamiltonian QM/MM scheme introduced by R  thlisberger and coworkers [91] and implemented within the CPMD package [92, 93]. In this approach, a properly mod-

ified Coulomb potential is used to explicitly take into account the short-range electrostatic interactions between the QM and MM regions within a certain radius around every QM atom ($r = 10.58 \text{ \AA}$) while the long-range QM/MM interactions are taken into account via a multipole expansion. We refer the reader to the review in reference [94] for a detailed description of the QM/MM MD approach.

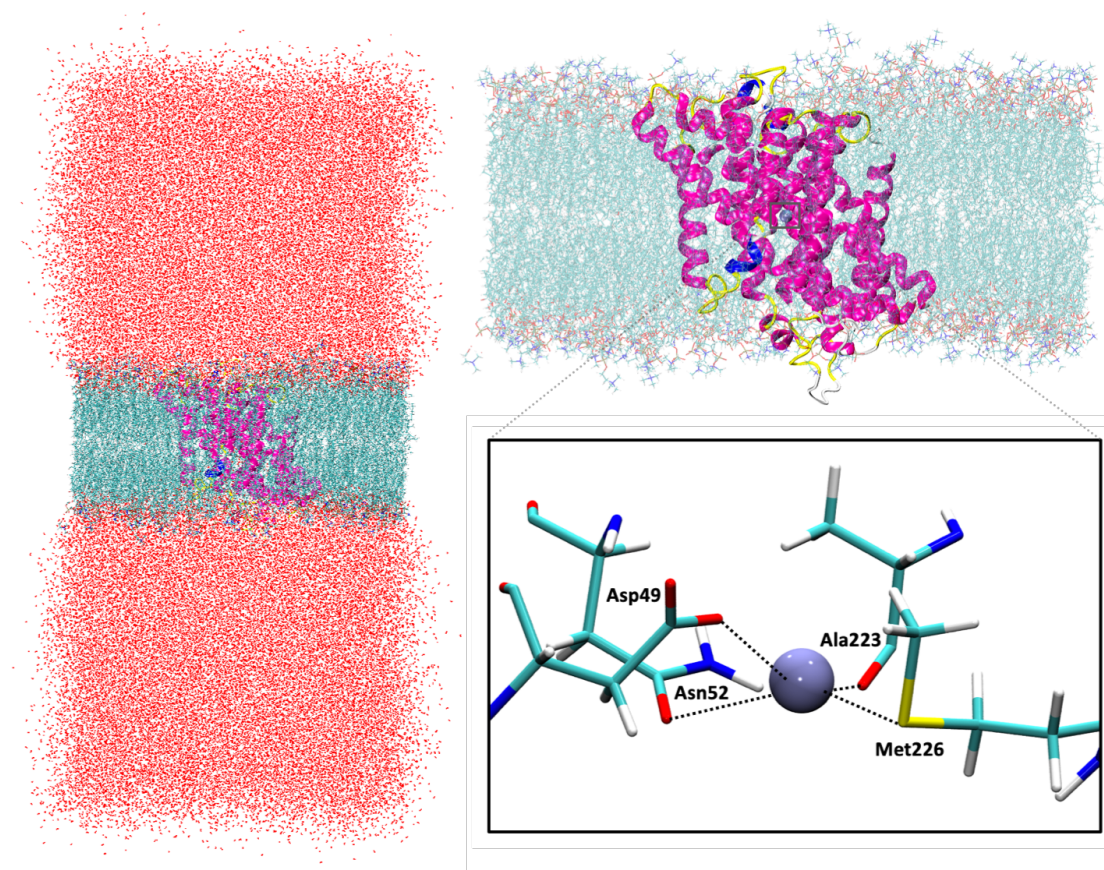


Figure 3.4: Frontal view of the whole computational model comprising the three component CHL-POPC-POPE membrane, the ScaDMT-IFC protein embedded into the lipid bilayer and in complex with the Mn(II) cation and the pool of solvent (H_2O) and monovalent ions K^+ and Cl^- around the membrane (**left**); zoom into the membrane-embedded ScaDMT-IFC- Mn^{2+} protein (**top right**); magnification of the Mn^{2+} active site region (**bottom right**). All the pictures have been obtained using VMD [60, 61].

In this study, the metal-ion binding-site, representing the active part of the system as well as the region of interest around the active site, has been treated at the QM level. The QM part was described at the Density Functional Theory (DFT) level [95–97] using a BLYP [98, 99] GGA (Generalized Gradient Approximation) exchange-correlation functional. A plane-wave basis set with a cutoff of 90 Ry was employed for the expansion of the single-particle wavefunctions and the core electrons were described through norm-conserving pseudopotentials of the Martins-Troullier type [100]. For the solution of the Poisson equation for the periodically-decoupled QM system the Hockney method [101] was chosen. A spin-unrestricted formalism was used since the ground state of the system is a sextet (for further details please refer to table S1 in the SI section). The rest of the system was included in the MM region and treated using the same

force fields as the ones mentioned above for the classical MD simulation. The dangling bonds at the QM/MM boundary were saturated using a monoatomic pseudopotential for C_{sp3} on C atoms placed in nonpolar environments [102]. QM/MM ab-initio Born-Oppenheimer and Car-Parrinello MD simulations were performed using the CPMD software [92, 93] (version 4.1). A hydrogen bond network analysis was performed using the VMD specific plugin [103] and providing as criteria for the formation of a hydrogen bond a donor-acceptor cutoff distance equal to 3.5 Å and a donor-hydrogen-acceptor angle cutoff equal to 30° [104].

3.4 Results

Classical and QM/MM molecular dynamics simulations have been used to determine the coordination structure of the Mn^{2+} metal ion in ScaDMT. Already from classical MD (in which the 4 coordinating ligands identified in the crystal structure were kept fixed at their crystallographic positions) we observe that water molecules participate to the coordination of the metal ion. This is evident from the time-dependent analysis of the distances between the manganese ion and the molecules approaching its coordination sphere (Fig. 3.5). The average values of these distances are given in SI.

Figure 3.5 shows that three water molecules coordinate Mn^{2+} during the entire classical MD simulation (1.6 μs) (including thermalization of 600 ns followed by a production run of 1000 ns). In particular, three water molecules (WAT24524-light blue line, WAT1701-red line and WAT56324-blue line) enter the Mn^{2+} coordination shell already during the thermalization phase of classical MD. While WAT56324 and WAT1701 keep their positions until the end of the simulation, WAT24524 exchanges with WAT3067 at around ~ 1100 ns. WAT3067 (purple line) approaches the active site around 650-700 ns, keeping a distance of $\sim 4-5$ Å. When WAT24524 leaves its coordination site (around 1088 ns) WAT3067 takes its place and remains there until the end of the simulation. Further details regarding the access of waters to the Mn^{2+} coordination sphere, the exchange between WAT24524 and WAT3067 and the average values of distances between Mn^{2+} and the four mentioned water molecules are given in SI. The presence of water molecules in the ion coordination sphere seems reasonable considering that the protein active site in its inward facing conformation is in direct contact with a pool of bulk water. As a result, classical MD simulations predict seven ligands to participate to the Mn^{2+} coordination: three water molecules and the four amino acid residues (Asp49, Asn52, Ala223 and Met226) already observed by X-ray crystallography.

It is also interesting to observe that distances between Mn^{2+} and Gln389, Gly46 and Asp49($O^{\delta 1}$) (predicted as metal ligands in the computational study by Pujol-Gimenéz and co-workers on ScaDMT in complex with a Cd^{2+} ion [105]) clearly never drop below 4 Å, thus indicating that those ligands never enter the first coordination shell of the ion. Similarly, our results are in contrast with the coordination sphere hypothesized for ScaDMT- Mn^{2+} by Bozzi et al. in ref. [49] who suggested a heptacoordinated Mn^{2+} coordination sphere composed of Asn52($O^{\delta 1}$), Ala223(O), Met226(S^{δ}), Asp49 acting as a bidentate ligand ($O^{\delta 1}$ and $O^{\delta 2}$) and

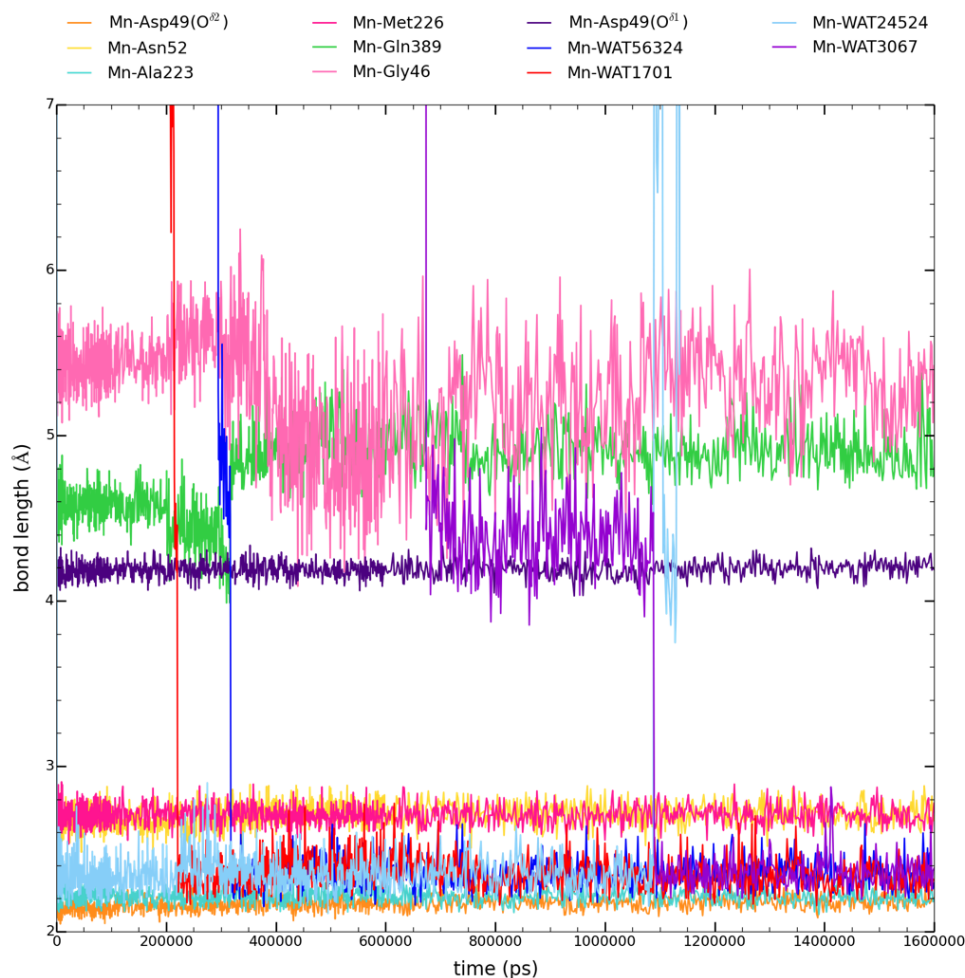


Figure 3.5: Distances between Mn^{2+} and the molecules approaching its coordination sphere during the entire classical MD simulation (thermalization: 600ns + production: 1000 ns). Distances are reported in Å, time is reported in ps. The distances between Mn^{2+} and the four ligands identified in the crystal structure Asp49, Asn52, Ala223 and Met226 are represented in orange, yellow, turquoise and magenta, respectively. The fluctuations in their bond length values are smaller with respect to that of the other molecules as these four ligands were kept restrained in their positions during the entire simulation. The four water molecules WAT56324, WAT1701, WAT24524 and WAT3067 are represented as blue, red, light blue and purple lines, respectively. Lastly, distances between Mn^{2+} and Gln389, Gly46 and Asp49($\text{O}^{\delta 1}$) are represented as green, pink and dark violet lines, respectively.

two water molecules. It is also important to notice that both models are not in agreement with the experimental data. In the reference crystal structure indeed Asp49 was observed to coordinate Mn^{2+} with only one of its carboxylic oxygens, Met226 was observed to coordinate Mn^{2+} and Gln389 and Gly46 were instead not observed to coordinate the metal ion. The Mn^{2+} coordination sphere models proposed by Pujol-Giménez et al. and by Bozzi et al. are shown in Fig. 3.12.

The unusual heptacoordinated structure we obtained is probably an artifact due to the limitation of the classical force field in describing the metal ion-water interactions. It suggests nevertheless that there seems to be a strong tendency for water coordination and an extension

of the coordination sphere beyond the 4 ligands detected in the crystal structure. To probe this further, we performed DFT based QM/MM simulations. We choose the last frame of classical MD (shown in the left panel of Fig. 3.6) as a representative configuration to start the subsequent QM/MM simulations. For this, the metal-ion, the Asp49, Asn52, Ala223 and Met226 side chains and the three water molecules in the Mn^{2+} coordination sphere (WAT56324 = WAT1, WAT1701 = WAT2 and WAT3067 = WAT3) were included in the QM region with a total charge equal to +1. A cubic cell with an edge length of 40.0 Å was used to ensure that the electron density of the QM atoms was completely contained inside the QM simulation box. To obtain the most representative structure from the QM/MM trajectory, we extracted the frame whose atomic coordinates were closest to the average coordinates calculated over the whole QM/MM MD trajectory (shown in the right panel of Fig. 3.6).

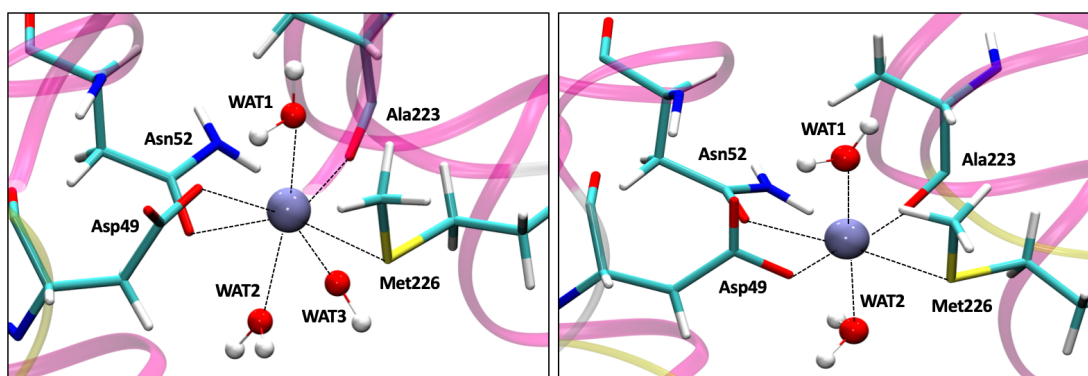


Figure 3.6: Snapshot of the Mn^{2+} coordination sphere as obtained from the classical MD simulations (**left**) and from QM/MM MD simulations (**right**). In both cases ScaDMT α -helices are depicted as magenta ribbons. Asp49, Asn52, Ala223 and Met226 are represented as sticks, water molecules as balls and sticks model while Mn^{2+} is represented as a blue sphere. C atoms are colored in cyan, O in red, N in blue, S in yellow and H in white.

In a first equilibration phase of the QM/MM simulations the system has been gradually thermalized to a temperature of 310 K by *ab initio* Born-Oppenheimer molecular dynamics (BOMD) (38.7 ps). During this step we observed a gradual yet remarkable rearrangement of the four amino acid residues coordinating Mn^{2+} . In particular, we noticed that the four amino acid residues reorganize compared to the crystal structure to have the four coordinating atoms (Asp49: $\text{O}^{\delta 2}$, Asn52: $\text{O}^{\delta 1}$, Ala223: O and Met226: S^{δ}) arranged in a nearly perfect planar configuration. Simultaneously with the rearrangement of the four amino acid residues, WAT3 leaves the Mn^{2+} coordination sphere thus leading the system into a stable and geometrically favorable octahedral coordination geometry for the Mn^{2+} active site (Fig. 3.6, right panel).

This coordination pattern was retained during the entire 29 ps of *ab initio* Car-Parrinello molecular dynamics (CPMD) simulation performed after the initial BOMD (as shown by the time evolution of characteristic distances in Fig. 3.7). The horizontal lines represent the average values for every bond length, reported also in Table 1a. Mn^{2+} -Asp49($\text{O}^{\delta 2}$) (orange) and Mn^{2+} -Asn52($\text{O}^{\delta 1}$) (yellow) bonds fluctuate in a very narrow range of values proving that these two residues are always bound to Mn^{2+} , with average bond lengths values of 2.07 ± 0.09 and 2.13 ± 0.10 Å, respectively. Mn^{2+} -Ala223(O) (turquoise) and Mn^{2+} -Met226(S^{δ}) (magenta)

bonds fluctuate over wider ranges of values but they never leave the Mn^{2+} coordination sphere, as highlighted by average distances of 2.46 ± 0.40 and 2.76 ± 0.25 Å, respectively. Fig. 3.7 also highlights the constant presence of WAT1 and WAT2 in the two axial coordination sites: Mn^{2+} -WAT1 (blue) and Mn^{2+} -WAT2 (red) bonds oscillate in narrow ranges of values with average bond lengths of 2.27 ± 0.14 and 2.35 ± 0.19 Å, respectively. On the other side, the distance between Mn^{2+} and WAT3 (purple) never drops below 3.2 Å and reaches maximum values of around 5.6 Å, with an average value of 4.02 ± 0.38 Å. This suggests that this water molecule does not directly participate to the first coordination shell of Mn^{2+} .

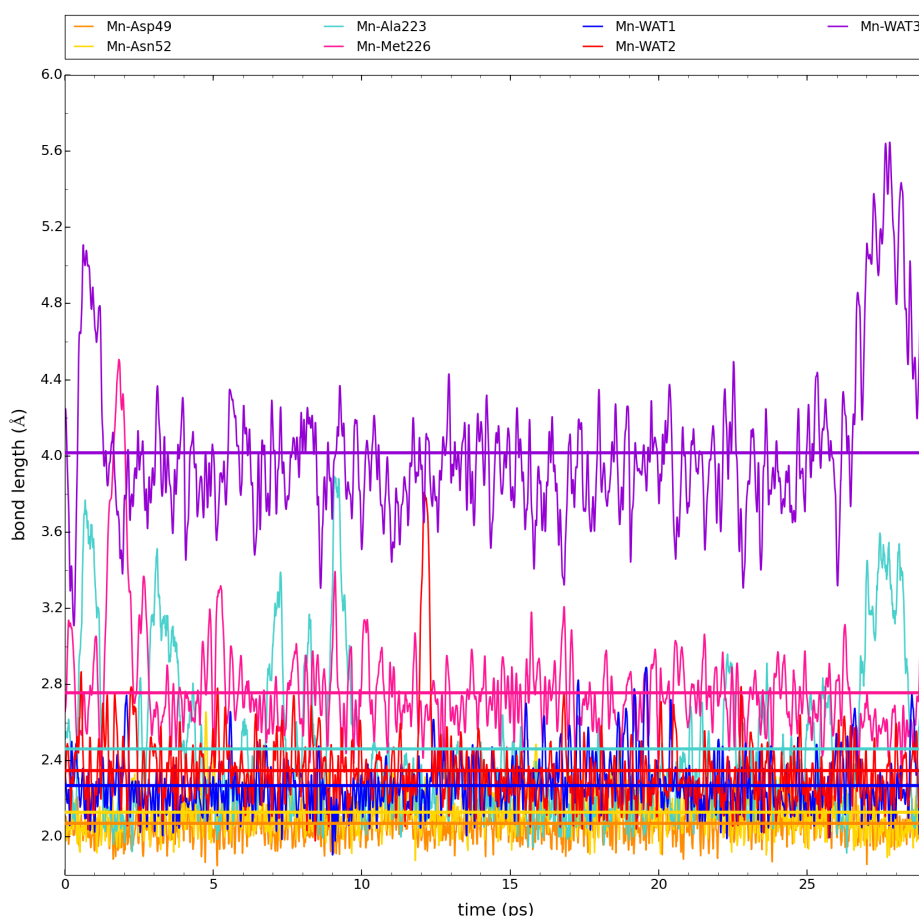


Figure 3.7: Fluctuations of the bond lengths for each of the six Mn^{2+} -coordinating ligands: Mn^{2+} -Asp49($\text{O}^{\delta 2}$) (orange), Mn^{2+} -Asn52($\text{O}^{\delta 1}$) (yellow), Mn^{2+} -Ala223(O) (turquoise), Mn^{2+} -Met226(S^{δ}) (magenta), Mn^{2+} -WAT1 (blue) and Mn^{2+} -WAT2 (red). Mn^{2+} -WAT3 distance is also represented in purple. The straight lines in the same color of the corresponding curve represent the average values during the whole 29 ps long CPMD QM/MM trajectory.

In Tables 3.1 (top panel) and 3.2 experimental and average QM/MM values of bond lengths and angles formed between Mn^{2+} and its ligands are presented, respectively. The obtained theoretical values of bond distances and angles are averaged over the whole QM/MM trajectory

and are presented together with their standard deviations. In Table 3.1 (top panel), the average value for the distance between Mn^{2+} and the non-coordinating WAT3 water molecule is also reported.

The results of the QM/MM simulation confirm the discrepancies already highlighted between our model and those presented by Pujol-Gimenéz [105] and Bozzi [49]. Although the former is in line with the prediction of an octahedral coordination geometry, it predicts different ligands (with Asp49 acting as bidentate ligand instead of Met226(S^δ) and Gln389($\text{O}^{\epsilon 1}$) and Gly46(O) as axial ligands instead of water molecules) coordinating the metal. On the other hand, Bozzi et al. propose a binding site model in which both Met226(S^δ) and two water molecules coordinate Mn^{2+} , in agreement with our model, but at the same time they propose Asp49 to coordinate the metal ion as a bidentate chelating ligand, thus suggesting a heptacoordinated Mn^{2+} .

It is important to notice that these predictions disagree with the experimental data, that indicates a coordination pattern more similar to our findings. All these observations support our results. In fact, while crystallographic experiments may miss the identification of water molecules, especially for difficult crystallization conditions such those required for transmembrane proteins and in the presence of strongly scattering metal centers, they are quite accurate in the characterization of amino acidic residues.

To additionally validate our results, we traced Mn^{2+} -Gln389($\text{O}^{\epsilon 1}$), Mn^{2+} -Gly46(O) and Mn^{2+} -Asp49($\text{O}^{\delta 1}$) distances during the whole QM/MM CPMD trajectory (Fig. 3.8). The Mn^{2+} -Asp49($\text{O}^{\delta 1}$) distance (dark violet) oscillates in a range of ~ 3.1 to 4.4 Å, with an average value of 3.84 ± 0.22 Å (Table 3.1, bottom panel). Mn^{2+} -Gln389($\text{O}^{\epsilon 1}$) (green) and Mn^{2+} -Gly46(O) (pink) distances are even larger with average values of 4.72 ± 0.18 and 5.14 ± 0.27 Å, respectively (Table 3.1, bottom panel). These values confirm that these three atoms do not participate in the first coordination shell of Mn^{2+} , as already suggested by the experimental crystallographic structure.

