Rodrigo Miguel de Sousa Barriga Alves

Graduated in Biochemistry

Simulating substrate binding sites in the *S. aureus* Type II NADH Dehydrogenase

Dissertation to obtain a Master's Degree
in Biochemistry for Health

Supervisor: Manuel N. Melo, PhD, ITQB NOVA
Co-Supervisor: Manuela M. Pereira, PhD, FCUL

November 2022

Rodrigo Miguel de Sousa Barriga Alves

Graduated in Biochemistry

Simulating substrate binding sites in the *S. aureus* Type II NADH Dehydrogenase

Dissertation to obtain a Master's Degree
in Biochemistry for Health

Supervisor: Manuel N. Melo, PhD, ITQB NOVA
Co-Supervisor: Manuela M. Pereira, PhD, FCUL

Members of the jury:
President: Professor Doctor Pedro Manuel Marques Henriques Matias
Supervisor: Professor Doctor Manuel Nuno de Sousa Pereira Simões de Melo
Opponent: Professor Doctor Bruno Lourenço da Silva Víctor
Secretary: Professor Doctor Ana Maria Varela Coelho

Instituto de Tecnologia Química e Biológica António Xavier

November 2022

COPYRIGHT

Some results in this thesis will be used for a future article, yet to be published. That information will be mentioned at the beginning of each relevant Chapter.

Additionally, some results in this thesis have already been presented in poster communications at both national and international congresses as follows:

Encontro de Jovens Investigadores de Biologia Computacional Estrutural VIII: Barriga, R., Lotfi, M., Pereira, M. M. & N. Melo, M. P2 - Parameterization of nucleotide cofactors and metabolites for the Martini 3 Coarse-Grain force field. http://ejibce.github.io/images/abstracts_2021.pdf (December 2021, University of Coimbra).

2nd Meeting of Young Portuguese Biophysicists: Barriga, R., Lotfi, M., Pereira, M. M. & N. Melo, M. P35 - The analysis of *Staphylococcus aureus*' NDH-2 Coarse-Grain simulations might be the beginning to unveil a future therapeutical target. in Book of abstracts from the 2nd Meeting of Young Portuguese Biophysicists – 2022 Biophysics Festival. in (Zenodo, 2022). DOI:10.5281/ZENODO.6772805. https://zenodo.org/record/6772805#.Y0HRG3bMK3B (June 2022, University of Aveiro).

The Biochemistry Global Summit 25th IUBMB Congress 46th FEBS Congress 15th PABMB Congress, 2022: Barriga, R., Lotfi, M., Pereira, M. M. & N. Melo, M. P-06.3-001 - The analysis of *Staphylococcus aureus*' NDH-2 Coarse-Grain simulations might be the beginning to unveil a future therapeutical target. in vol. **12** 67–336 (Wiley, 2022). DOI: 10.1002/2211-5463.13440. https://febs.onlinelibrary.wiley.com/doi/10.1002/2211-5463.13440 (July 2022, Centro de Congressos de Lisboa).

2nd Chem & Biochem Students Meeting: Barriga, R., Lotfi, M., Pereira, M. M. & N. Melo, M. P49 - The analysis of *Staphylococcus aureus*' NDH-2 Coarse-Grain simulations might be the beginning to unveil a future therapeutical target. http://chembiochem02.campus.ciencias.ulisboa.pt/assets/files/Book_of_Abstracts.pdf (July 2022, University of Lisbon).

Some results were also presented in two oral communications: Jornadas Intercalares dos Mestrados (February 2022, Nova University of Lisbon) and the Martini Developers Meeting (November 2022, Malta).

Other activities related to this thesis or done during this time, such as Tutoring in Summer Science ITQB NOVA (July 2022), organisation of the 3D-BioInfo-PT workshop in ITQB NOVA (September 2022) and others can be found at: https://www.cienciavitae.pt//pt/381C-1072-CC11

You may find all the supplementary material in the following link: https://drive.google.com/drive/folders/1dVqtz8L0lzGae23aAR6nWLKy6-sALITj?usp=sharing

AGRADECIMENTOS

Nesta Secção vou tentar mencionar todes aqueles que, de uma forma ou outra, me apoiaram e, sem os quais, nunca teria chegado até aqui. Vou ainda, hipoteticamente, entregar um presente, que possa descrever eventualmente um pouco ou muito a pessoa.

Primeiramente, gostaria de expressar o meu enorme agradecimento aos meus orientadores, ao Doutor Manuel Melo e à Professora Doutora Manuela Pereira. Os meus orientadores não só desempenharam um papel fundamental no meu desenvolvimento cognitivo, mas como cívico e social, igualmente. Ao Manuel, se pudesse dar-lhe-ia um novo e requintado *cluster*, da mais recente tecnologia e com as mais novas funcionalidades. A ele, um agradecimento muito grande por me ter aturado imenso, por sempre me ter incentivado a participar nos mais diversos simpósios e por ter dito que sim, a muitas vezes que eu queria participar. À professora Manuela, entregar-lhe-ia a possibilidade de estudar um microrganismo totalmente novo e com uma cadeia respiratória totalmente nova, nunca antes vista, para estudar e se deliciar com o desconhecido. À professora, devo um bocadinho deste meu "bichinho" pela Bioenergética, já da licenciatura. Por sempre ter encarado com um sorriso as nossas aulas experimentais de manhã, com quem aprendi muito no laboratório e na escrita de relatórios.

Segundamente, aos meus colegas de laboratório, sem os quais esta tese teria sido impossível. É um trabalho de equipa, esta tese também é vossa! Em especial um grande abraço ao Gonçalo que, não só é meu colega, mas também meu confidente nas suas horas vagas. Ao Fernando, porque nunca encontrei alguém tão engraçado e com um sentido de humor tão peculiar. Ao Doutor Luís, um dia quero saber tanto como tu, és o maior! À Maryam, por alguns resultados daqui também serem dela e por me ter ajudado, especialmente no meu início. À Mariana, pela amizade e simpatia constantes e por sempre me ter explicado tudo com tanta serenidade. Ao Karl, pela paciência em todos os dias me aguentar e me ter ensinado tanto de Inglês, em tão pouco tempo! E foi pouco este tempo! À Raquel, agradeço por toda a sinceridade, humildade e conforto que sempre me foram dados. Ao Guilherme, o agradecimento por me ter ajudado com uma pequena parte da minha tese e a hipótese de o poder ter orientado. Cheguei ao laboratório do Manuel sem saber nada. Nem conhecimentos base de Linux sabia. Todos me acolheram da melhor maneira e, sinceramente, não existem palavras suficientes para descrever o meu profundo agradecimento. Ao Luís dava-lhe uns ténis de corrida novinhos em folha para ele brilhar no atletismo. À Mariana, um modelo 3D da proteína que ela quisesse feito por ela e impresso. Ao Gonçalo, o lote de todos os novos livros do ensino básico e secundário, exclusivos para o senhor professor. Ao Fernando, a grande oportunidade de se estrear num palco de *stand-up comedy*. Ao Karl, uma nova viagem a Portugal. Ao Guilherme, dar-lhe-ia entradas grátis durante um ano para assistir aos jogos de futebol do seu clube. À Maryam e à Raquel, mas também a todas as mulheres, mais liberdade de expressão, de escolha e mais voz para serem devidamente ouvidas. Tínhamos sempre a nosso lado os nossos irmãos e a eles agradeço pela enormíssima amizade. Aos meus amigos Rita, Susana, João e Bárbara, isto sem vocês não teria tido qualquer sentido. Rita, a ti dava-te um teletransporte direto de tua casa para o ITQB, porque sofres mesmo muito com o tempo que despendes nos transportes. À minha querida Susana, que partilhamos o nosso amor por animais, especialmente cães, entregar-lhe-ia um brinquedo que o cão dela não conseguisse estragar. Ao João, pelo empenho excecional, dava-lhe um computador novo, ele bem merece! À Bárbara, dava-lhe uma bonequinha igualzinha a ela, para ela se relembrar que a beleza interior e exterior muitas vezes se entrelaçam na maior das perfeições.

Aos meus professores da licenciatura Ana Paula Paiva, Patrícia Faísca, Susana Marinho, Estrela Jorge, António Ferreira e Tânia Morais e aos meus professores do mestrado Teresa Catarino, Ricardo Louro, Pedro Matias, Ana Coelho, Margarida Archer e Paula Videira, por me terem ensinado tanto e me terem despertado e continuar a despertar o "bichinho" das Ciências, sejam elas Física, Matemática, Química, Bioquímica, Imunologia, Bioenergética, etc. Agradeço ainda ao José Brito, pela dúvidas que me tirou! A eles desejo-lhes uma excelente continuação de carreira e se pudesse construir-lhes-ia uma nova faculdade totalmente remodelada para poderem dar aulas e prosseguir com as suas investigações. Ao meu professor de Matemática A do secundário José Joaquim Borges, o maior dos