From this analysis, we are quite confident to conclude that the coordination present in the active site of the ScaDMT protein is octahedral with the Asp49($\text{O}^{\delta 2}$), Asn52($\text{O}^{\delta 1}$), Ala223(O), Met226(S^δ), WAT1(O) and WAT2(O) atoms participating in the direct interaction with Mn^{2+} .

To further characterize the proposed structure of the Mn^{2+} coordination sphere, we also performed an analysis of the hydrogen bond network linking the first and the second coordination shells during the QM/MM MD simulations. In particular, we found that most of the hydrogen-bonding interactions existing between the ligands that interact directly with Mn^{2+} and the molecules of the second coordination shell persist during the main part of the QM/MM MD trajectory of 29ps. This is evident for instance for the hydrogen bonds connecting WAT1 to Gln389($\text{O}^{\epsilon 1}$) and WAT41196(O) which persist respectively for 91.6% and 68.5% of the total time (see Fig. 3.9a). WAT41196 is substituted for a certain period of time by Asp49, as shown in Fig. 3.9b. In total WAT1 stays bound either to WAT41196 or Asp49 for 82.3% of the whole trajectory.

WAT2 is also involved in the formation of hydrogen bonds. In particular, one of its hydrogen

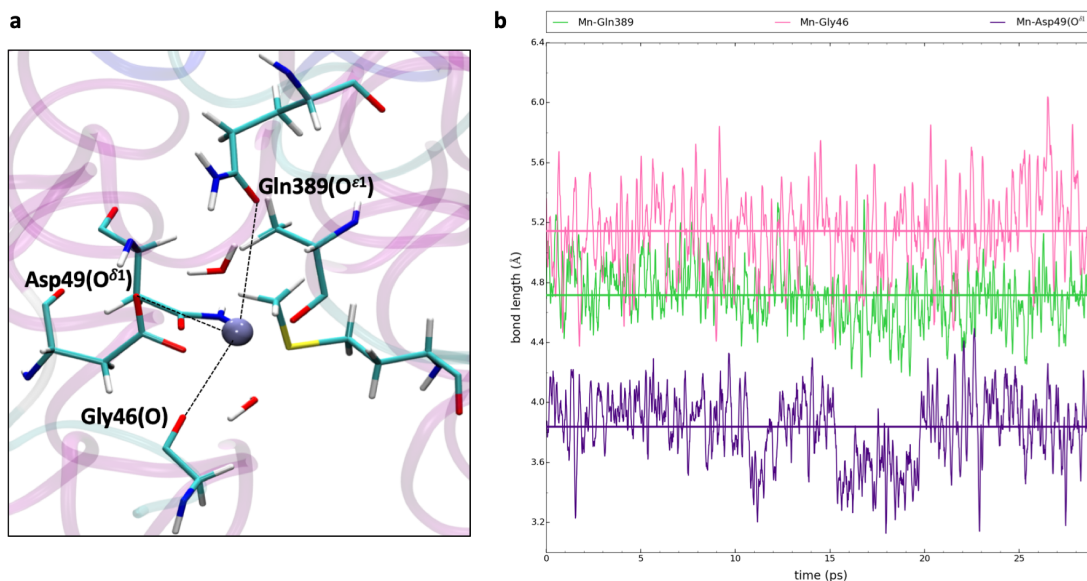


Figure 3.8: (a) Expanded view of the region around the protein active site. The position of the three atoms Gln389(O^{ε1}), Gly46(O) and Asp49(O^{δ1}) with respect to the Mn²⁺ ion is highlighted by the labels and the black dashed lines. (b) Oscillations of Mn²⁺-Asp49(O^{δ1}) (dark violet), Mn²⁺-Gln389(O^{ε1}) (green) and Mn²⁺-Gly46(O) (pink) distances over the whole QM/MM CPMD trajectory (29 ps). Horizontal lines represent the average values.

BOND	EXPERIMENTAL	QM/MM
Mn(II)-Asp49(O ^{δ2})	2.26	2.07 ± 0.09
Mn(II)-Asn52(O ^{δ1})	2.90	2.13 ± 0.10
Mn(II)-Ala223(O)	2.10	2.46 ± 0.40
Mn(II)-Met226(S ^δ)	2.66	2.76 ± 0.25
Mn(II)-WAT1(O)		2.27 ± 0.14
Mn(II)-WAT2(O)		2.35 ± 0.19
Mn(II)-WAT3(O)		4.02 ± 0.38
Mn(II)-Gln389(O ^{ε1})	4.49	4.72 ± 0.18
Mn(II)-Gly46(O)	5.59	5.14 ± 0.27
Mn(II)-Asp49(O ^{δ1})	4.27	3.84 ± 0.22

Table 3.1: (top) Experimental and average QM/MM values of bond lengths for each bond formed by Mn²⁺ and its six ligands. The average distance between Mn²⁺ and the non-coordinating water molecule WAT3 is also reported. (bottom) Experimental and average QM/MM values for the distances between Mn²⁺ and Gln389(O^{ε1}), Gly46(O) and Asp49(O^{δ1}) atoms. QM/MM values are presented in Å, together with their standard deviation errors. Furthermore, each bond is indicated in the same color with which it appears in Fig. 3.7 and in Fig. 3.8b, respectively.

atoms is bonded to the Val45 carbonyl oxygen (90.6% of the total time) while the other interacts with either WAT5 (71.3%), WAT9704 (6.0%) or WAT4458 (10.1%) (Fig. 3.10). In total WAT2 remains bonded to a water molecule for 87.4% of the simulation. The formation of long-lasting hydrogen bonds is a further confirmation of the crucial role played by WAT1 and WAT2 in the Mn²⁺ coordination sphere. It is important to highlight that at difference from what has been predicted by Pujol-Gimenéz and co-workers in ref. [105], in our simulations Gln389 seems to coordinate WAT1 (91.6% of the total time) instead of directly interacting

ANGLE	EXPERIMENTAL	QM/MM
Asp49(O ^{δ2})-Mn(II)-Asn52(O ^{δ1})	72.8°	95.0° ± 7.4°
Asn52(O ^{δ1})-Mn(II)-Ala223(O)	95.2°	88.2° ± 7.3°
Ala223(O)-Mn(II)-Met226(S ^δ)	101.0°	87.4° ± 6.3°
Met226(S ^δ)-Mn(II)-Asp49(O ^{δ2})	108.5°	89.5° ± 6.3°
WAT1(O)-Mn(II)-WAT2(O)		166.8° ± 7.0°
Asp49(O ^{δ2})-Mn(II)-WAT1(O)		80.8° ± 5.9°
Asn52(O ^{δ1})-Mn(II)-WAT1(O)		93.3° ± 9.1°
Ala223(O)-Mn(II)-WAT1(O)		93.2° ± 7.3°
Met226(S ^δ)-Mn(II)-WAT1(O)		91.1° ± 8.0°
Asp49(O ^{δ2})-Mn(II)-WAT2(O)		97.1° ± 9.6°
Asn52(O ^{δ1})-Mn(II)-WAT2(O)		91.7° ± 6.7°
Ala223(O)-Mn(II)-WAT2(O)		88.5° ± 7.3°
Met226(S ^δ)-Mn(II)-WAT2(O)		82.9° ± 6.4°

Table 3.2: Experimental and average QM/MM values of angles formed between Mn²⁺ and its six ligands (in degrees). The QM/MM values are presented together with their standard deviations.

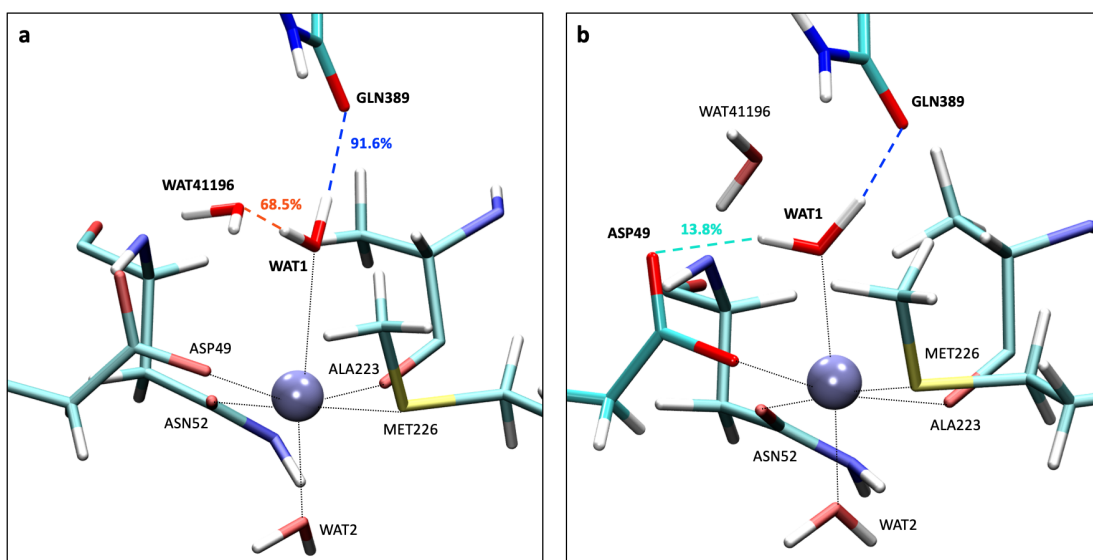


Figure 3.9: (a) Representation of the hydrogen bonds linking WAT1 hydrogens with Gln389(O^{ε1}) and WAT41196(O). Hydrogen bonds are indicated as blue and orange dashed lines, respectively. (b) Representation of the hydrogen bond between WAT1 and one of the Asp49 carboxylic oxygens as a turquoise dashed line. The percentages of persistence of each of these hydrogen bonds are also given in the same color. Mn²⁺ is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style.

with Mn²⁺. In the same way, also Gly46, expected to directly bind Mn²⁺ in ref. [105], seems to interact with Asn272 and WAT5 for respectively 93.1% and 58.5% of the total simulation. These results are shown in Fig. 3.11. Previous studies (ref. [105] and [49]) also proposed that Asp49 might coordinate the Mn²⁺ ion in a bidentate configuration. This situation never occurs in our simulations in which the Asp49 residue appears to be coordinated via O^{δ1} to either Gly51 (46.5%), WAT41196 (58.3%) or WAT1319 (34.6%) during the whole QM/MM trajectory,

as represented in Fig. 3.11.

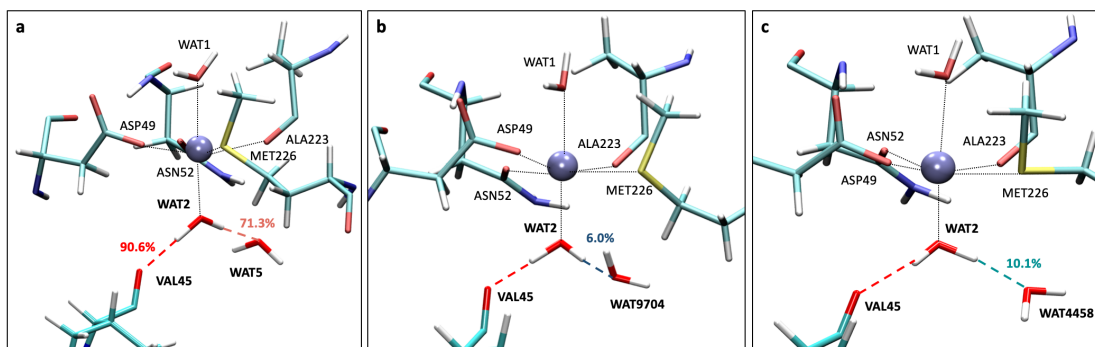


Figure 3.10: (a) (b) (c) Representation of the hydrogen bond linking WAT2 and Val45(O) as a red dashed line with the persistence percentage expressed in the same color. Representation of the hydrogen bonds linking WAT2 and either WAT5 (a), WAT9704 (b) or WAT4458 (c) as coral, steel blue and light sea green dashed lines respectively, together with their persistence percentages in the same color. Mn^{2+} is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style.

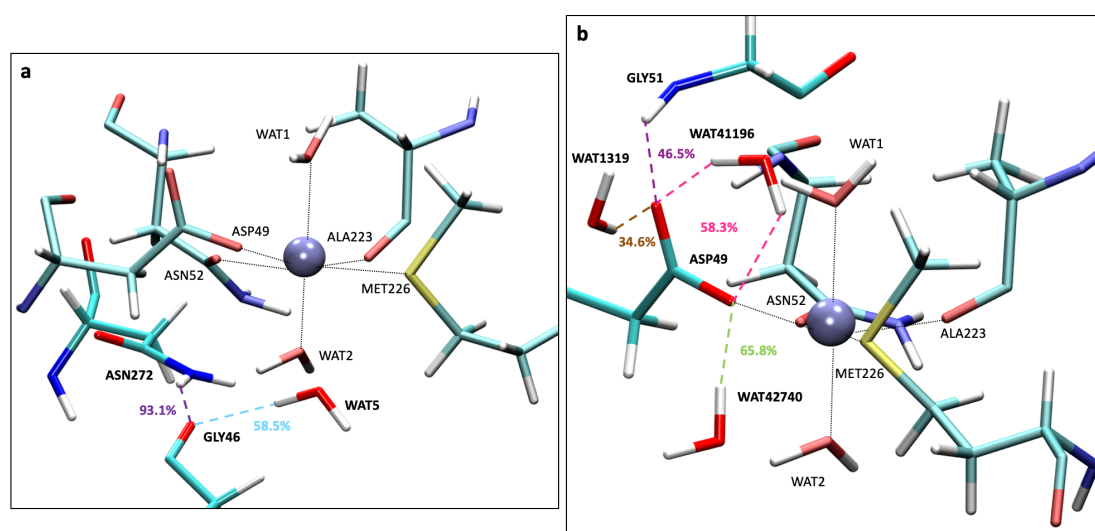


Figure 3.11: (a) Representation of the hydrogen bonds connecting Gly46 to Asn272 and WAT5 as violet and light blue dashed lines, respectively. (b) Representation of the hydrogen bonds formed by the two carboxylic oxygens of Asp49. In particular, Asp49($\text{O}^{\delta 2}$) forms hydrogen bonds with either WAT41196 or WAT42740 (magenta and green dashed lines respectively) while Asp49($\text{O}^{\delta 1}$) forms hydrogen bonds with either WAT41196, WAT1319 or the Gly51 residue (magenta, brown and purple dashed lines, respectively). The persistence percentage of each hydrogen bond is given in the correspondent color. Mn^{2+} is represented as a gray sphere whereas all other molecules and amino acidic residues are represented in licorice style.

3.5 Conclusions

In this work, the ScaDMT protein in complex with a Mn^{2+} ion has been investigated by a combination of classical and first-principles QM/MM MD simulations. The obtained results suggest an octahedral coordination geometry for Mn^{2+} . The four amino acidic residues pre-

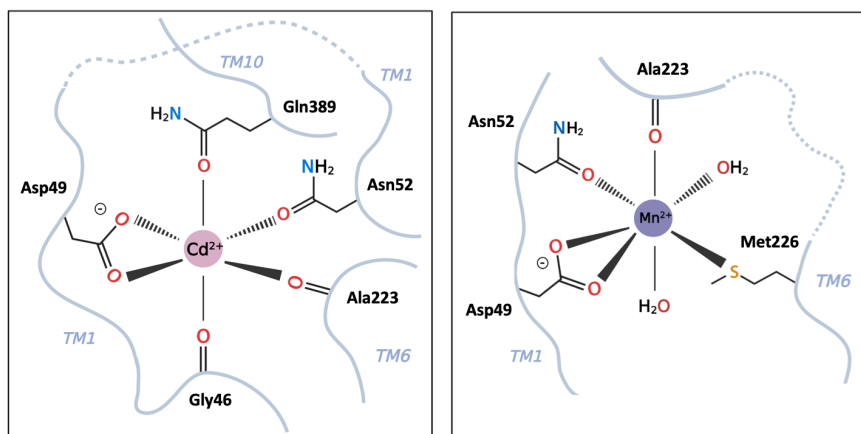


Figure 3.12: Schematic depiction of the Mn^{2+} coordination sphere as described by Pujol-Giménez and co-authors [105] (**left**) and Bozzi and co-authors (**right**) [49]. In the former, the Cd^{2+} metal ion has an octahedral coordination with the four residues observed in the X-ray structure (namely Asp49, Asn52, Ala223 and Met226) plus the side chain of Gln389 and the backbone carbonyl oxygen of Gly46 acting as ligands. In the latter instead, besides the four residues observed in the crystal structure the metal ion is coordinated by two water molecules. However, Asp49 acts as a bidentate ligand imparting therefore a heptahedral coordination. The transmembrane α -helices are represented in light gray, the Cd^{2+} and Mn^{2+} ions are represented as pink and gray spheres while N, O, and S atoms are represented in blue, red and orange, respectively.

dicted to coordinate Mn^{2+} in the crystallographic structure (Asp49, Asn52, Ala223 and Met226) are confirmed by our findings and occupy four equatorial coordination sites. The two axial sites left empty in the experimental structure are filled with two water molecules that never leave the first coordination shell during the whole QM/MM trajectory.

At difference to previously proposed coordination geometries [49, 105], our model is in agreement with the experimental crystallographic data. Our simulations are not only able to reproduce the experimental findings but they could also propose refinements for the coordination sites missing in the crystallographic measurements and readjust the small displacements in the location of the coordinating ligands present in the 3.40 Å resolution data to enable a nearly perfect octahedral geometry. Both classical and *ab initio* QM/MM MD indicate that Gln389 and Gly46 do not participate to the coordination of the Mn^{2+} ion, in disagreement with previous theoretical calculations [105]. Our simulations also exclude a possible coordination of Asp49 as a bidentate chelating ligand. These evidences are further strengthened by the analysis of the hydrogen bond network connecting the first and the second coordination shell during the QM/MM MD. We find indeed that Gln389 and Gly46 and Asp49($\text{O}^{\delta 1}$) are involved in the formation of hydrogen bonds that last for the main part of the dynamics. Furthermore, the two water molecules coordinating Mn^{2+} are also involved in the formation of permanent hydrogen bonds, further establishing their high stability in the coordination shell.

Notice that the coordination geometry proposed by Bozzi and colleagues (Fig. 3.12, right panel) is merely a hypothetical model based on the coordination sphere observed for the crystal structure of DraDMT in the substrate-bound OFC. On the other hand, the structure proposed by Pujol-Giménez et al. (left panel of Fig. 3.12) has been obtained through computa-

tional studies at the classical force field level. The theoretical model obtained in this way is clearly different from the one we obtained here using a more sophisticated QM/MM approach. However, this does not explain why we did not observe the same coordination pattern during the classical MD simulation. A possible explanation to this could be the fact that they used Cd^{2+} instead of Mn^{2+} in their simulations. The different size and electronic properties/parameterization of the investigated metal ion could affect the nature of the interactions and, therefore, the composition of its coordination sphere. For this reason, the application of the same computational protocol used here to other metal ions would be important to understand whether the predicted binding mode is conserved for all metal ions or the protein binding site is capable to adapt itself according to the nature of the substrate.

The identification of the complete composition of the Mn^{2+} coordination sphere of ScaDMT represents a key piece in the puzzle about the mechanistic details regarding the proton-coupled divalent transition-metal ion transport mechanism of SLC11/NRAMP transporters. This information could serve in fact as starting point for the study of the mechanism of metal ion release and the subsequent conformational changes. In this regard, it could be possible that the presence of water molecules in the coordination sphere of the metal ion could facilitate its release into the cellular cytoplasm and the associated conformational switch from the OFC to the IFC. In these conditions in fact water can solvate the metal ion inducing the transport. On the contrary, the presence of solely amino acidic residues in the coordination sphere of the metal ion could make the solvation process more difficult hindering therefore the release of the substrate. The application of enhanced sampling techniques might help addressing this issue.

Since ScaDMT represents the only currently available crystal structure of a Nramp/SLC11 transporter in a substrate-bound IFC, the application of the same computational procedure to another member of the family in the same conformational state is not possible. However, because of the high degree of sequence conservation between the members of the family (percentages already reported in section 3.2) the binding mode can hopefully be generalised to other transporters of the family.

References

- [1] D. Radisky and J. Kaplan. "Regulation of Transition Metal Transport across the Yeast Plasma Membrane". In: *J. Biol. Chem.* 274.8 (1999), pp. 4481–4484. DOI: <https://doi.org/10.1074/jbc.274.8.4481>.
- [2] N. Nelson. "Metal ion transporters and homeostasis". In: *EMBOJ.* 18.16 (1999), pp. 4361–4371. DOI: <https://doi.org/10.1093/emboj/18.16.4361>.
- [3] A. Rolfs and M. A. Hediger. "Metal ion transporters in mammals: structure, function and pathological implications". In: *J. Physiol.* 518.1 (1999), pp. 1–12. DOI: <https://doi.org/10.1111/j.1469-7793.1999.0001r.x>.
- [4] L. Mezzaroba et al. "The role of zinc, copper, manganese and iron in neurodegenerative diseases". In: *NeuroToxicol.* 74 (2019), pp. 230–241. DOI: <https://doi.org/10.1016/j.neuro.2019.07.007>.
- [5] K. Jomova et al. "Metals, oxidative stress and neurodegenerative disorders". In: *Mol. Cell. Biochem.* 345 (2010), pp. 91–104. DOI: <https://doi.org/10.1007/s11010-010-0563-x>.
- [6] W. Li et al. "Enhanced Expression of Natural Resistance-Associated Macrophage Protein 1 in Atherosclerotic Lesions May Be Associated with Oxidized Lipid-Induced Apoptosis". In: *Ann. New York Acad. Sci.* 1030.1 (2004), pp. 202–207. DOI: <https://doi.org/10.1196/annals.1329.026>.
- [7] M. Farina et al. "Metals, oxidative stress and neurodegeneration: A focus on iron, manganese and mercury". In: *Neurochem. Int.* 62.5 (2013), pp. 575–594. DOI: <https://doi.org/10.1016/j.neuint.2012.12.006>.
- [8] T. Moos and E. H. Morgan. "The Metabolism of Neuronal Iron and Its Pathogenic Role in Neurological Disease: Review". In: *Ann. New York Acad. Sci.* 1012.1 (2004), pp. 14–26. DOI: <https://doi.org/10.1196/annals.1306.002>.
- [9] P. Zatta et al. "The role of metals in neurodegenerative processes: aluminum, manganese, and zinc". In: *Brain Res. Bull.* 62.1 (2003), pp. 15–28. DOI: [https://doi.org/10.1016/S0361-9230\(03\)00182-5](https://doi.org/10.1016/S0361-9230(03)00182-5).
- [10] A. B. Bowman et al. "Role of manganese in neurodegenerative diseases". In: *J. Trace Elem. Med. Biol.* 25.4 (2011), pp. 191–203. DOI: <https://doi.org/10.1016/j.jtemb.2011.08.144>.
- [11] G. F. Kwakye et al. "Manganese-Induced Parkinsonism and Parkinson's Disease: Shared and Distinguishable Features". In: *Int. J. Environ. Res. Public Health* 12.7 (2015), pp. 7519–7540. DOI: [10.3390/ijerph120707519](https://doi.org/10.3390/ijerph120707519).
- [12] Y. Nevo and N. Nelson. "The NRAMP family of metal-ion transporters". In: *Biochim. Biophys. Acta* 1763.7 (2006), pp. 609–620. DOI: <https://doi.org/10.1016/j.bbamcr.2006.05.007>.

- [13] N. Montalbetti et al. “Mammalian iron transporters: Families SLC11 and SLC40”. In: *Mol. Asp. Med.* 34.2 (2013), pp. 270–287. DOI: <https://doi.org/10.1016/j.mam.2013.01.002>.
- [14] M. F. M. Cellier et al. “Polyphyletic origins of bacterial Nramp transporters”. In: *Trends Gen.* 17.7 (2001), pp. 365–370. DOI: [https://doi.org/10.1016/S0168-9525\(01\)02364-2](https://doi.org/10.1016/S0168-9525(01)02364-2).
- [15] M. Cellier et al. “Nramp defines a family of membrane proteins.” In: *Proc. Natl. Acad. Sci.* 92.22 (1995), pp. 10089–10093. DOI: [10.1073/pnas.92.22.10089](https://doi.org/10.1073/pnas.92.22.10089).
- [16] B. Mackenzie et al. “Divalent metal-ion transporter DMT1 mediates both H⁺-coupled Fe²⁺ transport and uncoupled fluxes”. In: *Pflügers Arch.* 451.4 (2006), pp. 544–558. DOI: <https://doi.org/10.1007/s00424-005-1494-3>.
- [17] A. Shawki and B. Mackenzie. “Interaction of calcium with the human divalent metal-ion transporter-1”. In: *Biochem. Biophys. Res. Commun.* 393.3 (2010), pp. 471–475. DOI: <https://doi.org/10.1016/j.bbrc.2010.02.025>.
- [18] S. M. Vidal et al. “Natural resistance to infection with intracellular parasites: Isolation of a candidate for Bcg”. In: *Cell* 73.3 (1993), pp. 469–485. DOI: [https://doi.org/10.1016/0092-8674\(93\)90135-D](https://doi.org/10.1016/0092-8674(93)90135-D).
- [19] M. Wessling-Resnick. “Nramp1 and Other Transporters Involved in Metal Withholding during Infection”. In: *J. Biol. Chem.* 290.31 (2015), pp. 18984–18990. DOI: <https://doi.org/10.1074/jbc.R115.643973>.
- [20] M. F. Cellier, P. Courville, and C. Champion. “Nramp1 phagocyte intracellular metal withdrawal defense”. In: *Microb. Infect.* 9.14 (2007), pp. 1662–1670. DOI: <https://doi.org/10.1016/j.micinf.2007.09.006>.
- [21] E. E. Johnson and M. Wessling-Resnick. “Iron metabolism and the innate immune response to infection”. In: *Microbes and Infection* 14.3 (2012), pp. 207–216. DOI: <https://doi.org/10.1016/j.micinf.2011.10.001>.
- [22] F. Canonne-Hergaux et al. “The Nramp2/DMT1 iron transporter is induced in the duodenum of microcytic anemia mk mice but is not properly targeted to the intestinal brush border”. In: *Blood* 96.12 (2000), pp. 3964–3970. DOI: <https://doi.org/10.1182/blood.V96.12.3964>.
- [23] F. Canonne-Hergaux et al. “Cellular and Subcellular Localization of the Nramp2 Iron Transporter in the Intestinal Brush Border and Regulation by Dietary Iron”. In: *Blood* 93.12 (1999), pp. 4406–4417. DOI: <https://doi.org/10.1182/blood.V93.12.4406>.
- [24] A. Donovan et al. “Positional cloning of zebrafish *ferroportin1* identifies a conserved vertebrate iron exporter”. In: *Nature* 403.6771 (2000), pp. 776–781. DOI: <https://doi.org/10.1038/35001596>.
- [25] K. Gkouvatsos, G. Papanikolaou, and K. Pantopoulos. “Regulation of iron transport and the role of transferrin”. In: *Biochim. Biophys. Acta* 1820.3 (2012), pp. 188–202. DOI: <https://doi.org/10.1016/j.bbagen.2011.10.013>.

- [26] A. Shawki et al. "Chapter Five - H^+ -Coupled Divalent Metal-Ion Transporter-1: Functional Properties, Physiological Roles and Therapeutics". In: *Co-Transport Systems*. Vol. 70. Curr. Top. Membr. Academic Press, 2012, pp. 169–214. DOI: <https://doi.org/10.1016/B978-0-12-394316-3.00005-3>.
- [27] N. C. Andrews. "The iron transporter DMT1". In: *Int. J. Biochem. Cell Biol.* 31.10 (1999), pp. 991–994. DOI: [https://doi.org/10.1016/S1357-2725\(99\)00065-5](https://doi.org/10.1016/S1357-2725(99)00065-5).
- [28] V. Picard et al. "Nramp 2 (DCT1/DMT1) Expressed at the Plasma Membrane Transports Iron and Other Divalent Cations into a Calcein-accessible Cytoplasmic Pool". In: *J. Biol. Chem.* 275.46 (2000), pp. 35738–35745. DOI: <https://doi.org/10.1074/jbc.M005387200>.
- [29] A. C. Illing et al. "Substrate Profile and Metal-ion Selectivity of Human Divalent Metal-ion Transporter-1". In: *J. Biol. Chem.* 287.36 (2012), pp. 30485–30496. DOI: <https://doi.org/10.1074/jbc.M112.364208>.
- [30] H. Gunshin et al. "Cloning and characterization of a mammalian proton-coupled metal-ion transporter". In: *Nature* 388 (1997), pp. 482–488. DOI: <https://doi.org/10.1038/41343>.
- [31] M. Knöpfel, L. Zhao, and M. D. Garrick. "Transport of Divalent Transition-Metal Ions Is Lost in Small-Intestinal Tissue of *b/b* Belgrade Rats". In: *Biochem.* 44.9 (2005), pp. 3454–3465. DOI: 10.1021/bi048768+.
- [32] E. Bardou-Jacquet et al. "A novel N491S mutation in the human *SLC11A2* gene impairs protein trafficking and in association with the G212V mutation leads to microcytic anemia and liver iron overload". In: *Blood Cells Mol. Dis.* 47.4 (2011), pp. 243–248. DOI: <https://doi.org/10.1016/j.bcmd.2011.07.004>.
- [33] C. Beaumont et al. "Two new human *DMT1* gene mutations in a patient with microcytic anemia, low ferritinemia, and liver iron overload". In: *Blood* 107.10 (2006), pp. 4168–4170. DOI: <https://doi.org/10.1182/blood-2005-10-4269>.
- [34] S. Lam-Yuk-Tseung et al. "A novel R416C mutation in human DMT1 (*SLC11A2*) displays pleiotropic effects on function and causes microcytic anemia and hepatic iron overload". In: *Blood Cells Mol. Dis.* 36.3 (2006), pp. 347–354. DOI: <https://doi.org/10.1016/j.bcmd.2006.01.011>.
- [35] A. Iolascon et al. "Microcytic anemia and hepatic iron overload in a child with compound heterozygous mutations in DMT1 (*SCL11A2*)". In: *Blood* 107.1 (2006), pp. 349–354. DOI: <https://doi.org/10.1182/blood-2005-06-2477>.
- [36] M. D. Fleming et al. "*Nramp2* is mutated in the anemic Belgrade (*b*) rat: Evidence of a role for *Nramp2* in endosomal iron transport". In: *Proc. Natl. Acad. Sci.* 95.3 (1998), pp. 1148–1153. DOI: 10.1073/pnas.95.3.1148.
- [37] M. D. Fleming et al. "Microcytic anaemia mice have a mutation in *Nramp2*, a candidate iron transporter gene". In: *Nat. gen.* 16.4 (1997), pp. 383–386. DOI: <https://doi.org/10.1038/ng0897-383>.

- [38] F. Canonne-Hergaux et al. "Characterization of the iron transporter DMT1 (NRAMP2/DCT1) in red blood cells of normal and anemic *mk/mk* mice". In: *Blood* 98.13 (2001), pp. 3823–3830. DOI: <https://doi.org/10.1182/blood.V98.13.3823>.
- [39] M. A. Su et al. "The G185R Mutation Disrupts Function of the Iron Transporter Nramp2". In: *Blood* 92.6 (1998), pp. 2157–2163. DOI: <https://doi.org/10.1182/blood.V92.6.2157>.
- [40] E. Blanco et al. "Not all DMT1 mutations lead to iron overload". In: *Blood Cells Mol. Dis.* 43.2 (2009), pp. 199–201. DOI: <https://doi.org/10.1016/j.bcmd.2009.05.003>.
- [41] M. P. Mims et al. "Identification of a human mutation of *DMT1* in a patient with microcytic anemia and iron overload". In: *Blood* 105.3 (2005), pp. 1337–1342. DOI: <https://doi.org/10.1182/blood-2004-07-2966>.
- [42] O. Jardetzky. "Simple Allosteric Model for Membrane Pumps". In: *Nature* 211 (1966), pp. 969–970. DOI: <https://doi.org/10.1038/211969a0>.
- [43] L. R. Forrest, R. Krämer, and C. Ziegler. "The structural basis of secondary active transport mechanisms". In: *Biochim. Biophys. Acta - Bioenergetics* 1807.2 (2011), pp. 167–188. DOI: <https://doi.org/10.1016/j.bbabi.2010.10.014>.
- [44] L. R. Forrest et al. "Mechanism for alternating access in neurotransmitter transporters". In: *Proc. Natl. Acad. Sci.* 105.30 (2008), pp. 10338–10343. DOI: [10.1073/pnas.0804659105](https://doi.org/10.1073/pnas.0804659105).
- [45] D. Drew and O. Boudker. "Shared Molecular Mechanisms of Membrane Transporters". In: *Ann. Rev. Biochem.* 85.1 (2016), pp. 543–572. DOI: [10.1146/annurev-biochem-060815-014520](https://doi.org/10.1146/annurev-biochem-060815-014520).
- [46] I. A. Ehrnstorfer et al. "Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport". In: *Nat. Struct. Mol. Biol.* 21.11 (2014), pp. 990–996. DOI: [10.1038/nsmb.2904](https://doi.org/10.1038/nsmb.2904).
- [47] I. A. Ehrnstorfer et al. "Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family". In: *Nature Commun.* 8.1 (2017), p. 14033. DOI: [10.1038/ncomms14033](https://doi.org/10.1038/ncomms14033).
- [48] A. T. Bozzi et al. "Crystal Structure and Conformational Change Mechanism of a Bacterial Nramp-Family Divalent Metal Transporter". In: *Structure* 24.12 (2016), pp. 2102–2114. DOI: <https://doi.org/10.1016/j.str.2016.09.017>.
- [49] A. T. Bozzi et al. "Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter". In: *eLife* 8 (2019), e41124. DOI: [10.7554/eLife.41124](https://doi.org/10.7554/eLife.41124).
- [50] S. Schulze et al. "Structural basis of Na⁺-independent and cooperative substrate/product antiport in CaiT". In: *Nature* 467 (2010), pp. 233–236. DOI: <https://doi.org/10.1038/nature09310>.
- [51] S. Ressler et al. "Molecular basis of transport and regulation in the Na⁺/betaine symporter BetP". In: *Nature* 458 (2009), pp. 47–52. DOI: <https://doi.org/10.1038/nature07819>.

- [52] S. Weyand et al. "Structure and Molecular Mechanism of a Nucleobase-Cation-Symport-1 Family Transporter". In: *Science* 322.5902 (2008), pp. 709–713. DOI: 10.1126/science.1164440.
- [53] S. Faham et al. "The Crystal Structure of a Sodium Galactose Transporter Reveals Mechanistic Insights into Na^+ /Sugar Symport". In: *Science* 321.5890 (2008), pp. 810–814. DOI: 10.1126/science.1160406.
- [54] X. Gao et al. "Structure and Mechanism of an Amino Acid Antiporter". In: *Science* 324.5934 (2009), pp. 1565–1568. DOI: 10.1126/science.1173654.
- [55] Y. Fang et al. "Structure of a prokaryotic virtual proton pump at 3.2 Å resolution". In: *Nature* 460 (2009), pp. 1040–1043. DOI: <https://doi.org/10.1038/nature08201>.
- [56] A. Yamashita et al. "Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters". In: *Nature* 437.7056 (2005), pp. 215–223. DOI: <https://doi.org/10.1038/nature03978>.
- [57] O. C. Gagné and F. C. Hawthorne. "Bond-length distributions for ions bonded to oxygen: results for the transition metals and quantification of the factors underlying bond-length variation in inorganic solids". In: *IUCrJ* 7.4 (2020), pp. 581–629. DOI: 10.1107/S2052252520005928.
- [58] P. W. Atkins et al.
- [59] M. N. Collomb and A. Deronzier. "Manganese: Inorganic & Coordination Chemistry Based in part on the article Manganese: Inorganic & Coordination Chemistry by Charles A. McAuliffe, Stephen M. Godfrey, & Michael Watkinson which appeared in the *Encyclopedia of Inorganic Chemistry, First Edition*." In: *Encyclopedia of Inorganic Chemistry*. John Wiley & Sons, Ltd, 2006. ISBN: 9780470862100. DOI: <https://doi.org/10.1002/0470862106.ia126>.
- [60] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [61] *VMD Visual Molecular Dynamics*. URL: <https://www.ks.uiuc.edu/Research/vmd/>.
- [62] H. M. Berman et al. "The Protein Data Bank". In: *Nucl. Acids Res.* 28.1 (2000), pp. 235–242. DOI: <https://doi.org/10.1093/nar/28.1.235>.
- [63] H. Berman, K. Henrick, and H. Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nat. Struct. Mol. Biol.* 10 (2003), p. 980. DOI: <https://doi.org/10.1038/nsb1203-980>.
- [64] *RCSB, PDB Protein Data Bank*. URL: <https://www.rcsb.org/>.
- [65] J. C. Gordon et al. "H++: a server for estimating pKa s and adding missing hydrogens to macromolecules". In: *Nucl. Acids Res.* 33.2 (2005), W368–W371. DOI: <https://doi.org/10.1093/nar/gki464>.
- [66] *H++*. URL: <http://newbiophysics.cs.vt.edu/H++/>.

- [67] M. A. Lomize et al. "OPM database and PPM web server: resources for positioning of proteins in membranes". In: *Nucl. Acids Res.* 40.D1 (2012), pp. D370–D376. DOI: <https://doi.org/10.1093/nar/gkr703>.
- [68] *Orientations of Proteins in Membranes (OPM) database*. URL: <https://opm.phar.umich.edu/>.
- [69] M. M. Ghahremanpour et al. "MemBuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the GROMACS biomolecular simulation program". In: *Bioinformatics* 30.3 (2014), pp. 439–441. DOI: <https://doi.org/10.1093/bioinformatics/btt680>.
- [70] *MemBuilder II*. URL: <http://bioinf.modares.ac.ir/software/mb2/builder.php>.
- [71] R. Cao et al. "Role of Extracellular Loops and Membrane Lipids for Ligand Recognition in the Neuronal Adenosine Receptor Type 2A: An Enhanced Sampling Simulation Study". In: *Molecules* 23.10 (2018), p. 2616. DOI: [10.3390/molecules23102616](https://doi.org/10.3390/molecules23102616).
- [72] T. E. Andreoli et al. *Physiology of Membrane Disorders*. Boston, MA: Springer, 1986. ISBN: 978-1-4613-2097-5. DOI: <https://doi.org/10.1007/978-1-4613-2097-5>.
- [73] M. G. Wolf et al. "g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation". In: *J. Comput. Chem.* 31.11 (2010), pp. 2169–2174. DOI: <https://doi.org/10.1002/jcc.21507>.
- [74] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation". In: *Comp. Phys. Commun.* 91.1 (1995), pp. 43–56. DOI: [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E).
- [75] D. van Der Spoel et al. "GROMACS: Fast, flexible, and free". In: *J. Comput. Chem.* 26.16 (2005), pp. 1701–1718. DOI: <https://doi.org/10.1002/jcc.20291>.
- [76] M. J. Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (2015), pp. 19–25. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>.
- [77] K. Lindorff-Larsen et al. "Improved side-chain torsion potentials for the Amber ff99SB protein force field". In: *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958. DOI: <https://doi.org/10.1002/prot.22711>.
- [78] C. J. Dickson et al. "Lipid14: The Amber Lipid Force Field". In: *J. Chem. Theory Comput.* 10.2 (2014), pp. 865–879. DOI: [10.1021/ct4010307](https://doi.org/10.1021/ct4010307).
- [79] W. L. Jorgensen et al. "Comparison of simple potential functions for simulating liquid water". In: *J. Chem. Phys.* 79.2 (1983), pp. 926–935. DOI: [10.1063/1.445869](https://doi.org/10.1063/1.445869).
- [80] I. S. Joung and T. E. Cheatham III. "Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations". In: *J. Phys. Chem. B* 112.30 (2008), pp. 9020–9041. DOI: [10.1021/jp8001614](https://doi.org/10.1021/jp8001614).
- [81] G. M. Bradbrook et al. "X-Ray and molecular dynamics studies of concanavalin-A glucoside and mannoside complexes Relating structure to thermodynamics of binding". In: *J. Chem. Soc. Faraday Trans.* 94 (11 1998), pp. 1603–1611. DOI: [10.1039/A800429C](https://doi.org/10.1039/A800429C).

- [82] Bryce Group: *Computational Biophysics and Drug Design*, AMBER parameter database. URL: <http://amber.manchester.ac.uk/>.
- [83] B. Hess et al. "LINCS: A linear constraint solver for molecular simulations". In: *J. Comput. Chem.* 18.12 (1997), pp. 1463–1472. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [84] H. G. Petersen. "Accuracy and efficiency of the particle mesh Ewald method". In: *J. Chem. Phys.* 103.9 (1995), pp. 3668–3679. DOI: 10.1063/1.470043.
- [85] T. Darden, D. York, and L. Pedersen. "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems". In: *J. Chem. Phys.* 98.12 (1993), pp. 10089–10092. DOI: 10.1063/1.464397.
- [86] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. DOI: 10.1063/1.328693.
- [87] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". In: *Phys. Rev. A* 31 (3 1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695.
- [88] S. Nosé. "A unified formulation of the constant temperature molecular dynamics methods". In: *J. Chem. Phys.* 81.1 (1984), pp. 511–519. DOI: 10.1063/1.447334.
- [89] S. Nosé. "A molecular dynamics method for simulations in the canonical ensemble". In: *Mol. Phys.* 52.2 (1984), pp. 255–268. DOI: 10.1080/00268978400101201.
- [90] A. Warshel and M. Levitt. "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme". In: *J. Mol. Biol.* 103.2 (1976), pp. 227–249. DOI: [https://doi.org/10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9).
- [91] A. Laio, J. VandeVondele, and U. Röthlisberger. "A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations". In: *J. Chem. Phys.* 116.16 (2002), pp. 6941–6947. DOI: 10.1063/1.1462041.
- [92] W. Andreoni and A. Curioni. "New advances in chemistry and materials science with CPMD and parallel computing". In: *Parallel Computing* 26.7 (2000), pp. 819–842. DOI: [https://doi.org/10.1016/S0167-8191\(00\)00014-4](https://doi.org/10.1016/S0167-8191(00)00014-4).
- [93] CPMD. URL: <https://www.cpmd.org/wordpress/>.
- [94] E. Brunk and U. Rothlisberger. "Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States". In: *Chem. Rev.* 115.12 (2015), pp. 6217–6263. DOI: <https://doi.org/10.1021/cr500628b>.
- [95] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Oxford: Oxford University Press, 1989. DOI: 10.1093/oso/9780195092769.001.0001.
- [96] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Phys. Rev.* 136 (3B 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.

- [97] W. Kohn and L. J. Sham. “Self-Consistent Equations Including Exchange and Correlation Effects”. In: *Phys. Rev.* 140 (4A 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.
- [98] A. D. Becke. “Density-functional exchange-energy approximation with correct asymptotic behavior”. In: *Phys. Rev. A* 38 (6 1988), pp. 3098–3100. DOI: 10.1103/PhysRevA.38.3098.
- [99] C. Lee, W. Yang, and R. G. Parr. “Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density”. In: *Phys. Rev. B* 37 (2 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.
- [100] N. Troullier and J. L. Martins. “Efficient pseudopotentials for plane-wave calculations”. In: *Phys. Rev. B* 43 (3 1991), pp. 1993–2006. DOI: 10.1103/PhysRevB.43.1993.
- [101] R. W Hockney. “The potential calculation and some applications”. In: *Methods Comput. Phys.* 9 (1970), pp. 135–211.
- [102] O. A. von Lilienfeld et al. “Variational optimization of effective atom centered potentials for molecular properties”. In: *J. Chem. Phys.* 122.1 (2005), p. 014113. DOI: 10.1063/1.1829051.
- [103] *VMD HBonds Plugin, Version 1.2*. URL: <https://www.ks.uiuc.edu/Research/vmd/plugins/hbonds/>.
- [104] *Hydrogen bonds — GROMACS 2021.4 documentation*. URL: <https://manual.gromacs.org/documentation/current/reference-manual/analysis/hydrogen-bonds.html>.
- [105] J. Pujol-Giménez, M. A. Hediger, and G. Gyimesi. “A novel proton transfer mechanism in the SLC11 family of divalent metal ion transporters”. In: *Sci. Rep.* 7.1 (2017), p. 6194. DOI: 10.1038/s41598-017-06446-y.

4 Computational studies provide new insights into Nramps proton transport

Chapter 4 is a pre-submission version of the article: "Computational studies provide new insights into Nramps proton transport" by **Maria Letizia Merlini**, Alessandra Magistrato, Paolo Carloni and Ursula Röthlisberger. My contribution: computational studies, results elaboration and discussion, graphics and picture production, draft writing.