obrigados por me ter ensinado de forma tão perfeita matemática e ter feito com que fosse a melhor disciplina que alguma vez tenha tido. Ao professor por ser tão meu amigo e tão devoto à sua profissão e por saber tanto de matemática. A ele dar-lhe-ia a possibilidade de escolher mudar algo atualmente no rumo da sua vida. O mesmo conferiria à minha professora de Biologia e Geologia, Teresa Catrau, não só pelo ensino, mas por toda a psicologia, envolvência nas nossas vidas enquanto diretora de turma e carinho por nós. Não menos importante, às minhas amigas Olga Chervonovska, Joana Pimenta e Inês Paccetti, um agradecimento por toda a partilha, entreajuda, diversão, companheirismo e confidência. À Olga se pudesse dava-lhe uma paz imediata no seu país de origem e o cessar da guerra. À Joana, uma trela que ela conseguisse puxar sem esforço o seu cão. À Inês, uma casa cheia de gatinhos com todo o amor que lhe pudessem conferir, assim como eu o faço diariamente. Aos amigos que ainda hoje trago da licenciatura, Inês Alves, Débora Castro, Jéssica Irina, José Pedro Dinis, Margarida Martins, Catarina Caldeira, Ana Sofia Augusto, Pedro Pereira, Matilde Barbosa, Duarte Breia Clemente, Pedro Fanica, André Silva, Duarte Almeida, Vanessa Costa, Neuza Salgado, Jéssica Cerqueira, Mariana Ribeiro, Cassandra Cardoso, Catarina Ferreira e Sofia Matias e tantos outros mais, por tanto, mas tanto… Por todas as magníficas memórias formadas, por aquelas que ainda se formarão. Vocês são do melhor, mesmo. Obrigado por me ampararem quando cheguei à faculdade e por estarem lá quando me ri, chorei, passei, chumbei,… Vocês conhecem-me verdadeiramente. A vocês uma viagem de férias todos juntos para nos divertirmos imenso como nos bons velhos tempos. Aos meus amigos Diogo Birra, Diogo Silva, André Luís, Catarina Silva (e restantes amigas do bar do ITQB) e Mafalda Mateus obrigado por todo o vosso excecional carinho e constante presença. A vocês concedo-vos a hipótese de um grande amor.

Aos meus amigos que vêm do secundário, Carolina Piano, João Simões, Mariana Antunes, Maria Nunes e Sara Dezerto, obrigado por me deixarem sentir que, pela primeira vez, estava integrado nalgum sítio. E, como vocês bem sabem, não há palavras para agradecer o que vocês fizeram por mim no meu período negro. O vosso suporte foi incondicional, como bem sabem. Em especial à Sarinha, a minha melhor amiga, já lá vão uns loucos quase 15 anos e a minha vida sem ti não teria sido a mesma coisa. Não há dúvidas que és a melhor amiga que alguém pode ter. Obrigado mesmo. A ti desejo-te que sejas muito, muito feliz e que sejas a maior chefe de clínica dentária do país. A ti e aos restantes, entrego-vos a possibilidade de um dia, em conjunto, termos uma casa de férias para todos nós! À minha anterior psicóloga Ana Galhardo Simões e à minha atual psicóloga, Catarina Portela, porque sem vocês nunca teria chegado até aqui e, porque graças a vocês, fui e vou enfrentando os meus medos, desejo que tenham a clínica com que sempre sonharam e que tenham mesmo muito sucesso. À minha família, em especial aos meus primos Filipe e Margarida Moura Reis, Rafaela Barriga por serem os meus segundos irmãos e por terem partilhado toda a infância comigo até chegarmos a esta etapa e que assim continue. Às minhas tias, Ana Barriga, Filipa Barriga, Cristina Barriga e tios, Carlos Quintas, Hélder Pereira e Rui Alves, e à minha avó Teresa Pereira, porque estão sempre lá, sempre presentes e isso nunca se esquece. A vocês, uma oportunidade de estarmos todos novamente reunidos mais uma vez. Aos meus atuais amigos, companheiros, caninos Yoshi, Kika e Amy, e aos meus queridos, maravilhosos, falecidos Putxi e Kiko, porque vocês são mesmo o melhor amigo do Homem, O amor incondicional, o companheirismo… Não se pagam! Aos primeiros, o maior osso do mundo. Aos últimos, a oportunidade vos poder voltar a ver um dia. Ao meu pai, um obrigado pelo enorme esforço que tens feito agora nesta fase. Obrigado por quereres melhorar, por ti, por mim e pela minha irmã. A ti, dar-te-ia a hipótese de um dia poderes ver o estádio do teu clube de cima na visão da águia. À minha pirralha, minha irmã, porque te ver crescer é sempre difícil e um aperto para mim, porque é um orgulho ver-te tornar na mulher que és hoje, leal, linda, assertiva e empática com os teus. A ti dar-te-ia o melhor computador *gamer* à face da terra.

Por fim, à minha mãe. A ti dava-te, se pudesse, o meu coração. Tu criaste-o, tu é que o construíste. Nunca teria sido alguém sem ti. Ensinaste-me tudo, entre os quais os meus valores, crenças, alegrias, e acima de tudo o que é ter coragem. Nunca vi alguém tão corajosa na minha vida. Nunca alguém me despertou tantos sentimentos de amor, admiração, preocupação, carinho e de sonhar alto. Obrigado por tudo aquilo que me deste, seja imaterial ou material. O meu coração é teu e que sempre bata em uníssono por nós dois.

ABSTRACT

Type II NADH Oxidoreductase (NDH-2) from *Staphylococcus aureus* was established as a therapeutic target against the virulency of this bacterium and an alternative to treat Complex I-derived diseases. To accurately model interactions of NDH-2 with its substrates such as menaquinones and NADH, Coarse-Grain (CG) simulations were employed.

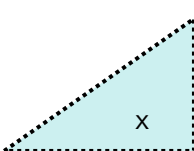
We used the Martini 3 CG force field, for which relevant molecules were parameterised. Martini follows a building-block approach; our parameterisation thus yielded a set of 35 molecules including other quinones (parameterised following a bottom-up approach) and nucleotides (which followed a top-down approach). Model validation compared atomistic and CG Connolly surfaces, their solvent accessible surface area (SASA) differences and the calculation of octanol-water partition coefficients (logPs). Overall, SASA differences were below the generally accepted limit of 5%. LogP analysis showed phosphorylated compounds, and alcohols to a lesser extent, are likely too hydrophilic in Martini 3. We employed mitigation strategies.

Aqueous simulations showed the expected *in vivo* interactions and selectivity of NDH-2 towards menaquinones. These quinones were also seen to prefer more extensive binding sites of all quinones. Furthermore, we established that D302 would bind NADH by its adenine and enable its bending to interact with FAD, NDH-2 cofactor.

For larger simulations, a model of the *S. aureus* membrane was built. Its fluidity is kept by the existence of methyl branches in its constituting lipids. We mimicked this fluidising effect with a degree of tail bending and were able to recover a gel-to-fluid transition of 293 K. We also observed that NDH-2 was able to pull menaquinones out of the membrane more than their usual fluctuations, highlighting how it is able to catalyse electron transfer monotopically.

We hope this work contributes to future research to unveil new potential targets to inhibit NDH-2 and as well for the Martini community.

Keywords: Coarse-Grain, Martini 3, Nucleotides, Quinones, Type II NADH Dehydrogenase, *Staphylococcus aureus*


x

RESUMO

O NADH Oxidorredutase do tipo II (NDH-2) de *Staphylococcus aureus* foi estabelecido como um alvo terapêutico contra a virulência desta bactéria e uma alternativa para tratar doenças relacionadas com o Complexo I. Para modelar as interações do NDH-2 com os seus substratos como as menaquinonas e o NADH, implementaram-se simulações Coarse-Grain (CG).

Usámos o campo-de-forças CG Martini 3, para o qual moléculas relevantes foram parametrizadas. Este baseia-se numa metodologia *building-block* resultando em 35 moléculas incluindo outras quinonas (seguindo uma abordagem *bottom-up*) e nucleótidos (*top-down*). A validação destes foi conseguida pela comparação das superfícies Connolly atomística e CG, as diferenças de área de superfície acessível ao solvente (SASA) e cálculo dos coeficientes de partição octanol-água (logPs). Em geral, a maioria das diferenças de SASA estão abaixo dos 5% aceitáveis. A análise de logPs mostrou que os compostos fosforilados e os álcoois, em menor extensão, estão demasiado hidrofílicos no Martini 3. Estratégias de mitigação foram implementadas.

As simulações aquosas demonstraram as interações expectáveis *in vivo* e uma seletividade do NDH-2 para as menaquinonas. Estas foram vistas preferindo eventos de ligação mais morosos entre todas as quinonas. Adicionalmente, o D302 foi visto a fixar o NADH pela sua adenina, permitindo o seu dobramento e interação com o FAD, cofator do NDH-2.

Em simulações de larga escala um modelo da membrana de *S. aureus* foi construído. A fluidez desta é mantida devido à existência de grupos metilo nos lípidos. Mimetizámos o seu efeito fluidificante usando um ângulo na cauda. Estes transitaram do estado de gel para fluído a 293 K. Também observámos que o NDH-2 conseguia puxar as menaquinonas da membrana mais do que o usual, evidenciando a sua capacidade de transferência eletrónica monotopicamente.

Esperemos que este trabalho contribua para futuras investigações na descoberta de alvos de inibição do NDH-2 e na comunidade Martini.

Termos-chave: Coarse-Grain, Martini 3, Nucleótidos, Quinonas, NADH Desidrogenase do tipo II, *Staphylococcus aureus*