4.1 Abstract

In Nramp/SLC11 secondary active transporters the transport of divalent transition metal ions across cellular membranes is coupled with that of protons in the same direction (i.e. from the extracellular matrix to the cellular cytoplasm), resulting therefore in a symport phenomenon. However, experimental evidence proved that proton uniport, in which the proton transport occurs in the absence of the metal ion substrate, is also possible. This suggests a noncanonical symporter behaviour for Nramps in which the two substrates are not tightly coupled. In the past years, different and sometimes inconsistent interpretations of this unconventional proton transport mechanism have been proposed. In this work, we performed molecular dynamics (MD) simulations on both *Staphylococcus Capitis* and *Eremococcus Coleocola* Divalent Metal-ion Transporters, whose crystal structures have been obtained in an outward- and an inward-facing conformation, respectively. Since proton uniport in the absence of bound metal has been reported in several works, the apo (substrate free) form of the two proteins was used for the simulations. Furthermore, in order to assess the role played by a conserved histidine residue indicated as potential key player in the proton transport process, both proteins were simulated in the unprotonated and protonated form. Our findings suggest a conserved glutamate residue close to the metal ion binding site facing a narrow, presumably water filled channel formed by α -helices 3 and 9 as the likely primary proton acceptor. We also identified the likely pathway followed by water molecules during their transport into the cytosol and found that it involves four amino acidic residues belonging to a conserved hydrophilic network connecting the binding site to the above mentioned channel. The protons need to pass through at least three of this four residues, thus explaining why upon mutation of these residues, the proton transport is interrupted or severely reduced.

4.2 Introduction

Nramps ("natural resistance-associated macrophage proteins") also called SLC11 ("solute carrier 11") are a family of secondary active transporters responsible for the controlled influx of divalent transition metal ions (such as Fe^{2+} , Mn^{2+} , Cu^{2+} , Zn^{2+} , etc.) across cellular membranes [1–6]. In these proteins the transfer of divalent transition metal ions (primary substrates) from the extracellular matrix to the cellular cytoplasm is coupled to a proton transport in the same direction (symport) [3, 7]. However, unlike in canonical symporters, protons and metal ions are not tightly coupled and their respective stoichiometry can vary as a function of metal ion nature, pH gradient and transmembrane voltage [3, 7].

In the past years, several experimental studies outlined some consistent general features for Nramp prokaryotic and eukaryotic homologs: the transport of protons from the extracellular matrix to the cellular cytoplasm, down the electrochemical gradient, can occur either in the presence or absence of the metal ion substrate, thus leading to "metal-proton symport" or "proton uniport" phenomena, respectively. However, in the majority of cases, the presence of metal ions seems to enhance the proton transport and vice-versa [3, 8–24].

The crystallographic characterization of *Staphylococcus Capitis* Divalent Metal Ion Transporter (ScaDMT) [25], *Eremococcus Coleocola* DMT (EcoDMT) [22] and *Deinococcus Radiodurans* DMT (DraDMT) [23, 26] provided fundamental insights of the proton transport mechanism since they allowed for the identification of some amino acidic residues potentially involved in the process. From structural analysis, it was indeed possible to identify a network of conserved protonatable residues that are located in the core of the protein and which are not present in structurally related LeuT-fold [27, 28] outgroups [7]. Within this network three amino acidic residues (an aspartate on α -helix 1 (TM1), a glutamate on TM3 and a histidine on TM6b) connect the metal binding site to a pair of salt bridges linking TM3 and TM9. This hydrophilic network allows water molecules to pass from the binding site, exposed to the bulk solvent in both the inward-facing conformation (IFC) and the outward-facing conformation (OFC), to the pair of salt bridges formed between TM3 and TM9. These latter bind an aspartate and a glutamate on TM3 with two facing arginines on TM9 and are located at the center of the lipidic bilayer and close to the cytosol, respectively. Moreover, a group of polar and charged residues flanks TM3, TM9 and the closeby TM4 and TM8, forming a narrow hydrophilic channel likely filled with water molecules and, therefore, potentially suitable for proton transfer. The conserved network of putative ionizable residues involved in the proton transfer mechanism is schematically represented in Fig. 4.1.

For the sake of clarity, the specific residues forming the hydrophilic network in ScaDMT, EcoDMT and DraDMT are presented in Table 4.1.

Residue	ScaDMT	EcoDMT	DraDMT
Asp on TM1	Asp49	Asp51	Asp56
His on TM6b	His228	His236	His232
Glu(1) on TM3	Glu127	Glu129	Glu134
Asp on TM3	Asp124	Asp126	Asp131
Glu(2) on TM3	Glu117	Glu119	Glu124
Arg(1) on TM9	Arg355	Arg368	Arg352
Arg(2) on TM9	Arg360	Arg373	Arg353

Table 4.1: Ionizable residues forming the hydrophilic network in ScaDMT [25], EcoDMT [22] and DraDMT [23, 26] crystallographic structures. Notably, the Asp residue on TM1 is also involved in the coordination of the divalent transition metal ion. The Asp and Glu(2) residues on TM3 (lines 4 and 5) form two salt bridges with two facing Arg residues on TM9 (lines 6 and 7), thus linking these two α -helices to shape a narrow hydrophilic channel.

Recently, experimental and computational studies provided different and sometimes inconsistent hypotheses about the mechanism behind the unconventional proton transport occurring in Nramps. In particular, Ehrnstorfer and colleagues indicated His228/236 (ScaDMT and EcoDMT numbering, respectively) on TM6b as the most likely primary proton acceptor [22]. In their mechanistic model, this residue binds the proton in the OFC and, after the conformational switch to the IFC, releases it either via the wide inward-facing vestibule connecting the metal ion binding site to the cytoplasm or through the narrow hydrophilic channel formed by TM3 and TM9. The proposed mechanism is illustrated in Fig. 4.2.

In a subsequent computational study, Pujol-Giménez and colleagues identified Glu127/129

(ScaDMT and EcoDMT numbering, respectively) on TM3 as the most likely proton acceptor [10]. They hypothesized that, after the release of the proton through the hydrophilic side channel taking place in the OFC, an allosteric Coulombic interaction occurring between the deprotonated negatively charged Glu127/129 residue and Asp49/51 triggers the conformational change from the OFC to the IFC and the subsequent metal ion release via the wide inward-open vestibule, as shown in Fig. 4.3.

In a more recent work, Bozzi and colleagues proposed Asp56 (DraDMT numbering) as the most likely proton acceptor [23]. In their mechanistic model, the incoming metal ion binds to Asp56 displacing the proton and transferring it to Asp131 via either Glu134 or His232. From Asp131 the proton is finally released through the narrow side channel. The binding of the metal ion and the simultaneous displacement of the proton triggers the conformational switch. Furthermore, they also speculated that, at contrast to the metal ion, whose release requires the intracellular gate opening (i.e. OFC to IFC conformational switch), the proton transport can eventually occur completely in the OFC. More details about these mechanistic studies can be found in Section 1.2.6 of Chapter 1.

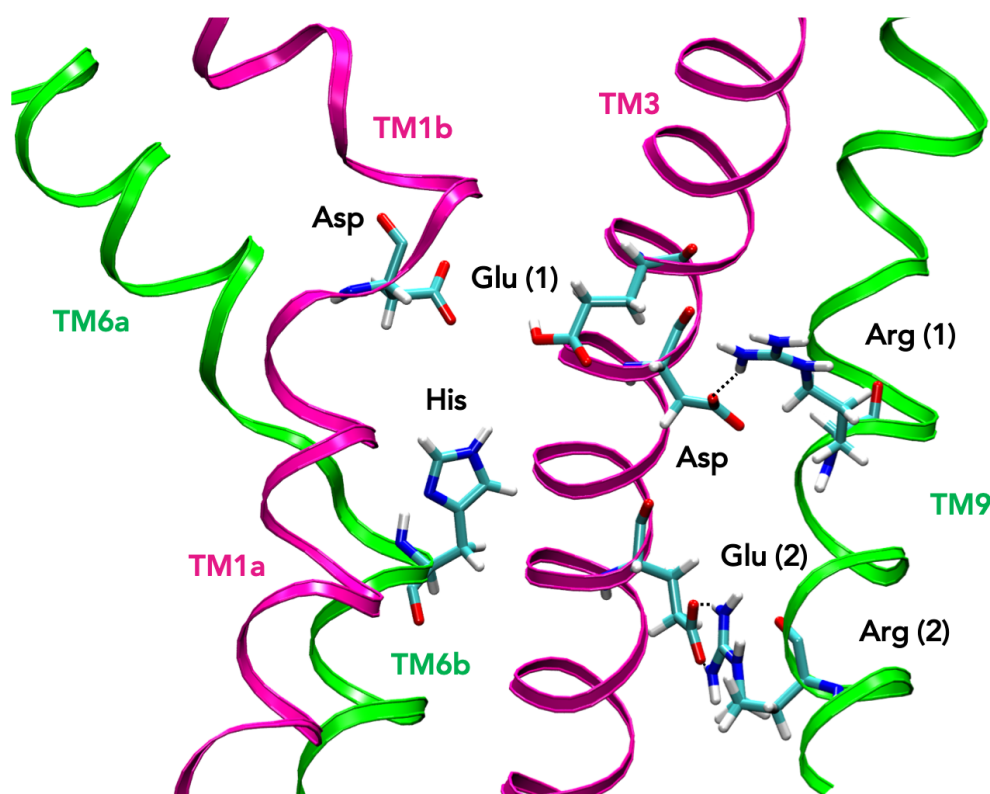


Figure 4.1: Schematic representation of the conserved network of protonatable residues potentially involved in the proton transfer [7, 10, 22, 23]. Transmembrane α -helices TM1 and TM3, belonging to the first protein subdomain, and TM6 and TM9, belonging to the second protein subdomain, are represented as magenta and green ribbons, respectively. Amino acidic residues are represented in a stick style, with C, N, O and H atoms represented in cyan, blue, red and white, respectively. The Asp residue on the unwound region of TM1, together with the His residue on TM6b and the Glu on TM3 connect the metal binding site, located at the unwound regions of TM1 and TM6, with the narrow hydrophilic channel formed by TM3 and TM9, linked together by two salt bridges between Asp and Glu residues located on TM3 and two Arg located on TM9. The figure has been obtained using VMD [29, 30].

Although a wide range of experimental studies have provided fundamental structural and functional evidence, Nramps proton transport mechanism remains still ambiguous, as illustrated by the different and inconsistent mechanistic models described above [10, 22, 23]. Such knowledge is however of crucial importance to rationalize the reasons of the highly variable proton/metal transport stoichiometry and the still undetermined thermodynamic coupling between the two substrates.

In this work, molecular dynamics (MD) simulations have been used to identify the primary proton acceptor as well as to determine the possible pathway taken by protons during the release into the cellular cytoplasm. Both the inward-facing conformation of ScaDMT [25] and the outward-facing conformation of EcoDMT [22] have been investigated. Information about the proton transport have been extracted by a comparison of the results obtained for the unprotonated and protonated forms of His228 and His236 while, among the other residues belonging to the hydrophilic network, Asp51/49, Asp126/124, and Glu119/117 are deprotonated and Glu129/127 and Arg368/355 and Arg373/360 are protonated. Since Nramps can also transport protons in the absence of the metal ion primary substrate, only the apo form of these proteins have been studied. Since Glu129/127 resulted to be protonated in both of the two investigated proteins, we will refer to this residue from now on as Glh129/127.

In our analyses, we mainly focus on the four amino acidic residues belonging to the hydrophilic network that could be potentially involved in the proton transport, namely Asp49/51, Glh127/129, His228/236 and Asp124/126 in ScaDMT/EcoDMT, respectively.

4.3 Computational details

Classical molecular dynamics simulations of four selected systems (apo-ScaDMT-IFC with His228 in both its unprotonated/protonated forms and apo-EcoDMT-OFC with His236 residue in both its unprotonated/protonated forms) were performed in a lipid bilayer of composition CHL (cholesterol) ~ 25%, POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylcholine) ~ 41% and POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphatidylethanolamine) ~ 34%, mimicking the nervous system cell membrane surrounded by a water solution containing Cl^- and K^+ ions in physiological concentrations (given in Table 4.3). For the ScaDMT [25] and EcoDMT [22] proteins, the crystal structures determined by Ehrnstorfer and co-workers available in the RCSB Protein Data Bank [32–34] (PDB IDs: 4WGW [25] and 5M87 [22], with resolutions of 3.40 and 3.30 Å, respectively) have been used. In all four systems the N- and C- termini of the proteins were capped with N-acetyl (ACE) and N-methyl amide (NME) groups respectively, in order to neutralize their charge. Protonation states of all titratable residues of the two proteins were evaluated at the physiological value of pH 7.4, using the H^{++} web-server [35, 36]. However, protonation states of apo-ScaDMT-IFC His228 respectively apo-EcoDMT-OFC His236 were assigned independently from the results of these *in silico* pKa predictions by performing explicit MD simulation for both protonation states.

Since ScaDMT and EcoDMT are both transmembrane proteins, we estimated their initial

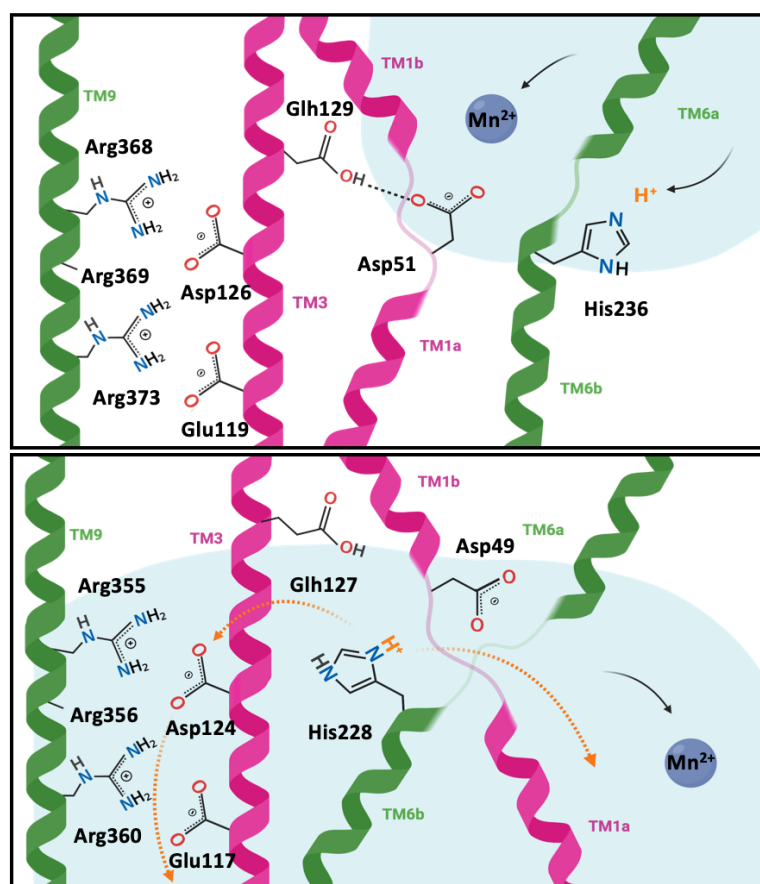


Figure 4.2: Schematic representation of the proton transport mechanism as hypothesized by Ehrnstorfer and colleagues [22]. The residue numbering corresponds to those of EcoDMT and ScaDMT [25] in the **top** and **bottom** panel, respectively. In the OFC (**top**), Asp51 and His236 are both in close contact with the aqueous cavity (represented in light blue). The proton (represented in orange) binds to His236, which acts as the primary proton acceptor. After the conformational change to the IFC (**bottom**) the proton can be released into the cytoplasm either directly, through the wide inward-directed aqueous cavity, which is also the metal ion exit pathway, or via the narrow water-filled channel formed by TM3 and TM9. The two possible exit pathways are indicated as orange dashed arrows. TM1 and TM3, belonging to the first protein subdomain, are represented in magenta while TM6 and TM9, belonging to the second protein subdomain, are represented in green. The metal ion (Mn^{2+}) is represented as a gray sphere. Within the amino acid residues, O and N atoms are represented in red and blue, respectively, while C and H atoms are represented in black. The pictures have been obtained with Biorender [31].

orientation with respect to the hydrocarbon core of the lipid bilayer with the help of the OPM (Orientations of Proteins in Membranes) database [37, 38]. For this, the protein is treated as a rigid body able to freely float in a hydrophobic slab of adjustable thickness. Its spatial arrangement is calculated by minimizing the free energy needed for the transfer from the bulk water to the lipid bilayer. The three-component membrane containing POPC, POPE and CHL was created using the MemBuilder II web-based graphical user interface [39, 40]. A symmetric pre-thermalized membrane (i.e. same number of lipids in the two leaflets), with a square base length of 105 Å and thickness of 51 Å has been used for all four systems. Before insertion of the proteins it contained a total of 450 lipid molecules: 190 POPC, 152 POPE and 108 cholesterol molecules [41, 42]. However, some of the lipid molecules have been removed to incorporate

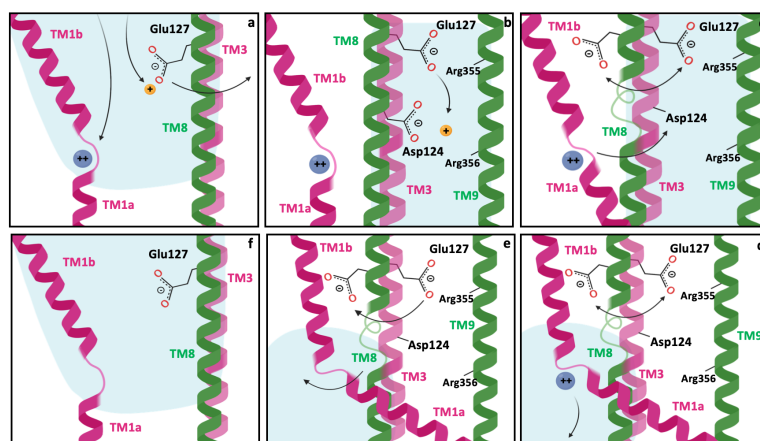


Figure 4.3: Schematic representation of the overall transport mechanism as described by Pujol-Giménez and colleagues [10]. (a) The substrates both bind in the high-affinity OFC, where Glu127 is in a high-pKa state and accepts the proton. (b) After a conformational switch to an inward-occluded state, the proton is released by Glu127 likely along the narrow channel formed by TM3 and TM9. (c) The negative charge on Glu127 after the proton release is the "signal" that triggers the intracellular gate opening while, simultaneously, a partial unwinding of TM8 confers high flexibility to Glu127. (d) In the IFC the metal ion is solvated and released into the cytoplasm. (e) Glu127 reorients itself into the high-pKa state and the gate closes, thus recovering the initial OFC state (f). The proton and the divalent transition metal ion are represented as orange and gray spheres, respectively. Their charge is reported next to it. TM1 and TM3 and TM8 and TM9 are represented in magenta and green, respectively. The solvent accessible region of the protein is represented in light blue. The pictures have been obtained using Biorender [31].

the proteins by using the gmx membed tool [43] available in the Gromacs simulation package [44–46]. The number of POPC, POPE and cholesterol molecules composing the membrane at the end of the embedding process and the final percentage composition of the membrane in each of the four systems is given in Table 4.2.

System	CHL	POPC	POPE	% CHL	% POPC	% POPE
apo-ScaDMT-His228u	99	165	137	24.7	41.1	34.2
apo-ScaDMT-His228p	99	165	137	24.7	41.1	34.2
apo-EcoDMT-His236u	98	164	138	24.5	41.0	34.5
apo-EcoDMT-His236p	98	164	138	24.5	41.0	34.5

Table 4.2: Total number of CHL, POPC and POPE molecules and final percentage composition of the membrane in each of the four studied systems. The letters u and p stand for unprotonated and protonated His residues, respectively.

The systems were then solvated using the gmx solvate option and K^+ and Cl^- ions were added using the gmx genion option, both of them included in the Gromacs software package [44–46]. The total number of water molecules and K^+ and Cl^- ions as well as the concentration of these ions in each of the four systems are given in Table 4.3.

The four final systems apo-ScaDMT-His228u, apo-ScaDMT-His228p, apo-EcoDMT-His236u and apo-EcoDMT-His236p (where u stands for unprotonated residue while p stands for protonated residue) were composed of a total number of atoms equal to 266567, 266566, 267716

System	H ₂ O	K ⁺	Cl ⁻	[K ⁺]	[Cl ⁻]
apo-ScaDMT-His228u	71159	119	123	92.9	96.0
apo-ScaDMT-His228p	71158	119	124	92.9	96.8
apo-EcoDMT-His236u	71076	116	120	90.7	93.8
apo-EcoDMT-His236p	71075	116	121	90.7	94.6

Table 4.3: Number of total water molecules, K⁺ and Cl⁻ ions, as well as K⁺ and Cl⁻ concentrations in each of the four studied systems (reported in mM).

and 267715, respectively.

The Amber99SB-ILDN [47] and Amber lipid 14 [48] force fields (FF) have been used to describe the proteins and the membrane lipids, respectively. The 3-site rigid TIP3P [49] water model has been used to treat the solvent while K⁺ and Cl⁻ ions were described using the parameters from Joung and Cheatham [50]. Harmonic position restraints (force constants $f_x = f_y = f_z = 10000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$) have initially been applied on the protein heavy atoms (i.e. protein backbone and side chains, excluding hydrogen atoms) and successively slowly removed, in order to thermalize the system.

Classical MD simulations have been performed using the Gromacs package (version 2016-4). For each of the four investigated systems a long equilibration run of 500 ns followed by a production run of 1.0 μs was performed using an integration time step of 2 fs. All covalent bonds involving hydrogen atoms have been constrained with the LINCS algorithm [51]. The Particle Mesh Ewald (PME) [52, 53] method was used to account for long-range electrostatic interactions while a switch function with a real space cutoff of 1.2 nm was used to account for van der Waals interactions. Simulations were performed in the isothermal-isobaric NTP ensemble using the Parrinello-Rahman barostat [54, 55] (1 bar) and the Nosé-Hoover thermostat [56–58] (310K).

The time-dependent analysis of the pKa values of the four selected residues was performed (for each of the four investigated systems) using the PropKa software [59–61] (version 3.0).