INDEX OF CONTENTS

INDEX OF CONTENTS

INDEX OF FIGURES

INDEX OF FIGURES

INDEX OF FIGURES

INDEX OF TABLES

LIST OF ABBREVIATIONS

a – Hydrogen bonding acceptor label
AA – Atomistic or All-Atom
ADN - Adenine
ADOS - Adenosine
ADP – Adenosine Phosphate
AMP – Adenosine Monophosphate
AMR – Antimicrobial Resistance
ATP – Adenosine Triphosphate
BCFAs - Branched Fatty Acids
C – Apolar bead
CID – Compound Identifier
CG – Coarse-Grain or Coarse-Grained
CL – Cardiolipin
CM - Continuum Mechanics
COW - Centre-Of-Weight
Cpl – Complex I
Cyd - Cytochrome *bd* Menaquinol: Oxygen Oxidoreductase or Cytochrome *bd* Oxidase
D – Divalent ion bead
d – Hydrogen bonding donor label
DAG – Diacylglycerol
DHNA - 2,4-dihydroxynaphtoic acid
DMQ – Demethylmenaquinone
DMNFH - 2,3-Dimethyl-1,4-naphthalenediol
DMNFQ - 2,3-Dimethyl-1,4-naphthoquinone
DNA - Deoxyribonucleic Acid
e – electron-richer label
EC - Enzyme Commission (number)
ETs - Exfoliative Toxins
EU – European Union
FAD/FADH$_2$ – Flavin Adenine Dinucleotide (oxidised/reduced forms)
SI - International System of Units
FeS - Iron-Sulfur
FMN – Flavin Mononucleotide
GBZ - Methyl-Geranylbenzoquinone
G8YLG - Geran-8-yl Geran
HAQO - 4-hydroxyquinoline 1-oxide
HAQOH - 1,4-dihydroxyquinolin-1-ium
HLS - Haemolysins
HQNO - 2-heptyl-4-quinolinol 1-oxide
HQNOH - 2-heptyl-1-hydroxyquinolin-1-ium-4-ol
IUPAC - International Union of Pure and Applied Chemistry
LJ – Lennard-Jones
LPG – Lysyl Phosphatidylglycerol
MBAR - Multistate Bennett Acceptance Ratio
MD – Molecular Dynamics
MM – Molecular Mechanics
MND – Menadione
MNDOL - Menadiol
MNQ – Menaquinone
MNQOL - Menaquinol

MRSA - Methicillin-Resistant *Staphylococcus aureus*
N – Intermediate/Non-polar bead
NADH/NAD$^+$ – Nicotinamide Adenine Dinucleotide (reduced/oxidised)
NADPH/NADP$^+$ - Nicotinamide Adenine Dinucleotide Phosphate (reduced/oxidised)
NarGH – Menaquinol : Nitrate Oxidoreductase
NDH-1 – Type I NADH Dehydrogenase or Type I NADH : Quinone Oxidoreductase
NDH-2 – Type II NADH Dehydrogenase or Type II NADH :Menaquinone Oxidoreductase
OCT - Octanol
P – Polar bead
PBC – Periodic Boundary Conditions
PDB – Protein Data Bank
PES - Potential Energy Surface
PG – Phosphatidylglycerol
PQ – Plastoquinone
PQOL - Plastoquinol
Q – Monovalent ion bead
QM – Quantum Mechanics
Qox - *aa*$_3$-type Menaquinol: Oxygen Oxidoreductase
R – Regular bead
RBFL – Riboflavin
RBOS - Ribose
ROS – Reactive Oxygen Species
ReB – Restricted Bending Potential
RNA - Ribonucleic Acid
RQ – Rhodoquinones
S – Small bead
S. aureus - *Staphylococcus aureus*
SAS - Solvent Accessible Surface
SASA - Solvent Accessible Surface Area
SES - Solvent Excluded Surface
SDH – Succinate Dehydrogenase or Succinate : Quinone Oxidoreductase
SMILES - Simplified Molecular Input Line Entry System
T – Tiny bead
tDBDF - Two-Dinucleotide Binding Domains Flavoproteins
THI – Thiamine
TMP – Thiamine Monophosphate
TPP - Thiamine Pyrophosphate
TSST - Toxic Shock Syndrome Toxin
U – Dummy bead
UA – United-Atom
UBQH - Ubiquinol
UBQN – Ubiquinone
UBQ0 – Ubiquinone-0
UBQ0H – Ubiquinol-0
v – electron-poor label
V-Rescale – Velocity-Rescale
VRSA - Vancomycin-Resistant *Staphylococcus aureus*
VMD – Visual Molecular Dynamics
W – Water bead
WHO - World Health Organization
X – Halo-compound bead
XQ - *o*-Xyloquinone
XQH - *o*-Xylene-3,6-diol

LIST OF SYMBOLS

Vectors will be displayed in **bold**.

T – Absolute Temperature
a – Acceleration vector
ϑ, ϕ and φ – Angles (angle bond, torsion and improper dihedral)
Q – Charge
[X] – Concentration of the X specie
τ - Coupling parameter
r_{cut} - Cut-off radius distance
d, r or ℓ - Distance
K – Force constant
F – Force vector
R - Ideal gas constant
σ_{ij} – LJ collision diameter
ε_{ij} – LJ depth of the well
m – Mass of the particle
ϕ_d – Phase factor
V – Potential energy
P - Pressure
ε_ℓ - Relative dielectric constant or Relative permittivity
r - Space positions vector
n – Step or multiplicity
t – Time
ε_0 - Vacuum permittivity
Δt – Variation in time
ΔG – Variation in the Gibbs free energy
v – Velocity vector

1 BIOLOGICAL STATE-OF-THE-ART

1.1 Antimicrobial resistance – a new pandemic around the corner

Antimicrobial resistance (AMR) is by some described as the next pandemic[1]. It can be defined as the ability of a microorganism to, over time and through genetic changes, no longer respond to medicines - including antiparasitics, antifungals, antivirals and antibiotics – which might retard or prevent treatment of an infection caused by the same microorganism - possibly leading to countless consequences[2], *e.g.*, the patient's death. According to the World Health Organization (WHO), AMR is one of the top 10 global public health threats faced by humanity[3].