In order to verify whether these four residues are connected through hydrogen-bonded water molecules acting as a bridge a self-made Python script using the MDAnalysis Water Bridge Analysis module has been employed [62–64]. Here, a "water bridge" is defined as one or more water molecules forming hydrogen bonds between each other and with both the two selected residues (i.e. the two residues whose connection needs to be verified) so that the two residue are connected by consecutive hydrogen bonds going from the starting residue to the target residue. If the "bridge" connecting two residues involves two or more water molecules it is defined as a "n-th order water bridge", where n is the number of water molecules that participate. Different types of water bridges are represented in Fig. 4.4. For the hydrogen bonds detection the default geometric criteria of the software were used (i.e. donor-acceptor distance cutoff equal to 3.0 Å and donor-hydrogen-acceptor angle cutoff equal to 120°).

Finally, an analysis of the Solvent Accessible Surface Area (SASA) was performed using the

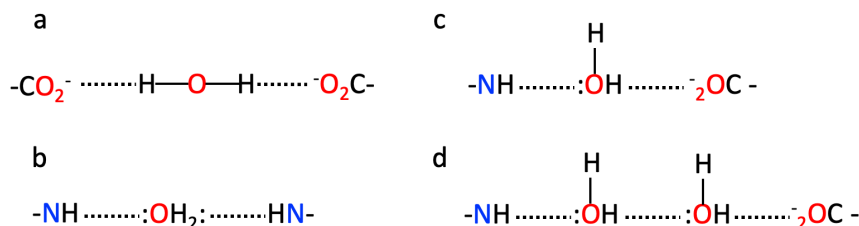


Figure 4.4: Examples of water bridges. (a) Water bridge formed between two hydrogen bond acceptors; (b) water bridge formed between two hydrogen bond donors; (c) water bridge formed between a hydrogen bond donor and a hydrogen bond acceptor; (d) second order water bridge composed by two water molecules.

Gromacs analysis tools [65, 66].

4.4 Results

In order to assess the ability of the selected residues to accept and release protons, we traced their pKa values during time along each of the four trajectories obtained from MD simulations [59, 60]. As already mentioned in section 4.3, the pKa analysis was performed using the PropKa software. In particular, the editable scripting version of the software was used to obtain the pKa values of the four investigated residues for each of the 1000 frames composing the four trajectories resulting from the MD simulations of apo-ScaDMT and apo-EcoDMT. At contrast to previous *in silico* pKa predictions performed on single crystallographic structures, this time-dependent analysis allowed to study the pKa as a dynamical property of amino acids and evaluate their tendency to accept or donate protons in the different local environments sampled during the MD simulations.

As previously mentioned, we performed MD simulations of apo-EcoDMT [22] and apo-ScaDMT [25] proteins in both the unprotonated/protonated states of their conserved histidine residue (His236/228 for EcoDMT and ScaDMT numbering, respectively). However, the average pKa values of this residue, calculated for all configurations sampled during the trajectories of both the unprotonated as well as the protonated form turns out to be consistently low (3.91 ± 0.20 and 3.72 ± 0.27 respectively for apo-EcoDMT and apo-ScaDMT with His236/228 in their unprotonated form and 3.94 ± 0.32 and 4.02 ± 0.24 for apo-EcoDMT and apo-ScaDMT with His236/228 in their protonated form respectively, as reported in Table 4.4 and represented in Fig. 4.5 and 4.6 and in SI Fig. B.1 and Fig. B.2) and not in agreement with the standard His pKa value in aqueous solution (~ 6.05 - 6.10). This means that at physiological pH, the probability to find this histidine residue in the protonated form for a significant interval of time during both simulations is very low and, therefore, from now on only the results regarding the simulations with His236/228 in the unprotonated form will be presented (while results regarding the two simulations with His236/228 in the protonated form are reported in SI). Furthermore, for the sake of clarity, we will refer to the apo-EcoDMT simulation with His236 in its unprotonated form in the OFC and to the apo-ScaDMT simulation with His228 in its unprotonated form in

the IFC as simply "OFC" and "IFC", respectively. The instantaneous pKa values of the four selected residues evaluated in IFC and in OFC with His228/236 deprotonated are shown in Figs. 4.5 and 4.6.

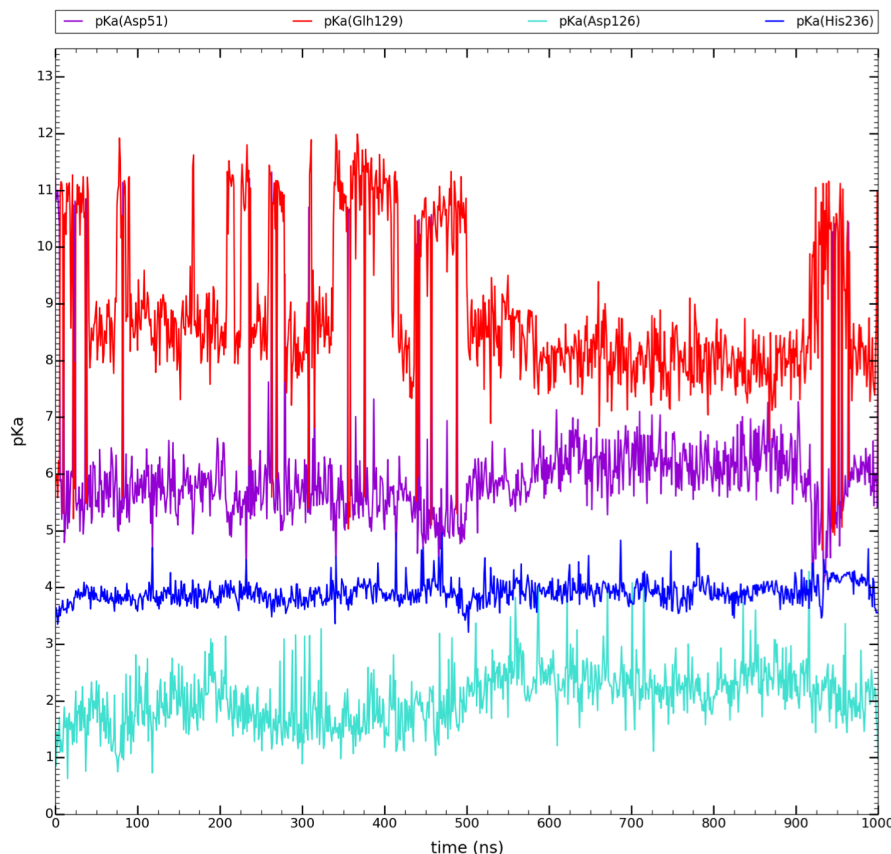


Figure 4.5: pKa values of residues Asp51 (violet), Glh129 (red), Asp126 (turquoise) and His236 (blue) during the OFC trajectory (1 μ s = 1000ns) with His236 in its deprotonated form. Time is reported in ns. Residue numbering from OFC.

Consistently for both OFC and IFC, 3 out of the 4 considered residues (Glh129/127, Asp51/49 and His236/228) have pKa values that are clearly shifted with respect to their standard values in aqueous solution (~ 4.25 , ~ 3.85 and ~ 6.05 for E, D and H, respectively) suggesting that their microenvironment might have evolved in such a way as to tune them for a specific functional role (Table 4.4, Fig. 4.5 and Fig. 4.6). The most extreme case is represented by Glh129/127 owing an average pKa value of 8.80 ± 1.35 in the OFC and 9.21 ± 0.52 in the IFC, i.e. upshifted by as much as ~ 4 -5 pKa units with respect to its standard value in aqueous solution.

Interestingly, in the OFC (Fig. 4.5) the pKa value of Glh129 undergoes large variations over 7 pKa units, sampling values in the range ~ 5 to 12, reminiscent of a three-state behaviour with pKa ranges around 10-12, 8-9 and 5-6. The high average pKa value clearly indicates that Glh129 is mostly protonated at neutral pH and could indeed act as highly efficacious primary proton acceptor. Furthermore, the simultaneous existence of high pKa states and

those slightly below physiological pH indicates a strong tendency of this residue to accept protons and release them under physiological conditions.

Remarkable shifts are also present in the pKas of Asp51, which can vary from maximum values of ~ 11 to minimum values of ~ 4 (Table 4.4) with an average value of 5.99 ± 0.98 , as shown in Fig. 4.5. This suggests that Asp51 is also likely to undergo protonation and deprotonation processes. In particular, the overlap between the instantaneous pKas of Glh129 and Asp51 occurring several times during the simulation, suggests a likely exchange of protons between the two residues.

At contrast to Asp51 and Glh129, the pKa of His236 is significantly downshifted (average pKa 3.91 ± 0.20) with respect to its bulk value but remains more or less constant during the dynamics, with maximum and minimum values oscillating in a much smaller range ($\Delta pKa \sim 1.75$), as shown in Fig. 4.5. However, occasionally configurations are sampled which bring the pKa of His236 close to physiological pH (7.4) and in close overlap with the one of Asp51 opening the possibility of a further proton exchange from Asp51 to His236.

On the other side, the pKa values of Asp126, ranging from maximum values of ~ 5 to minimum values of ~ 0.5 , with a ΔpKa range of 3.65 (further details are given in Table 4.4), display scarce overlap with either His236 and the other two residues. An exchange of protons involving this residue is therefore less likely in the OFC, although it cannot be excluded. The average pKa value of this residue of 2.05 ± 0.50 seems to indicate that it remains in its deprotonated form for most of simulation time, however possible rapid protonation/deprotonation events could explain the full proton uniport observed in the OFC in the absence of the metal ion.

Focussing on Fig. 4.6, where the pKa values of the four selected residue in the inward-facing conformation of apo-ScaDMT are represented, the situation is quite different. Here, the pKa values of Glh127 oscillate over a range of ~ 3 units, with maximum values ~ 11 and minimum values ~ 8 (as shown in Table 4.4 and in Fig. 4.6) without displaying noticeable downward shifts. With an average pKa value equal to 9.21 ± 0.52 Glh127 remains throughout the trajectory in a high-pKa state suggesting that, in this conformation, it tends to maintain a protonated state. The three other residues (Asp49, His228 and Asp124) have pKa values in a similar range (average pKa values 4.83 ± 0.41 , 3.72 ± 0.27 and 3.27 ± 0.50 , respectively) that would facilitate isoenergetic proton exchange among them (Fig. 4.6). However, they all lie at too low pKa values to be present in both protonated and unprotonated forms at neutral pH. One possible explanation could be that the pKa values of these three residues might be changed significantly in the presence of the metal ion. This would suggest that proton uniport in the apo form of the transporter happens mainly in the OFC while proton transport in the IFC is indeed coupled to the metal. Another explanation could be that overall all these three residues are in their deprotonated forms but rapid proton shuttling, in which the protonated form are very short lived, can happen, especially once Asp49 (and maybe already His228) has already been protonated in the OFC.

In particular, Asp49 pKa values vary in a range of around ~ 3 units, with average pKa value of 4.83 ± 0.41 while His228 pKa oscillates in a range of around ~ 1.6 units, with average value of 3.72 ± 0.27 , as possible to see in Table 4.4 and in Fig. 4.6. This residue displays a remarkable

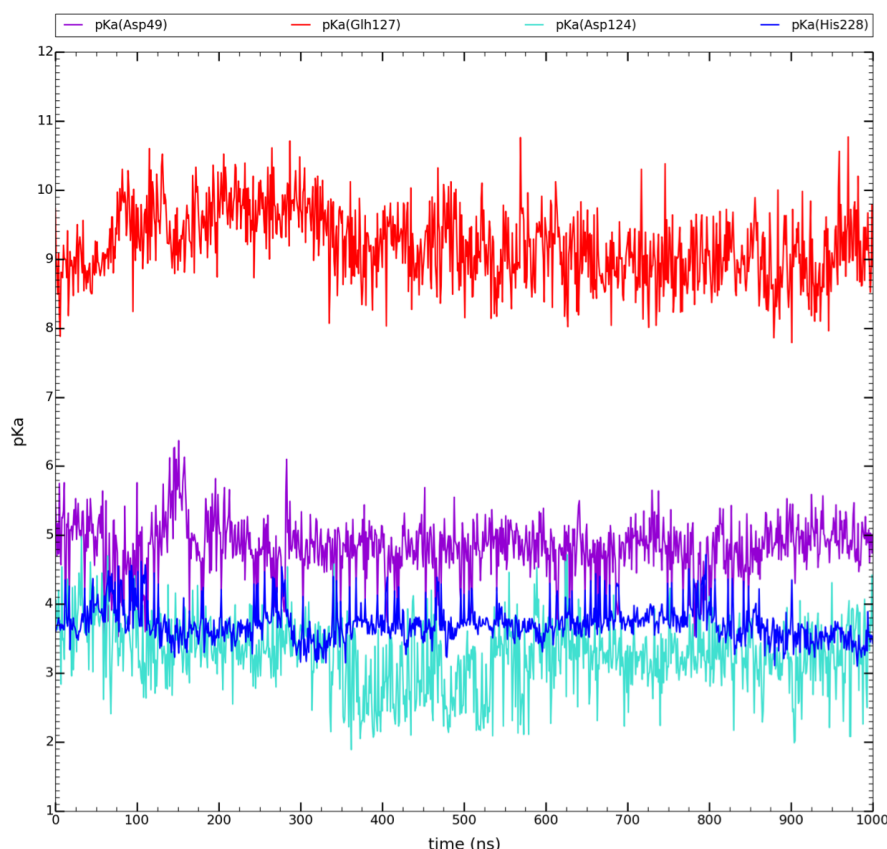


Figure 4.6: pKa values of residues Asp49 (violet), Glh127 (red), Asp124 (turquoise) and His228 (blue) during IFC trajectory (1 μ s = 1000ns) with His228 in its deprotonated form. Time is reported in ns.

number of upward shifts occurring simultaneously with Asp49 downward shifts. The overlap between Asp49 and His228 pKa values suggests a likely exchange of protons between these residues. Lastly, Asp124 pKa values oscillate in a range of around ~ 3 units with average value of 3.27 ± 0.50 . Detailed values are given in Table 4.4. Asp124 pKa curve superimposes with that of His228 during the whole trajectory, suggesting that exchange of protons between these two residues can eventually occur.

Summarizing, Glh129 most likely acts as the primary proton acceptor although an initial proton transfer from Asp51 cannot be entirely excluded. Protons exchange can occur between Glh129 and Asp51 and, to some extent, between Asp51 and His236 (and His236 and Asp126) in the OFC and between Asp49 and His228 and His228 and Asp124 in the IFC.

PropKa used to perform the pKa analysis is based on a set of empirical rules accounting for the hydrogen bonding, the desolvation and the intra-protein (charge-charge) interactions effects affecting the standard pKa values of amino acidic residues. More rigorous pKa predictions can be obtained using methods based on the explicit solution of the Poisson-Boltzmann equation [35, 67–71]. However, all of these software packages can only be used through webserver, in which single structures are processed at a time making therefore the time-dependent pKa

analysis that we performed much more cumbersome.

The results of this time-dependent pKa analysis suggest a primary proton acceptor role for either Glh129 (or Asp51). To further assess the ability of the four selected residues to accept/release protons we performed an analysis of the Solvent Accessible Surface Area (SASA) [66]. OFC Asp51, Glh129, Asp126, His236 SASA values over time are reported in the left graph of Fig. 4.7 while IFC Asp49, Glh127, Asp124 and His228 SASA values over time are reported in the right graph of Fig. 4.7. Surprisingly, all four residue have similar SASA values in outward-versus inward-facing conformations of apo-EcoDMT and apo-ScaDMT. Despite its position, within the metal-ion binding site, that would suggest Asp51/49 as the residue with the largest surface area exposed to the solvent, interestingly Glh129/127 and His236/His228 result to be the residues with largest solvent exposed superficial area. However, since the SASA is directly proportional to the size of the residue and, consequently, to the number of atoms composing it, we decided to further verify these results by computing the SASA only for the carboxylic groups of Asp51/49 and Glh129/127 side chains of OFC and IFC, as shown in Fig. 4.8. In fact, according to this, the two residues exhibit similar SASA values for their carboxylic groups with slightly higher values for Glh, suggesting that both of them are similarly solvent accessible and thus eligible as primary proton acceptors, as already suggested from the pKa analysis.

residue	$pK_{aAV} \pm \sigma(pKa)$	ΔpKa	δpKa
Asp51	5.99 ± 0.98	7.32	2.13
Glh129	8.80 ± 1.35	7.33	4.73
Asp126	2.05 ± 0.50	3.65	-1.81
His236	3.91 ± 0.20	1.75	-2.19
Asp49	4.83 ± 0.41	2.92	0.97
Glh127	9.21 ± 0.52	2.98	5.14
Asp124	3.27 ± 0.50	3.09	-0.59
His228	3.72 ± 0.27	1.62	-2.38

Table 4.4: Average pKa (pK_{aAV}) values with their standard deviations ($\sigma(pKa)$), ΔpKa (defined as the difference between the maximum and minimum pKa values assumed during the simulation) and δpKa (defined as the difference between the average pKa value and the standard pKa value in aqueous solution at physiological pH) for residues Asp51, Glh129, Asp126 and His236 of OFC (lines 1-4) and for residues Asp49, Glh127, Asp124 and His228 of IFC (lines 5-8). Reference standard pKa values in aqueous solution at physiological pH are taken from ref. [72]. Each residue is indicated in the same color used in Fig. 4.5 and 4.6.

The presence of solvent molecules around Asp51/49 and Glh129/127 and inside the hydrophilic narrow channel in both OFC and IFC is graphically represented in Fig. 4.9 illustrating both the accessibility of these two residues to water and the possibility for the proton to be released towards the cytoplasm through the channel.

Since the proton transport likely occurs through a Grotthus-like "proton hopping mechanism", we performed a statistical analysis to verify whether during the OFC and IFC MD simulations the four selected residues are connected through one or more water molecules forming hydrogen bonds between each other and with the two residues representing the starting and arrival points. The water molecule(s) eventually connecting two residues act(s) as a "bridge"

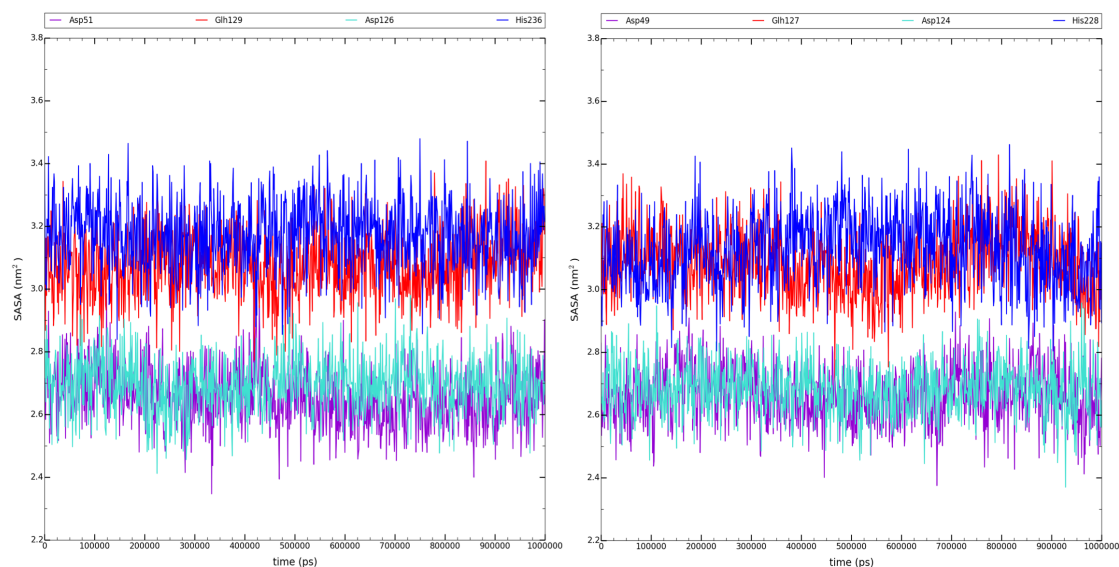


Figure 4.7: (left) Asp51, Glh129, Asp126 and His236 Solvent Accessible Surface Area (SASA) values over time ($1 \mu\text{s} = 1000 \text{ ns}$) in OFC trajectory. (right) Asp49, Glh127, Asp124 and His228 SASA values over time in IFC trajectory. SASA values are given in nm^2 while time is reported in ps.