Bacteria have been efficient organisms gaining resistance to antibiotics, and this resistance can be associated with dissimilar causes, such as genetic material transfer, self-medication, availability of new antibiotics and selective pressure[4]. In 2019, it was estimated that almost 5 million people worldwide died due to a direct or indirect association with bacterial AMR, most of them in Africa[5], as shown in **Figure 1.1-A**:

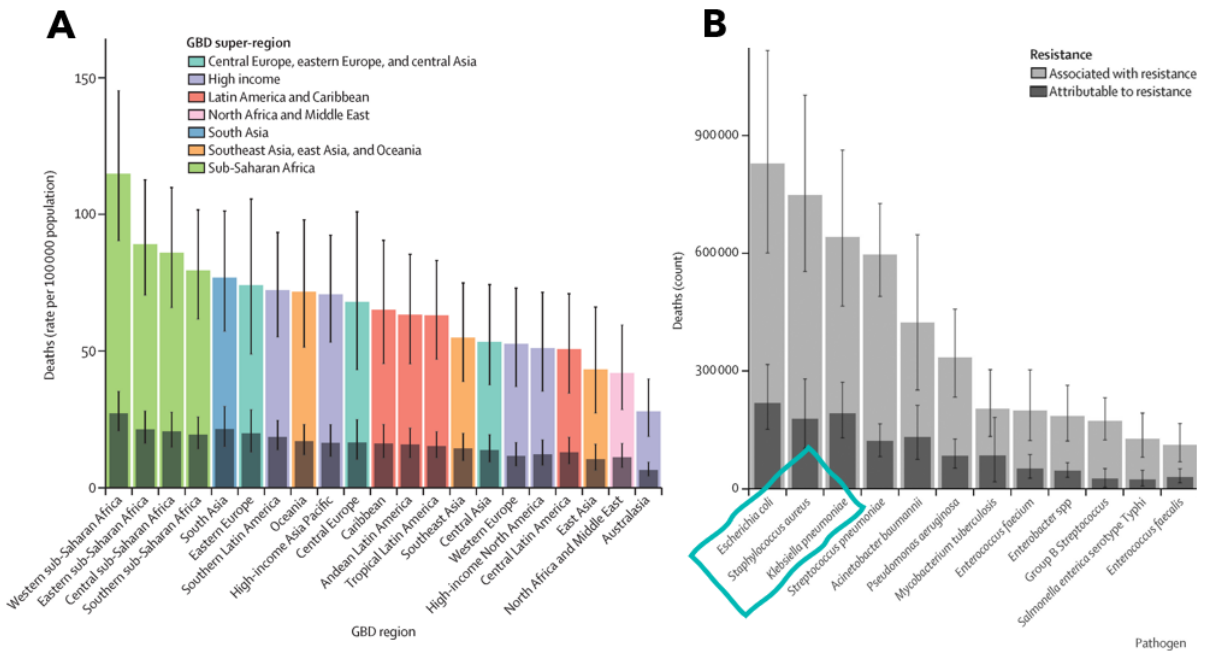


Figure 1.1 – (A) Graphic representation of the death rate per one hundred thousand people associated with AMR from different GBD (Global Burden of Diseases) regions around the world, 2019. (B) Graphic representation of the global death counts attributable to AMR bacteria by pathogen, 2019. The three bacteria that cause the most deaths are inserted in a blue rectangle. Error bars show 95% uncertainty intervals. Retrieved from Antimicrobial Resistance Collaborators study[5].

When the WHO released a report on AMR bacteria, it focused on some rather contagious and resistant bacteria, like *Klebsiella pneumoniae*, *Escherichia coli* and *Staphylococcus aureus* (*S. aureus*)[6]. The acquired resistance of these bacteria are the three main causes of death by bacterial AMR found in Antimicrobial Resistance Collaborators study[5], as shown in **Figure 1.1-B**.

Of these three bacteria, *S. aureus*, namely methicillin- and vancomycin-resistant strains (MRSA and VRSA), have the highest level of priority ("High – priority 2") in research to find an antibiotic, according to the WHO[7]. Therefore, studies that involve the development of drugs targeting this bacterium or the characterisation of their binding and function are of the utmost importance.

S. aureus infection, and particularly MRSA, is a pandemic and highly widespread in hospitals[8]. Portugal is currently the 5th European Union (EU) country with the most MRSA isolates, but was the first from 2010 to 2012, where the MRSA phenotype was almost 60% of isolates[9]. Despite the stabilisation trend in the EU and the declining tendency in our country, some European countries such as Norway, Croatia, Latvia, Cyprus and other Nordic countries are facing a growth in case numbers[10,11]. Besides, a recent study showed that in the United States, there has been a new increase in MRSA cases, in a place where a significant dwindling trend was occurring since 2005[12], because of the COVID-19 pandemic[13]. Fortunately, there are still many antibiotics available that can be used against MRSA infections and other *S. aureus*-resistant strains, such as VRSA[14].