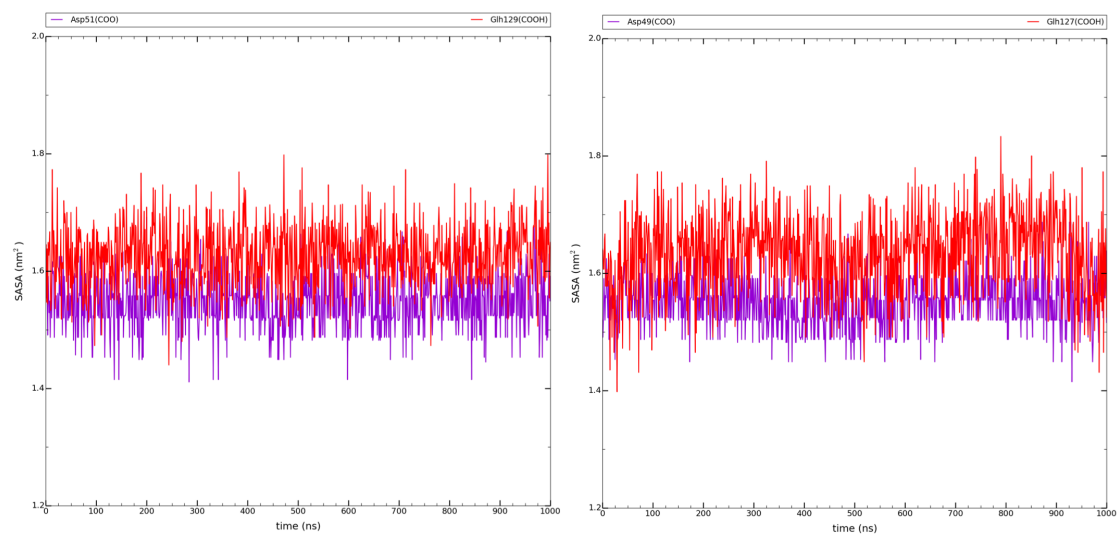


Figure 4.8: Solvent accessible surface area (SASA) over time for the side chain carboxylic groups of the two residues indicated as potential primary proton acceptors, namely Asp51 and Glh129 in OFC, (left) and Asp49 and Glh127 in IFC (right). Asp51/49 and Glh129/127 are represented in violet and red, respectively. The SASA values are given in nm^2 while time is given ns.

and, therefore, we will refer to them as "water bridges". In particular, we assessed the statistical presence (i.e. the percentage of time in which water bridges occur over the total simulation time) of water bridges for each of the three couples of amino acidic residues represented in Fig. 4.10: (1) Asp51/49-His236/228; (2) Asp51/49-Glh129/127 and (3) His236/228-Asp126/124 following OFC and IFC numbering, respectively. The three investigated couples are those for which we already identified a possible exchange of protons in the previously discussed pKa

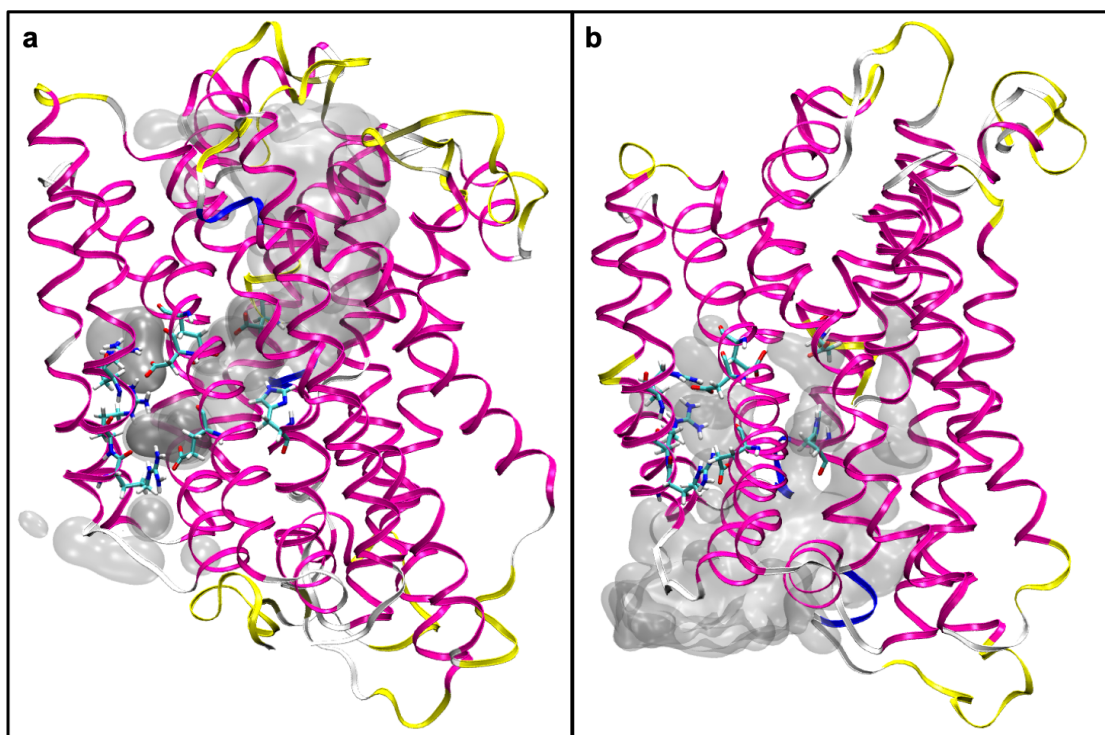


Figure 4.9: Graphical representation of the accessibility of solvent inside the narrow hydrophilic channel formed between TM3 and TM9 and in contact with Asp51/49 and Glh129/127 in OFC (a) and IFC (b), respectively. The two proteins are represented as magenta ribbons while the residues forming the hydrophilic network are represented as sticks with C, H, O, and N atoms colored in cyan, white, red and blue, respectively. The solvent is represented as a gray mesh. Both pictures have been obtained using VMD [29, 30].

analysis. This statistical analysis has been performed on both OFC and IFC trajectories (with, respectively, His236/228 in its unprotonated state), both of them 1 μ s long and composed by 10000 frames.

The results of this statistical analysis are summarized in Tables 4.5 and 4.6. Water bridges with consecutive hydrogen bonds connect Asp51 and His236 in the OFC and Asp49 and His228 in the IFC for respectively 66.1% and 63.7% of the total simulation. At contrary to Asp51/49-His236/228, exhibiting similar values for the percentage of existence of water bridges in the OFC and in the IFC, the Asp51-Glh129 pair displays a percentage of existence of water bridges of 34.8% in the OFC which is almost the double of what is found for the corresponding couple in the IFC (19.8%). This means that proton exchange between these two residues can occur in both conformational states but it is more probable in the OFC. However, it is important to outline here that Asp51/49 and Glh129/127 are very close to each other and can eventually be connected through a direct hydrogen bond formed between their side chain carboxylic groups. Since protons can be exchanged through either water-mediated or direct hydrogen bonds, the overall percentages in which the residues are somehow connected and proton exchange is possible are expected to be higher than those mentioned.

Finally, for the couple His236-Asp126, intact water bridges exist for only 3% of the whole

Amino Acid Pair	apo-EcoDMT-His236u
Asp51-His236	66.1%
Asp51-Glh129	34.8%
His236-Asp126	3.0%

Table 4.5: Percentage of existence (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the OFC MD simulations (His236 unprotonated).

Amino Acid Pair	apo-ScaDMT-His228u
Asp49-His228	63.7%
Asp49-Glh127	19.8%
His228-Asp124	12.8%

Table 4.6: Percentage of existence of water bridges (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the IFC MD simulation (His228 unprotonated).

OFC MD simulation. This percentage increases up to 12.8% in the IFC MD simulation. This evidence suggests that proton transfer between these two residues is more likely to occur in the IFC. Average distances (together with their standard deviations) between the residues forming the analysed pairs are given in Tables B.4 and B.5.

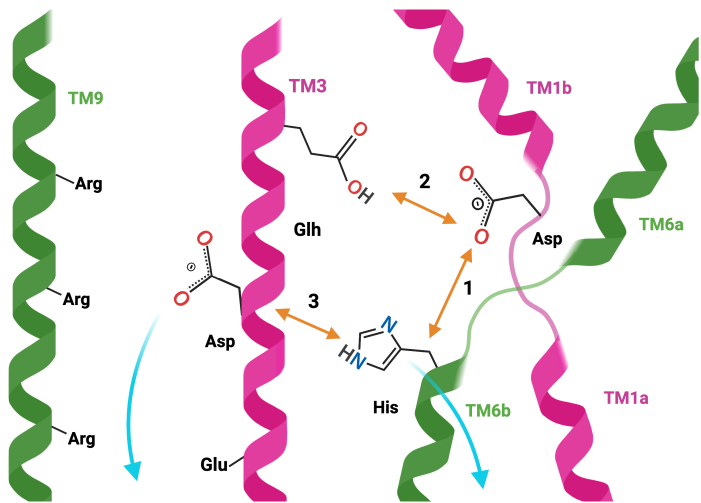


Figure 4.10: Pairs of amino acidic residues investigated to verify the eventual presence of water bridges during OFC and IFC MD simulations. As usual, TM1 and TM3 are colored in magenta while TM6 and TM9 are colored in green. C and H atoms of the residues are indicated in black while O and N are represented in red and blue, respectively. The three couples of residues analysed are represented with orange arrows. They are: (1) Asp51/49-His236/228; (2) Asp51/49-Glh129/127 and (3) His236/228-Asp126/124 in apo-EcoDMT and apo-ScaDMT numbering, respectively. The cyan arrows indicate the two possible release routes of protons: along the narrow hydrophilic channel or in the wide inward-directed vestibule connecting the active site and the cellular cytoplasm. The picture has been obtained using Biorender [31].

Although residues Asp51 and His236 (OFC) and the corresponding couple of residues Asp49 and His228 (IFC) are well connected (with similar percentages of existence of water bridges) in both the OFC and the IFC, indicating that proton exchange between the two residues is possible in both conformational states, the results of the pKa analysis suggest that these two residues likely exchange protons in the inward-facing conformation. On the other hand, both

the pKa and the water bridge analysis suggest proton exchange between Asp51 and Glh129 in the OFC and between His228 and Asp124 in the IFC.

Taken together, the outcomes of all the discussed analyses allows to suggest a potential pathway for proton transport. In fact, the proton pathway likely involves all four of the investigated residues and can occur either in the OFC alone or in combination between OFC and IFC. The pKa analysis clearly indicates that Glh129 is in its protonated state at neutral pH while the high and low pKa states visited by this residue during the simulation suggest a strong tendency to undergo rapid and sequential deprotonation/protonation events. Therefore we suggest that this residue plays the crucial role of primary proton acceptor initiating the proton transport process. It likely transfers its proton to Asp51 in the outward-facing conformation. After the conformational change leading the protein from the OFC to the IFC and presumably happening once Asp56 is protonated (it could in principle also happen after the proton exchange between Asp56 and His236) the proton is likely transferred to His236/228 which, in turn, relays it to the bulk water in the wide inward-directed vestibule connecting the metal ion binding site with the cytoplasm or, eventually, to the aspartate residue located on TM3, Asp124. From there the proton could be released into the cytoplasm along the hydrophilic narrow side channel existing between TM3 and TM9. Furthermore, in the OFC proton exchange from His236 to Asp126 and from there into the narrow hydrophylic side channel would explain the proton uniport observed in the absence of the metal ion substrate occurring completely in the OFC. The mechanistic models proposed for the proton transport involving both the OFC and the IFC or occurring entirely in the OFC are schematically represented in Fig. 4.11 and Fig. 4.12, respectively.

The aspartate residue located on TM1 (Asp51/49) is also suitable to act as potential (or secondary) proton acceptor, as indicated by the remarkable shifts and fluctuations of its pKa value. We suggest that it can eventually accept a proton and carry on the transport mechanism by just skipping the first step. This would explain why proton transport has been observed in E127A and E127D (ScaDMT numbering) mutants by Pujol-Giménez and co-authors [10]. In any case, the proton needs to pass by Asp51 (EcoDMT numbering) thus explaining the removal of proton transport observed by Bozzi and colleagues in the D56N and D56A mutants (DraDMT numbering) [7, 23]. The same reasoning can be applied to the His residue. In the suggested proton transport schemes, the proton indeed passes through His236/228 (EcoDMT and ScaDMT numbering, respectively) and this is in agreement with the outage of proton influx observed in the H236A mutant (EcoDMT numbering) by Ehrnstorfer and colleagues [22].

4.5 Conclusions

In this work, the OFC EcoDMT and IFC ScaDMT proteins in their apo form and with His236/228 (EcoDMT and ScaDMT numbering, respectively) in both their unprotonated and protonated states have been investigated through classical MD simulations. The obtained trajectories

were analysed with respect to multiple features aimed at gaining new atomistic-level insights into the proton transport mechanism occurring in Nramp transporters. We focused in particular on the role played by four amino acidic residues (Asp51/49, Glh129/127, His236/228 and Asp126/124, in apo-EcoDMT and apo-ScaDMT numbering, respectively) belonging to the conserved hydrophilic network and already indicated as potentially players involved in the proton transport process.

The time-dependent evolution of the pKa values of the four selected residues allows to evaluate their tendency to accept or donate protons in the different local environments sampled during the MD simulations. We were able therefore to identify Glh129/127 and Asp51/49 as the potential primary proton acceptors initiating the transport process as well as to indicate the pairs of residues which are more likely to exchange protons: namely Glh129-Asp51 and Asp51-His236 in the OFC and Asp49-His228, His228-Asp124 in the IFC. The solvent accessible surface area (SASA) analysis confirmed Glh129/127 and Asp51/49 as potential primary proton acceptors. Finally, since the proton transport presumably occurs through a Grotthus-like "proton hopping mechanism", the existence of chains composed by one or more water molecules bridging two of the concerned residues through hydrogen-bonds and allowing therefore the proton transfer between the selected couples of residues was assessed through a statistical analysis.

Putting together all the results of the analysis performed, we are able to suggest a potential proton transfer pathway. In the proposed model, the proton is transferred from the glutamate residue, undergoing rapid and consecutive deprotonation/protonation cycles and representing the most likely primary proton acceptor, to the aspartate residue located on TM1. After the conformational change from the OFC to the IFC (probably occurring after aspartate protonation) this residue transfers the proton to the other aspartate located on TM3 via the histidine on TM6b. From here the proton is finally released into the cytoplasm presumably through the narrow hydrophilic side channel. However, a direct release of the proton from the histidine through the wide inward-directed vestibule connecting the binding site with the cellular cytoplasm can not be excluded.

Such a long multistep pathway for proton transport could be due to the always present salt bridge interaction between the side chain carboxylic group of the aspartate on TM3 and the facing arginine. In this situation the involved residues can not take part to the proton transport.

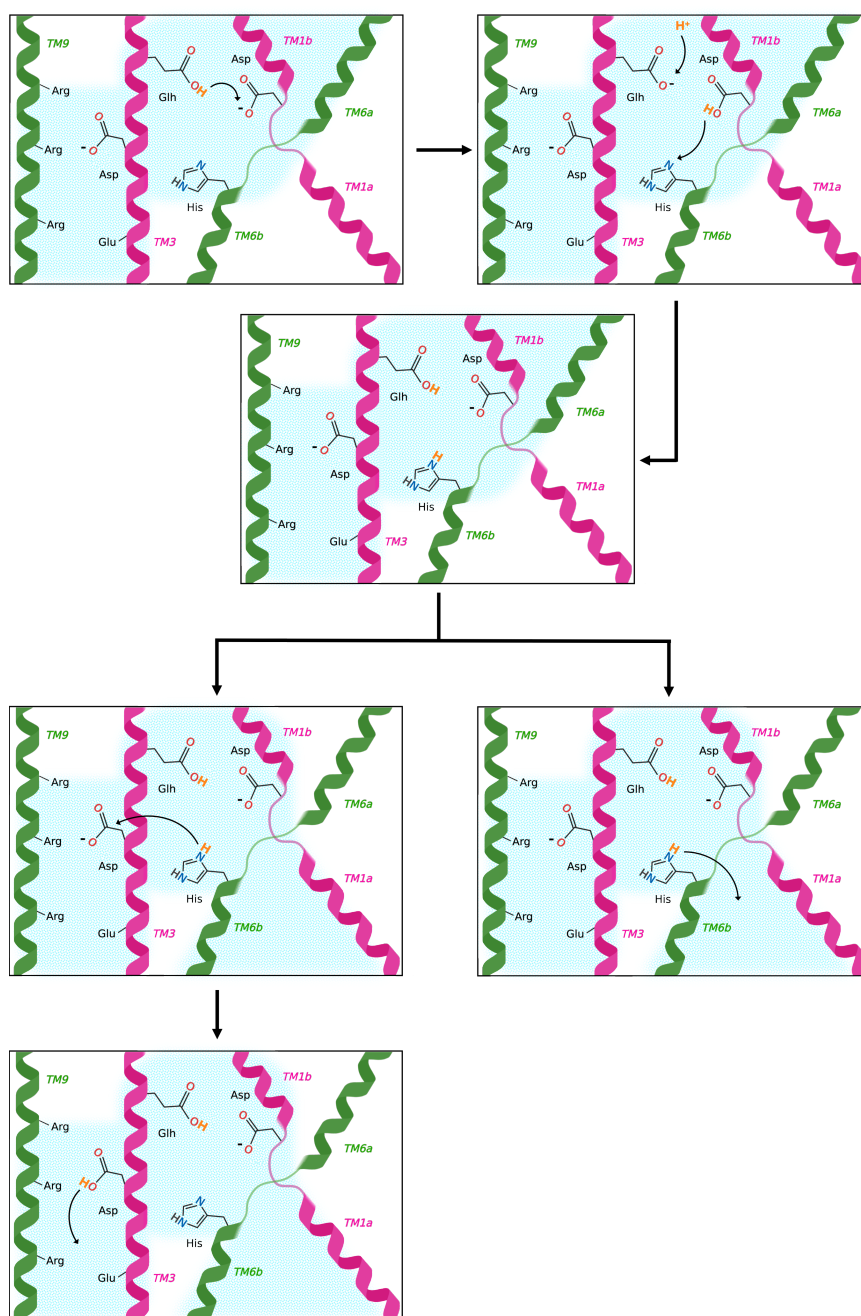


Figure 4.11: Mechanistic model proposed for the proton transport involving both OFC and IFC. The protonated Glu residue located on TM3 (Glh) likely acts as the primary proton acceptor initiating the proton transport process. In the OFC, the proton is transferred from Glh to the Asp residue located on TM1 (**top left**) and from there to the His residue located on TM6b (**top right center**) while the deprotonated Glh residue immediately recovers its protonated state. Once the proton is located on the His residue and the conformational change to the IFC has occurred the proton can be released into the cellular cytoplasm using two distinct routes: it is either directly released through the wide inward-directed vestibule connecting the metal-ion binding site and the cytoplasm (**bottom right**) or it is transported to the Asp residue located on TM3 and finally from there released into the cellular cytoplasm through the narrow hydrophilic channel formed between TM3 and TM9 (**bottom left**). The light blue regions indicate regions accessible to the solvent while the proton is represented in orange. The sequence of events is indicated by the black arrows. The picture has been obtained using Biorender [31].

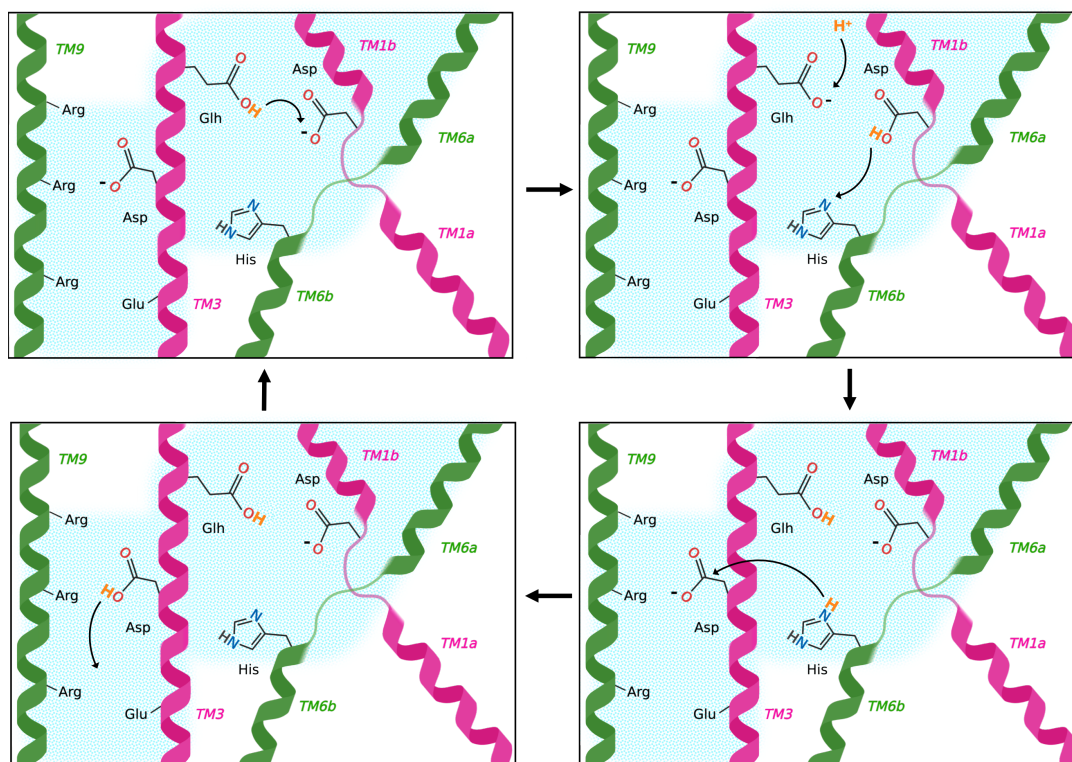


Figure 4.12: Mechanistic model proposed for the proton transport occurring entirely in the OFC. From the protonated Glu residue located on TM3 (Glh) likely acting as the primary proton acceptor, the proton is transferred to the Asp residue on TM1. From there it is transferred to the His residue located on TM6b which, in turn, transfers it to the Asp residue on TM3. Finally, the proton is released into the cellular cytoplasm through the narrow hydrophilic channel formed between TM3 and TM9. The light blue regions indicate regions accessible to the solvent while the proton is represented in orange. The sequence of events is indicated by the black arrows. The picture has been obtained using Biorender [31].

References

- [1] Y. Nevo and N. Nelson. “The NRAMP family of metal-ion transporters”. In: *Biochim. Biophys. Acta* 1763.7 (2006), pp. 609–620. DOI: <https://doi.org/10.1016/j.bbamcr.2006.05.007>.
- [2] M. Cellier et al. “Nramp defines a family of membrane proteins.” In: *Proc. Natl. Acad. Sci.* 92.22 (1995), pp. 10089–10093. DOI: [10.1073/pnas.92.22.10089](https://doi.org/10.1073/pnas.92.22.10089).
- [3] H. Gunshin et al. “Cloning and characterization of a mammalian proton-coupled metal-ion transporter”. In: *Nature* 388 (1997), pp. 482–488. DOI: <https://doi.org/10.1038/41343>.
- [4] J. R. Forbes and P. Gros. “Divalent-metal transport by NRAMP proteins at the interface of host–pathogen interactions”. In: *Trends Microbiol.* 9.8 (2001), pp. 397–403. DOI: [https://doi.org/10.1016/S0966-842X\(01\)02098-4](https://doi.org/10.1016/S0966-842X(01)02098-4).
- [5] M. F. M. Cellier. “Chapter Ten - Nramp: from Sequence to Structure and Mechanism of Divalent Metal Import”. In: *Metal Transporters*. Vol. 69. Curr. Topics Membr. Academic Press, 2012, pp. 249–293. DOI: <https://doi.org/10.1016/B978-0-12-394390-3.00010-0>.
- [6] B. Mackenzie and M. A. Hediger. “SLC11 family of H⁺-coupled metal-ion transporters NRAMP1 and DMT1”. In: *Pflügers Archiv.* 447 (2004), pp. 571–579. DOI: <https://doi.org/10.1007/s00424-003-1141-9>.
- [7] A. T. Bozzi and R. Gaudet. “Molecular Mechanism of Nramp-Family Transition Metal Transport”. In: *J. Mol. Biol.* 433.16 (2021), p. 166991. DOI: <https://doi.org/10.1016/j.jmb.2021.166991>.
- [8] X. Z. Chen et al. “Yeast SMF1 Mediates H⁺-coupled Iron Uptake with Concomitant Uncoupled Cation Currents”. In: *J. Biol. Chem.* 274.49 (1999), pp. 35089–35094. DOI: <https://doi.org/10.1074/jbc.274.49.35089>.
- [9] B. Mackenzie et al. “Functional properties of multiple isoforms of human divalent metal-ion transporter 1 (DMT1)”. In: *Biochem. J.* 403.1 (2007), pp. 59–69. DOI: <https://doi.org/10.1042/BJ20061290>.
- [10] J. Pujol-Giménez, M. A. Hediger, and G. Gyimesi. “A novel proton transfer mechanism in the SLC11 family of divalent metal ion transporters”. In: *Sci. Rep.* 7.1 (2017), p. 6194. DOI: [10.1038/s41598-017-06446-y](https://doi.org/10.1038/s41598-017-06446-y).
- [11] S. Tandy et al. “Nramp2 Expression Is Associated with pH-dependent Iron Uptake across the Apical Membrane of Human Intestinal Caco-2 Cells”. In: *J. Biol. Chem.* 275.2 (2000), pp. 1023–1029. DOI: <https://doi.org/10.1074/jbc.275.2.1023>.
- [12] Y. Nevo and N. Nelson. “The Mutation F227I Increases the Coupling of Metal Ion Transport in DCT1”. In: *J. Biol. Chem.* 279.51 (2004), pp. 53056–53061. DOI: <https://doi.org/10.1074/jbc.M408398200>.