1.2 *S. aureus*

1.2.1 Microbiological and virulence aspects

S. aureus is a non-motile, non-spore-forming, facultative anaerobe, Gram-positive bacterium[15,16] and a commensal microorganism capable of asymptomatically infecting about 30 % of the population, colonising mucosae and skin, but it can also be an opportunistic and virulent pathogen[17–19]. These spherical organisms are the cause of multiple infections (**Figure 1.2**), which are leveraged by various virulence factors:

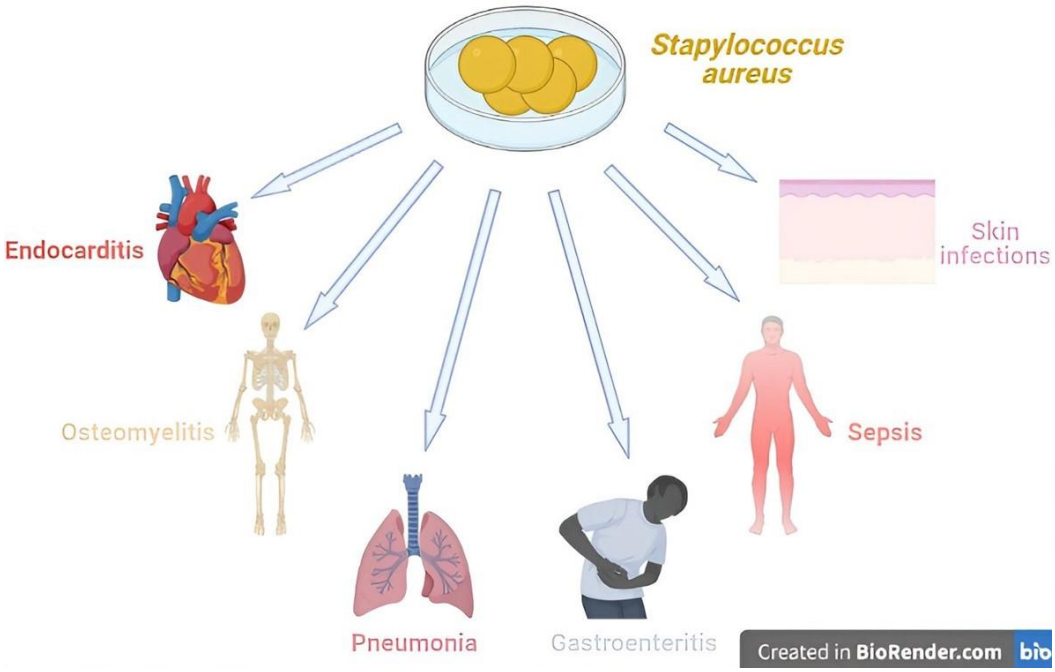


Figure 1.2 – Illustration that shows some of the multiple infections triggered by *S. aureus*, adapted from Merck Sharp and Dohme manual[20] and Kozajda *et al.*[21]: endocarditis - inflammation of the heart inner membrane, endocardium[22]; osteomyelitis - infection in a bone[23]; pneumonia – acute inflammation of pulmonary tissue[24]; gastroenteritis – inflammation of the digestive system[25]; sepsis – inflammation throughout the whole body[26]; skin infections. Created using BioRender[27].

The most prevalent are:

- Biofilms: an assemblage of organised microbial cells which contributes to the diffusion and establishment of these microorganisms[28,29] (**Figure 1.3**).
- Polysaccharide capsules: impair the opsonisation from the complement system and antibodies and the phagocytosis by phagocytes[30]. Some strains have also acquired surface-associated proteins to their cell wall structure, such as the Staphylococcal Protein-A, which attaches to the circulating Immunoglobulins G, maximising the protection against these two processes[30,31].
- Exfoliative toxins (ETs): serine proteases, which catalyse the detachment of desmosomal proteins (cadherins). This cleavage leads to skin exfoliation, dryness and blister formation, characteristics of the Staphylococcal Scalded Skin Syndrome[29–31].
- Pore-forming toxins: Haemolysins (Hls) -α (or α-toxin) and -β (or β-toxin) and leukotoxins (**Figure 1.3**), which catalyse the erythrocytes and leucocytes' lysis, respectively[29–31].
- Superantigens: antigens that overstimulate the immune system[32] (T cells, in specific) as shown in **Figure 1.3**, in particular, the toxic shock syndrome toxin (TSST), which causes the toxic shock syndrome, whose main downfalls consist of systemic and gastrointestinal disorders[29–31].
- Coagulase: catalyses the conversion of fibrinogen (factor I, soluble glycoprotein complex) to insoluble fibrin causing coagulation (**Figure 1.3**)[30,33].