- [13] A. Shawki and B. Mackenzie. “Interaction of calcium with the human divalent metal-ion transporter-1”. In: *Biochem. Biophys. Res. Commun.* 393.3 (2010), pp. 471–475. DOI: <https://doi.org/10.1016/j.bbrc.2010.02.025>.
- [14] A. Sacher, A. Cohen, and N. Nelson. “Properties of the mammalian and yeast metal-ion transporters DCT1 and Smf1p expressed in *Xenopus laevis* oocytes”. In: *J. Exp. Biol.* 204.6 (2001), pp. 1053–1061. DOI: <https://doi.org/10.1242/jeb.204.6.1053>.
- [15] N. Nelson, A. Sacher, and H. Nelson. “The significance of molecular slips in transport systems”. In: *Nat. Rev. Mol. Cell. Biol.* 3 (2002), pp. 876–881. DOI: <https://doi.org/10.1038/nrm955>.
- [16] S. Buracco et al. “*Dictyostelium* Nramp1, which is structurally and functionally similar to mammalian DMT1 transporter, mediates phagosomal iron efflux”. In: *J. cell sci.* 128.17 (2015), pp. 3304–3316. DOI: <https://doi.org/10.1242/jcs.173153>.
- [17] V. Picard et al. “Nramp 2 (DCT1/DMT1) Expressed at the Plasma Membrane Transports Iron and Other Divalent Cations into a Calcein-accessible Cytoplasmic Pool”. In: *J. Biol. Chem.* 275.46 (2000), pp. 35738–35745. DOI: <https://doi.org/10.1074/jbc.M005387200>.
- [18] P. Courville et al. “Determination of Transmembrane Topology of the *Escherichia coli* Natural Resistance-associated Macrophage Protein (Nramp) Ortholog”. In: *J. Biol. Chem.* 279.5 (2004), pp. 3318–3326. DOI: <https://doi.org/10.1074/jbc.M309913200>.
- [19] P. Courville et al. “Solute Carrier 11 Cation Symport Requires Distinct Residues in Transmembrane Helices 1 and 6”. In: *J. Biol. Chem.* 283.15 (2008), pp. 9651–9658. DOI: <https://doi.org/10.1074/jbc.M709906200>.
- [20] R. Chaloupka et al. “Identification of Functional Amino Acids in the Nramp Family by a Combination of Evolutionary Analysis and Biophysical Studies of Metal and Proton Cotransport in Vivo”. In: *Biochem.* 44.2 (2005), pp. 726–733. DOI: 10.1021/bi048014v.
- [21] W. Lan et al. “A facile transport assay for H⁺ coupled membrane transport using fluorescence probes”. In: *Anal. Methods* 4.1 (2012), pp. 44–46. DOI: <https://doi.org/10.1039/C1AY05549F>.
- [22] I. A. Ehrnstorfer et al. “Structural and mechanistic basis of proton-coupled metal ion transport in the SLC11/NRAMP family”. In: *Nature Commun.* 8.1 (2017), p. 14033. DOI: 10.1038/ncomms14033.
- [23] A. T. Bozzi et al. “Structures in multiple conformations reveal distinct transition metal and proton pathways in an Nramp transporter”. In: *eLife* 8 (2019), e41124. DOI: 10.7554/eLife.41124.
- [24] A. T. Bozzi et al. “Unique structural features in an Nramp metal transporter impart substrate-specific proton cotransport and a kinetic bias to favor import”. In: *J. Gen. Physiol.* 151.12 (2019), pp. 1413–1429. DOI: <https://doi.org/10.1085/jgp.201912428>.
- [25] I. A. Ehrnstorfer et al. “Crystal structure of a SLC11 (NRAMP) transporter reveals the basis for transition-metal ion transport”. In: *Nat. Struct. Mol. Biol.* 21.11 (2014), pp. 990–996. DOI: 10.1038/nsmb.2904.

- [26] A. T. Bozzi et al. "Crystal Structure and Conformational Change Mechanism of a Bacterial Nramp-Family Divalent Metal Transporter". In: *Structure* 24.12 (2016), pp. 2102–2114. DOI: <https://doi.org/10.1016/j.str.2016.09.017>.
- [27] A. Yamashita et al. "Crystal structure of a bacterial homologue of Na⁺/Cl[−]-dependent neurotransmitter transporters". In: *Nature* 437.7056 (2005), pp. 215–223. DOI: <https://doi.org/10.1038/nature03978>.
- [28] A. Vastermark et al. In: *Proteins: Structure, Function, and Bioinformatics* 82.10 (2014), pp. 2797–2811. DOI: <https://doi.org/10.1002/prot.24643>.
- [29] W. Humphrey, A. Dalke, and K. Schulten. "VMD: Visual molecular dynamics". In: *J. Mol. Graph.* 14.1 (1996), pp. 33–38. DOI: [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [30] *VMD Visual Molecular Dynamics*. URL: <https://www.ks.uiuc.edu/Research/vmd/>.
- [31] *BioRender*. URL: <https://biorender.com/>.
- [32] H. M. Berman et al. "The Protein Data Bank". In: *Nucl. Acids Res.* 28.1 (2000), pp. 235–242. DOI: <https://doi.org/10.1093/nar/28.1.235>.
- [33] H. Berman, K. Henrick, and H. Nakamura. "Announcing the worldwide Protein Data Bank". In: *Nat. Struct. Mol. Biol.* 10 (2003), p. 980. DOI: <https://doi.org/10.1038/nsb1203-980>.
- [34] *RCSB, PDB Protein Data Bank*. URL: <https://www.rcsb.org/>.
- [35] J. C. Gordon et al. "H++: a server for estimating pKa s and adding missing hydrogens to macromolecules". In: *Nucl. Acids Res.* 33.2 (2005), W368–W371. DOI: <https://doi.org/10.1093/nar/gki464>.
- [36] *H++*. URL: <http://newbiophysics.cs.vt.edu/H++/>.
- [37] M. A. Lomize et al. "OPM database and PPM web server: resources for positioning of proteins in membranes". In: *Nucl. Acids Res.* 40.D1 (2012), pp. D370–D376. DOI: <https://doi.org/10.1093/nar/gkr703>.
- [38] *Orientations of Proteins in Membranes (OPM) database*. URL: <https://opm.phar.umich.edu/>.
- [39] M. M. Ghahremanpour et al. "MemBuilder: a web-based graphical interface to build heterogeneously mixed membrane bilayers for the GROMACS biomolecular simulation program". In: *Bionformatics* 30.3 (2014), pp. 439–441. DOI: <https://doi.org/10.1093/bioinformatics/btt680>.
- [40] *MemBuilder II*. URL: <http://bioinf.modares.ac.ir/software/mb2/builder.php>.
- [41] R. Cao et al. "Role of Extracellular Loops and Membrane Lipids for Ligand Recognition in the Neuronal Adenosine Receptor Type 2A: An Enhanced Sampling Simulation Study". In: *Molecules* 23.10 (2018), p. 2616. DOI: [10.3390/molecules23102616](https://doi.org/10.3390/molecules23102616).
- [42] T. E. Andreoli et al. *Physiology of Membrane Disorders*. Boston, MA: Springer, 1986. ISBN: 978-1-4613-2097-5. DOI: <https://doi.org/10.1007/978-1-4613-2097-5>.

- [43] M. G. Wolf et al. “g_membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation”. In: *J. Comput. Chem.* 31.11 (2010), pp. 2169–2174. DOI: <https://doi.org/10.1002/jcc.21507>.
- [44] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. “GROMACS: A message-passing parallel molecular dynamics implementation”. In: *Comp. Phys. Commun.* 91.1 (1995), pp. 43–56. DOI: [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E).
- [45] D. van Der Spoel et al. “GROMACS: Fast, flexible, and free”. In: *J. Comput. Chem.* 26.16 (2005), pp. 1701–1718. DOI: <https://doi.org/10.1002/jcc.20291>.
- [46] M. J. Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1-2 (2015), pp. 19–25. DOI: <https://doi.org/10.1016/j.softx.2015.06.001>.
- [47] K. Lindorff-Larsen et al. “Improved side-chain torsion potentials for the Amber ff99SB protein force field”. In: *Proteins: Structure, Function, and Bioinformatics* 78.8 (2010), pp. 1950–1958. DOI: <https://doi.org/10.1002/prot.22711>.
- [48] C. J. Dickson et al. “Lipid14: The Amber Lipid Force Field”. In: *J. Chem. Theory Comput.* 10.2 (2014), pp. 865–879. DOI: [10.1021/ct4010307](https://doi.org/10.1021/ct4010307).
- [49] W. L. Jorgensen et al. “Comparison of simple potential functions for simulating liquid water”. In: *J. Chem. Phys.* 79.2 (1983), pp. 926–935. DOI: [10.1063/1.445869](https://doi.org/10.1063/1.445869).
- [50] I. S. Joung and T. E. Cheatham III. “Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations”. In: *J. Phys. Chem. B* 112.30 (2008), pp. 9020–9041. DOI: [10.1021/jp8001614](https://doi.org/10.1021/jp8001614).
- [51] B. Hess et al. “LINCS: A linear constraint solver for molecular simulations”. In: *J. Comput. Chem.* 18.12 (1997), pp. 1463–1472. DOI: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- [52] H. G. Petersen. “Accuracy and efficiency of the particle mesh Ewald method”. In: *J. Chem. Phys.* 103.9 (1995), pp. 3668–3679. DOI: [10.1063/1.470043](https://doi.org/10.1063/1.470043).
- [53] T. Darden, D. York, and L. Pedersen. “Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems”. In: *J. Chem. Phys.* 98.12 (1993), pp. 10089–10092. DOI: [10.1063/1.464397](https://doi.org/10.1063/1.464397).
- [54] M. Parrinello and A. Rahman. “Polymorphic transitions in single crystals: A new molecular dynamics method”. In: *J. Appl. Phys.* 52.12 (1981), pp. 7182–7190. DOI: [10.1063/1.328693](https://doi.org/10.1063/1.328693).
- [55] M. Parrinello, A. Rahman, and P. Vashishta. “Structural Transitions in Superionic Conductors”. In: *Phys. Rev. Lett.* 50 (14 1983), pp. 1073–1076. DOI: [10.1103/PhysRevLett.50.1073](https://doi.org/10.1103/PhysRevLett.50.1073).
- [56] S. Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. In: *J. Chem. Phys.* 81.1 (1984), pp. 511–519. DOI: [10.1063/1.447334](https://doi.org/10.1063/1.447334).
- [57] S. Nosé. “A molecular dynamics method for simulations in the canonical ensemble”. In: *Mol. Phys.* 52.2 (1984), pp. 255–268. DOI: [10.1080/00268978400101201](https://doi.org/10.1080/00268978400101201).

- [58] W. G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Phys. Rev. A* 31 (3 1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695.
- [59] C. R. Søndergaard et al. “Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values”. In: *J. Chem. Theory Comput.* 7.7 (2011), pp. 2284–2295. DOI: 10.1021/ct200133y.
- [60] M. H. M. Olsson et al. “PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions”. In: *J. Chem. Theory Comput.* 7.2 (2011), pp. 525–537. DOI: 10.1021/ct100578z.
- [61] *PropKa 3.0*. URL: <https://github.com/jensengroup/propka>.
- [62] N. Michaud-Agrawal et al. “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations”. In: *J. Comput. Chem.* 32.10 (2011), pp. 2319–2327. DOI: <https://doi.org/10.1002/jcc.21787>.
- [63] R. J. Gowers et al. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations”. In: *Proceedings of the 15th Python in Science Conference*. Ed. by Sebastian Benthall and Scott Rostrup. 2016, pp. 98–105. DOI: 10.25080/Majora-629e541a-00e.
- [64] *MDAnalysis Water Bridge*. URL: https://docs.mdanalysis.org/1.1.1/documentation_pages/analysis/wbridge_analysis.html.
- [65] *Hydrogen bonds — GROMACS 2021.4 documentation*. URL: <https://manual.gromacs.org/documentation/current/reference-manual/analysis/hydrogen-bonds.html>.
- [66] *gmx sasa*. URL: <https://manual.gromacs.org/documentation/current/onlinehelp/gmx-sasa.html>.
- [67] D. Bashford. “Macroscopic electrostatic models for protonation states in proteins”. In: *Front. Biosci. (Landmark Ed.)* 9.2 (2004), pp. 1082–1099. DOI: 10.2741/1187.
- [68] D. Bashford and M. Karplus. “pKa’s of ionizable groups in proteins: atomic detail from a continuum electrostatic model”. In: *Biochem.* 29.44 (1990), pp. 10219–10225. DOI: 10.1021/bi00496a010.
- [69] M. R. Gunner et al. “Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations”. In: *Biochim. Biophys. Acta* 1757.8 (2006), pp. 942–968. DOI: <https://doi.org/10.1016/j.bbabi.2006.06.005>.
- [70] J. M. Antosiewicz and D. Shugar. “Poisson–Boltzmann continuum-solvation models: applications to pH-dependent properties of biomolecules”. In: *Mol. BioSyst.* 7 (11 2011), pp. 2923–2949. DOI: 10.1039/C1MB05170A.
- [71] R. Anandakrishnan, B. Aguilar, and A. V. Onufriev. “H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations”. In: *Nucl. Acids Res.* 40 (W1 2012), W537–W541. DOI: <https://doi.org/10.1093/nar/gks375>.

- [72] G. Miclotte, K. Martens, and J. Fostier. “Computational assessment of the feasibility of protonation-based protein sequencing”. In: *PLOS ONE* 15.9 (2020), pp. 1–22. DOI: 10.1371/journal.pone.0238625.

Conclusions and outlook

Computational methods have nowadays become indispensable tools able to complete and improve interpretations of physical processes made by experimental techniques as they can provide crucial details regarding the molecular level mechanisms underlying such processes. This synergistic approach between experimental and computational techniques is particularly important in the investigation of biological processes, in which molecular simulations represent a crucial tool to furnish molecular level explanations to evidence gathered from experimental studies, or refine them, as seen in this thesis.

A combination of classical and QM/MM *ab initio* Born-Oppenheimer and Car-Parrinello molecular dynamics simulations has been used to refine the coordination sphere of Mn^{2+} in the binding site of ScaDMT. This represents to date the only available crystal structure of a member of the Nramp family in a substrate-bound inward-facing conformation, in which however the coordination sphere of the metal binding site has only been partially resolved. Our results suggest a nearly perfect octahedral coordination sphere for Mn^{2+} . The four amino acidic residues predicted to coordinate the metal ion in the crystallographic structure (namely Asp49, Asn52, Ala223 and Met226) are confirmed by our findings. However, they undergo substantial rearrangements with respect to the highly distorted coordination geometry found in the X-ray structure and occupy the four equatorial sites of the octahedral coordination polyhedron whereas, the two axial sites left empty in the experimental structure are filled with two water molecules that presumably escaped the crystallographic characterization. Our simulations also exclude the participation of Gln389 and Gly46 to the coordination of Mn^{2+} as proposed in previous theoretical models and a possible coordination of Asp49 as a bidentate ligand. The nature of the coordination of the metal binding site identified in this thesis has been further strengthened by the analysis of the hydrogen bond network connecting the first and the second coordination shell.

The identification of the complete composition of the Mn^{2+} coordination sphere of ScaDMT represents a key information that could serve as starting point for the study of the mechanism of metal ion release and the concomitant conformational changes. A similar study could also be applied to the characterisation of the complete coordination shell of other divalent transition metal ion, which would constitute another important step forward in the comprehension of metal coordination in Nramp transporters. This could for instance help to

understand whether the binding mode is common for all the substrates or the binding site is able to modulate itself depending on the nature of the ion.

In the second part of this thesis work, the respectively outward and inward facing conformations of the EcoDMT and ScaDMT proteins have been investigated using classical molecular dynamics (MD) simulations in their apo forms and with their histidine residue (His236/228 in EcoDMT and ScaDMT numbering, respectively) in both unprotonated and protonated states. Different analyses of the obtained trajectories have been performed in order to provide new molecular level insights into the proton transport mechanism occurring in Nramp transporters. We focused in particular on the role played by four amino acidic residues (Asp51/49, Glh129/127, His236/228 and Asp126/124, in apo-EcoDMT and apo-ScaDMT numbering, respectively) belonging to the conserved hydrophilic network and already indicated as potentially involved in the proton transport process. The analyses performed on the four selected residues allowed us to indicate Glh129/127 and Asp51/49 (EcoDMT and ScaDMT numbering) as the likely primary proton acceptor initiating the proton transport process and to indicate the potential pathway followed by the protons during their transport. According to our findings, the proton is presumably transferred from Glh129/127, undergoing rapid and consecutive deprotonation/protonation events, to Asp51/49. Then, after the conformational switch from the OFC to the IFC (presumably happening after Asp51 protonation) the proton is transferred to Asp126/124 via His236/228 on TM6b. From here the proton is finally released into the cytoplasm presumably through the narrow hydrophilic side channel via Asp124. However, a direct release of the proton from His228 through the wide, inward-directed vestibule connecting the binding site with the cellular cytoplasm can not be completely excluded.

In order to confirm or eventually refine the proton transport model proposed here it would be undoubtedly useful to further characterize the proton transport in the presence of the metal ion substrate. The metal ion could in fact strongly affect the pKa values of the investigated amino acidic residues providing therefore additional insights into the proton-metal co-transport involving both the outward and inward facing conformational states. Furthermore, it would be interesting to identify the event triggering the conformational switch using enhanced sampling techniques. In particular, it seems that the protonation of the Asp residue located on TM1 that is directly involved in metal ion coordination constitutes a key event that could trigger the conformational change from the OFC to the IFC and the subsequent release of the metal ion.

A Supporting Information - Coordination of Mn(II) in SLC11/NRAMP transporters uncovered by QM/MM simulations

Table A.1: Energy values (in a.u.) and energy differences (in kcal mol⁻¹) of a simple model of the experimental ScaDMT active site in the sextet (S = 5/2), quartet (S = 3/2) and doublet (S = 1/2) multiplicity states, as obtained from DFT calculations. The tested model contains the four amino acidic residues coordinating Mn²⁺ (Asp49, Asn52, Ala223, Met226) properly saturated on the -NH and -CO groups and the Mn²⁺ ion itself, for a total of 62 atoms. DFT calculations were performed using the hybrid B3LYP exchange-correlation functional [1–3], the 6-31G(d,p) basis set [4–9] and the Gaussian 16 software package [10].

Table A.2: Absolute energies (in a.u.)

Multiplicity	E(UB3LYP)(a.u.)
sextet	-2978.368
quartet	-2978.337
doublet	-2978.314

Table A.3: Energy differences (in kcal mol⁻¹).

	ΔE (kcal mol ⁻¹)
(E _S - E _Q)	-19.335
(E _Q - E _D)	-14.227

BOND	EQUILIBRATION (600ns)	PRODUCTION (1000ns)	TOTAL (1600ns)
Mn(II)-Asp49(O ^{δ2})	2.16 ± 0.04	2.18 ± 0.04	2.16 ± 0.04
Mn(II)-Asn52(O ^{δ1})	2.70 ± 0.07	2.72 ± 0.07	2.70 ± 0.07
Mn(II)-Ala223(O)	2.22 ± 0.04	2.21 ± 0.04	2.22 ± 0.04
Mn(II)-Met226(S ^δ)	2.70 ± 0.06	2.70 ± 0.06	2.70 ± 0.06
Mn(II)-Gln389(O ^{ε1})	4.66 ± 0.22	4.92 ± 0.14	4.76 ± 0.23
Mn(II)-Gly46(O)	5.33 ± 0.33	5.24 ± 0.34	5.30 ± 0.34
Mn(II)-Asp49(O ^{δ1})	4.19 ± 0.05	4.19 ± 0.05	4.19 ± 0.05

Table A.4: Average distances between Mn²⁺ and Asp49(O^{δ2}), Asn52(O^{δ1}), Ala223(O), Met226(S^δ), Gln389(O^{ε1}), Gly46(O) and Asp49(O^{δ1}) calculated over the equilibration phase of classical MD (600ns) (first column), the production phase (1000ns) (second column) and the whole classical MD trajectory (equilibration + production) (1600ns) (third column).

Supporting Information - Coordination of Mn(II) in SLC11/NRAMP transporters uncovered Chapter A

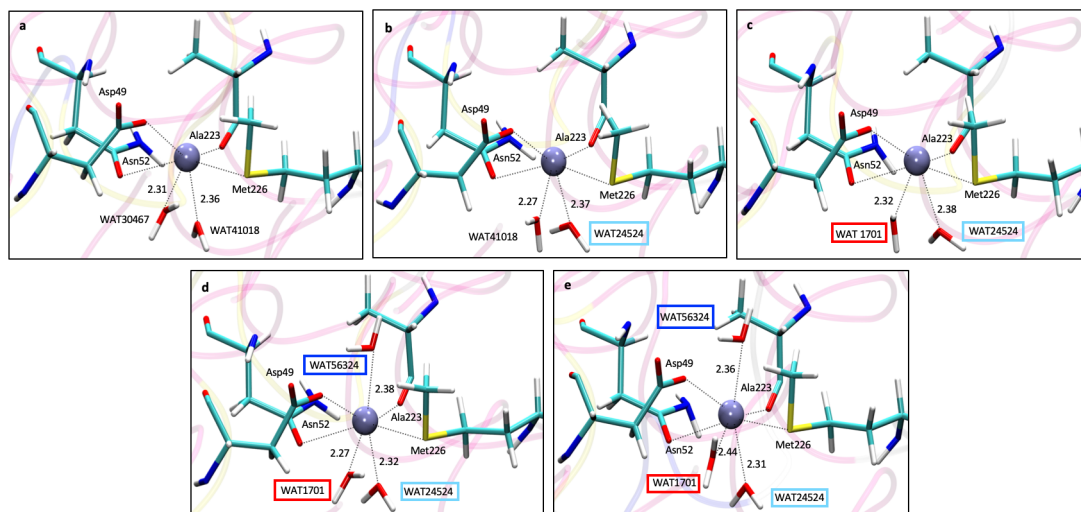


Figure A.1: Chronological sequence of access of water molecules inside the Mn^{2+} coordination sphere during the equilibration phase of classical MD. (a) At the very beginning of the classical MD simulation ($t = 0$ ps) two water molecules (WAT30467 and WAT41018) are already close to Mn^{2+} . This seems reasonable considering that ScaDMT in its inward-facing conformation is in direct contact with the water composed cytoplasmatic vestibule. The short minimization dynamics allowed water molecules to diffuse inside the cytoplasmatic vestibule and approach the Mn^{2+} binding site. However, in these first moments of the thermalization process water molecules approach Mn^{2+} only transiently; (b) $t = 686$ ps: WAT24524 (light-blue box) enters Mn^{2+} coordination sphere and remains there until it exchanges with WAT3067, at around half of the production run; (c) $t = 220200$ ps: WAT1701 (red box) enters Mn^{2+} coordination sphere and keeps its positions until the end of the production run; (d) $t = 317050$ ps: WAT56324 (blue box) enters Mn^{2+} coordination sphere and keeps its position until the end of the production run; (e) end of the equilibration phase of classical MD ($t = 600000$ ps), WAT24524 (light blue box), WAT1701 (red box) and WAT56324 (blue box) coordinate Mn^{2+} together with the four amino acidic residues Asp49, Asn52, Ala223 and Met226.

BOND	DISTANCE (Å)
Mn(II)-WAT24524(O)	2.35 ± 0.08
Mn(II)-WAT1701(O)	2.35 ± 0.09
Mn(II)-WAT56324(O)	2.35 ± 0.10
Mn(II)-WAT3067(O)	2.37 ± 0.09

Table A.5: Average distances between Mn^{2+} and the four water molecules participating to the coordination sphere during the classical MD simulation. Each average value has been calculated over the total permanence period of that water molecule in the coordination sphere. Mn^{2+} -WAT24524 average value has been calculated by averaging distance values between $t_0 = 686$ ps and $t_F = 1088100$ ps; Mn^{2+} -WAT1701 average value has been calculated by averaging distance values between $t_0 = 220200$ ps and $t_F = 1600000$ ps; Mn^{2+} -WAT56324 average value has been calculated by averaging distance values between $t_0 = 317050$ ps and $t_F = 1600000$ ps; Mn^{2+} -WAT3067 average value has been calculated by averaging distance values between $t_0 = 1088200$ ps and $t_F = 1600000$ ps.

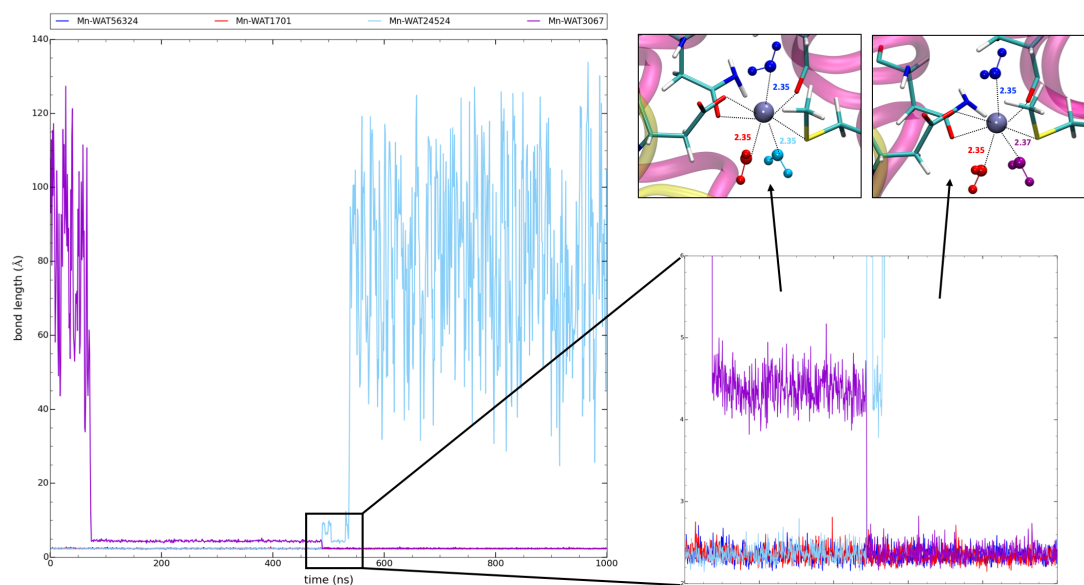


Figure A.2: Distances between Mn²⁺ and WAT56324, WAT1701, WAT24524 and WAT3067 during the production run (1000 ns) of classical MD simulation, represented as blue, red, light blue and purple lines respectively (**left panel**). The exchange between WAT24524 and WAT3067 is highlighted by the black box. Magnification of the waters exchange area (**right bottom panel**). Representation of the Mn²⁺ coordination spheres before and after the water molecules exchange (**top right panels**). Asp49, Asn52, Ala223 and Met226 ligands are represented in licorice style while Mn²⁺ is represented as a gray sphere. The water molecules coordinating the metal ion are represented as ball and sticks model in the same colors with which their distances are represented in the graphs. Average values of the Mn²⁺-waters distances are also given.

References

- [1] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange". In: *J. Chem. Phys.* 98.7 (1993), pp. 5648–5652. DOI: 10.1063/1.464913.
- [2] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". In: *Phys. Rev. B* 37 (2 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.
- [3] S. H. Vosko, L. Wilk, and M. Nusair. "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis". In: *Can. J. Phys.* 58.8 (1980), pp. 1200–1211. DOI: <https://doi.org/10.1139/p80-159>.
- [4] R. Ditchfield, W. J. Hehre, and J. A. Pople. "Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules". In: *J. Chem. Phys.* 54.2 (1971), pp. 724–728. DOI: 10.1063/1.1674902.
- [5] P. C. Hariharan and J. A. Pople. "The influence of polarization functions on molecular orbital hydrogenation energies". In: *Theor. chim. acta* 28 (1973), pp. 213–222. DOI: <https://doi.org/10.1007/BF00533485>.
- [6] W. J. Hehre, R. Ditchfield, and J. A. Pople. "Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital

- Studies of Organic Molecules". In: *J. Chem. Phys.* 56.5 (1972), pp. 2257–2261. DOI: 10.1063/1.1677527.
- [7] M. S. Gordon et al. "Self-consistent molecular-orbital methods. 22. Small split-valence basis sets for second-row elements". In: *J. Am. Chem. Soc.* 104.10 (1982), pp. 2797–2803. DOI: 10.1021/ja00374a017.
- [8] M. M. Francl et al. "Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements". In: *J. Chem. Phys.* 77.7 (1982), pp. 3654–3665. DOI: 10.1063/1.444267.
- [9] V. A. Rassolov et al. "6-31G* basis set for atoms K through Zn". In: *J. Chem. Phys.* 109.4 (1998), pp. 1223–1229. DOI: 10.1063/1.476673.
- [10] *Gaussian 16*. URL: <https://gaussian.com/>.

B Supporting Information - Computational studies provide new insights into Nramps proton transport

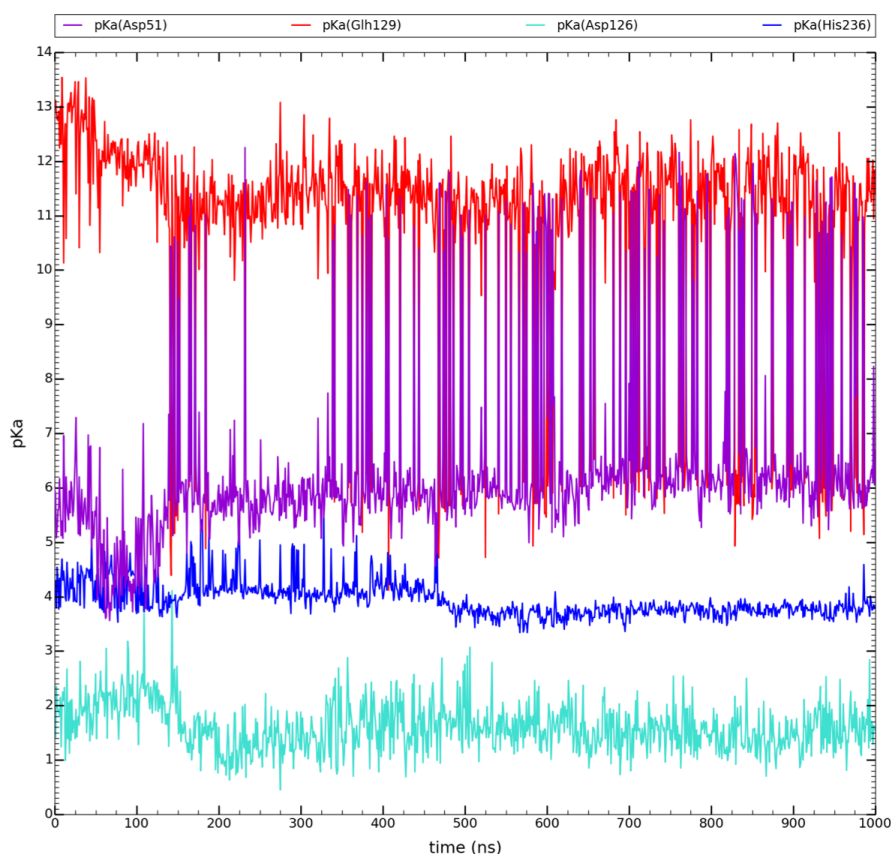


Figure B.1: pKa values of residues Asp51 (violet), Glh129 (red), Asp126 (turquoise) and His236 (blue) during apo-EcoDMT (OFC) trajectory (1 μ s = 1000ns) with His236 in its protonated form. Time is reported in ns. Residue numbering from apo-EcoDMT.

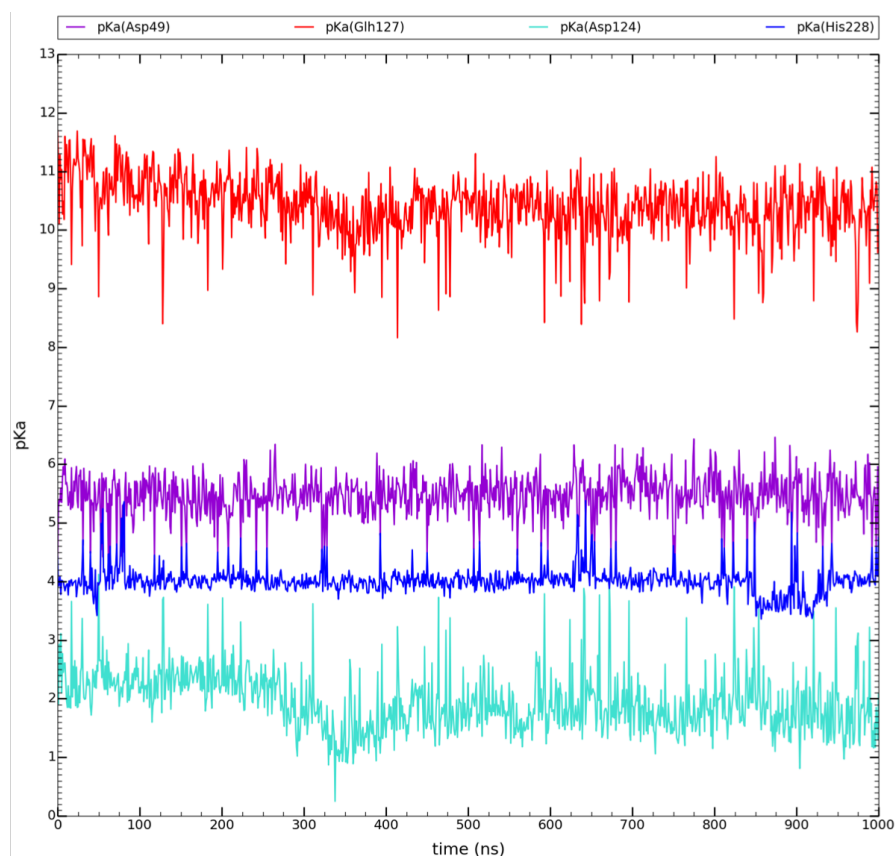


Figure B.2: pKa values of residues Asp49 (violet), Glh127 (red), Asp124 (turquoise) and His228 (blue) during apo-ScaDMT (IFC) trajectory ($1 \mu s = 1000 ns$) with His228 in its protonated form. Time is reported in ns. Residue numbering from apo-ScaDMT.

residue	$pK_{a,AV} \pm \sigma(pKa)$	ΔpKa
Asp51	6.51 ± 1.87	8.69
Glh129	10.80 ± 1.99	9.42
Asp126	1.61 ± 0.44	3.66
His236	3.94 ± 0.32	2.23
Asp49	5.46 ± 0.36	2.39
Glh127	10.40 ± 0.50	3.54
Asp124	1.95 ± 0.51	3.87
His228	4.02 ± 0.24	2.02

Table B.1: Average pKa ($pK_{a,AV}$) values with their standard deviations ($\sigma(pKa)$) and the (ΔpKa) for residues Asp51, Glh129, Asp126 and His236 of apo-EcoDMT (lines 1-4) and for residues Asp49, Glh127, Asp124 and His228 (lines 5-8). Each residue is indicated in the same color used in Fig. B.1 and B.2.

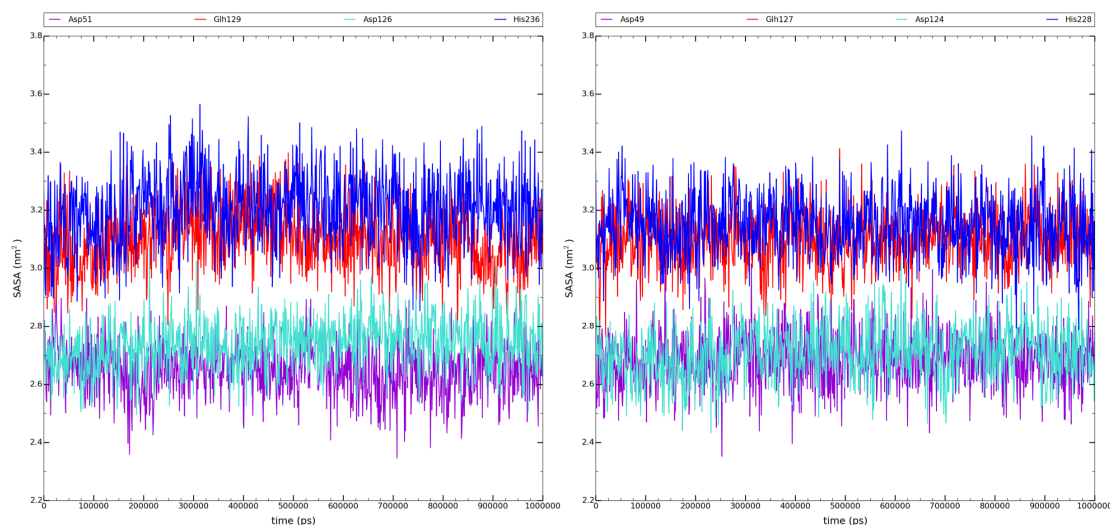


Figure B.3: (left) Asp51, Glh129, Asp126 and His236 Solvent Accessible Surface Area (SASA) values over time ($1 \mu\text{s} = 1000 \text{ ns}$) in apo-EcoDMT (OFC) trajectory with His236 protonated. (right) Asp49, Glh127, Asp124 and His228 SASA values over time in apo-ScaDMT (IFC) trajectory with His228 protonated. SASA values are given in nm^2 while time is reported in ps.

Amino Acid Pair	apo-EcoDMT-His236p
Asp51-His236	92.0%
Asp51-Glh129	58.6%
His236-Asp126	63.2%

Table B.2: Percentage of existence (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the apo-EcoDMT OFC MD simulations (His236 protonated).

Amino Acid Pair	apo-ScaDMT-His228p
Asp49-His228	94.0%
Asp49-Glh127	3.3%
His228-Asp124	0.9%

Table B.3: Percentage of existence of water bridges (given as fraction of the total number of frames contained in the trajectory) for each of the three investigated couples of residues in the apo-ScaDMT IFC MD simulation (His228 protonated).

Amino Acid Pair	apo-EcoDMT-His236u
Asp51(C^γ)-His236($\text{N}^{\epsilon 2}$)	$7.45 \pm 1.04 \text{ \AA}$
Asp51(C^γ)-Glh129(C^δ)	$6.40 \pm 1.38 \text{ \AA}$
His236($\text{N}^{\epsilon 2}$)-Asp126(C^γ)	$10.15 \pm 0.70 \text{ \AA}$

Table B.4: Average distances and corresponding standard deviations (in \AA) between the residues forming the pairs analysed in the "water bridge" analysis for apo-EcoDMT with His236 unprotonated.

Amino Acid Pair	apo-ScaDMT-His228u
Asp49(C^γ)-His228($\text{N}^{\delta 1}$)	$7.93 \pm 0.56 \text{ \AA}$
Asp49(C^γ)-Glh127(C^δ)	$6.10 \pm 0.37 \text{ \AA}$
His228($\text{N}^{\delta 1}$)-Asp124(C^γ)	$9.62 \pm 0.66 \text{ \AA}$

Table B.5: Average distances and corresponding standard deviations (in \AA) between the residues forming the pairs analysed in the "water bridge" analysis for apo-ScaDMT with His228 unprotonated.

Supporting Information - Computational studies provide new insights into Nramps proton transport
Chapter B

Amino Acid Pair	apo-EcoDMT-His236p
Asp51(C γ)-His236(N ϵ^2)	$7.57 \pm 0.81 \text{ \AA}$
Asp51(C γ)-His236(N δ^1)	$7.78 \pm 0.58 \text{ \AA}$
Asp51(C γ)-Gln129(C δ)	$4.46 \pm 0.18 \text{ \AA}$
His236(N ϵ^2)-Asp126(C γ)	$10.86 \pm 1.01 \text{ \AA}$
His236(N δ^1)-Asp126(C γ)	$10.06 \pm 0.61 \text{ \AA}$

Table B.6: Average distances and corresponding standard deviations (in \AA) between the residues forming the pairs analysed in the "water bridge" analysis for apo-EcoDMT with His236 protonated.

Amino Acid Pair	apo-ScaDMT-His228p
Asp49(C γ)-His228(N δ^1)	$7.02 \pm 0.32 \text{ \AA}$
Asp49(C γ)-His228(N ϵ^2)	$7.47 \pm 0.35 \text{ \AA}$
Asp49(C γ)-Gln127(C δ)	$8.28 \pm 0.41 \text{ \AA}$
His228(N δ^1)-Asp124(C γ)	$9.82 \pm 0.33 \text{ \AA}$
His228(N ϵ^2)-Asp124(C γ)	$10.02 \pm 0.46 \text{ \AA}$

Table B.7: Average distances and corresponding standard deviations (in \AA) between the residues forming the pairs analysed in the "water bridge" analysis for apo-ScaDMT with His228 protonated.

Maria Letizia Merlini

maria.merlini@epfl.ch marialetiziamerlini@gmail.com • +41 078 652 2906

Citizenship: Italian

Research interests

computational chemistry and biochemistry, biological systems, theoretical chemistry

Education

- 2016 – Present **Ecole Polytechnique Fédérale de Lausanne** – Lausanne, Switzerland
PhD in Computational Chemistry and Biochemistry
Advisor: Prof. Ursula Röthlisberger
Computational Studies of the Proton-Coupled Metal Ion Transport in the SLC11/NRAMP Family of Transporters.
- 2013 – 2016 **Università degli Studi di Perugia** – Perugia, Italy
Master of Science in Chemical Science
Advisor: Prof. Paola Belanzoni
Theoretical study of a Bio-Inspired Mn(II) Complex with Catalase-like Antioxidant Activity.
- 2006 – 2013 **Università degli Studi di Perugia** – Perugia, Italy
Bachelor Degree in Chemistry
Advisor: Prof. Paola Belanzoni
Theoretical study of a Iridium-based EDTA Catalyst for Water Oxidation.

Publications

- 2018 **Understanding the Catalase-Like Activity of a Bioinspired Manganese(II) Complex with a Pentadentate NSNSN Ligand Framework. A Computational Insight into the Mechanism**
Merlini M. L., Britovsek G. J. P., Swart M., Belanzoni P.
ACS Catalysis, 2018, 8 (4), 2944-2958

Teaching experience

Autumn 2019	Teaching assistant, MATH-111 : Linear Algebra (EPFL) Workload: 4hrs
Spring 2019	Teaching assistant, PREPA-031 (a): Mathematics 1a for Mise à Niveau (EPFL) Workload: 62 hrs
Spring 2018	Teaching assistant, PREPA-032(b): Mathematics 1b for Mise à Niveau (EPFL) Workload: 80 hrs
Autumn 2017	Teaching assistant, CH-163: Biooriented Chemical Chemistry (EPFL), Workload: 40 hrs
Autumn 2017	Teaching assistant, CH-311: Molecular and Cellular Biophysics 1 (EPFL), Workload: 45 hrs
Spring 2017	Teaching assistant, CH-222: Coordination Chemistry (EPFL) Workload: 72 hrs
Spring 2017	Teaching assistant, CH-359: Project of Computational Chemistry (EPFL), Woakload: 42 hrs

Schools and Conferences

July 2018	CECAM Workshop "Frontiers and Challenges of Computing Metals for Biochemical Medical and Technological Applications" (Paris) Poster presentation
-----------	---

May 2018	CECAM Workshop "Physiological Role of Ions in the Brain: Towards a Comprehensive View by Molecular Simulations" (Scuola Normale Superiore, Pisa) Poster presentation
May 2017	CECAM School "Hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) Approaches to biochemistry and Beyond" (EPFL, Lausanne)
January 2017	CECAM School "MolSim 2017: Understanding Molecular Simulations" (Amsterdam)

Professional memberships

2016 – Present	Swiss Chemical Society (SCS)
----------------	------------------------------

Technical skills

Computer administration

Mac OS, Linux, Windows

Programming languages

Python, Bash

Computational chemistry softwares

Gromacs, Amber, CPMD, Gaussian, ADF

Data analysis

Matplotlib

Visualization

VMD, Rasmol, PyMOL

Productivity softwares

Microsoft Word, Microsoft Office, L^AT_EX, Biorender

Languages

Italian (native), English (fluent), French (intermediate)

Other interests

Travelling, Drawing and painting, Reading