There are exhaustive lists of virulence factors in literature[34], nonetheless the most significant are mentioned above, and some consequences are hereby summed up in **Figure 1.3**:

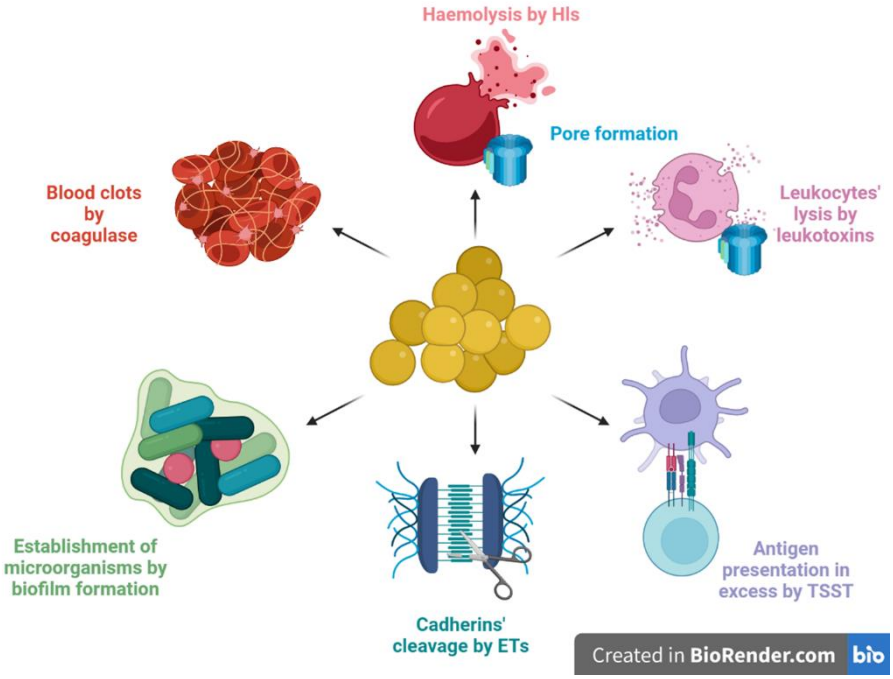


Figure 1.3 – Illustration that shows some consequences of *S. aureus* virulence factors. Created using Biorender[27]. Abbreviation meanings are in the text above.

1.3 NDH-2: a special enzyme in the *S. aureus* membrane

1.3.1 *S. aureus* membrane – characteristics and constitution

S. aureus membrane plays a key role in defining bacterial virulence and methyl branches in maintaining its fluidity.

We know that unsaturated fatty acids contribute to a greater fluidity of the phospholipidic membrane. They contain one or more *cis* double bond, thus occupying a larger space. When the temperature decreases, the membrane lipids press in on each other, but their kinks forced them to keep some space between each other. Thus, the membrane does not become rigid. Conversely, with saturated fatty acids, the membrane pressing would result in an increase in rigidity, possibly rupturing[35].

Most of Gram-positive bacteria contain many more saturated fatty acids (65 to 80 %) than unsaturated. Besides, *S. aureus* cannot produce most of the unsaturated fatty acids[36–39]. This way, the presence of branched fatty acids (BCFAs) with methyl branches, is key to membrane fluidity and, consequently to respond to nutrient conditions or environmental stimuli[36–39]. This is due to the fact that these BCFAs also have kinks mimicking the double bounds present in unsaturated fatty acids. These kinks – methyl branches – occupy enough space to keep the membrane fluidity when the temperature decreases[35].

BCFAs can be divided into two groups, according to the position of those methyl branches: *iso* or *anteiso,* as shown (**Figure 1.4**):

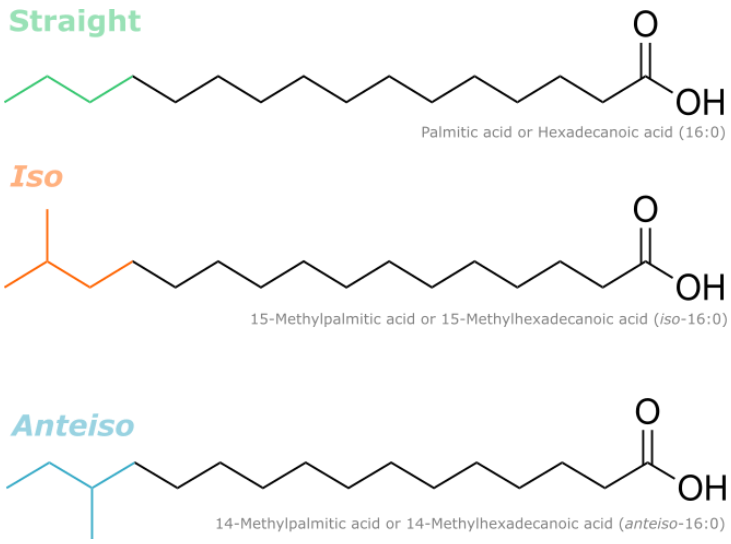


Figure 1.4 – Illustration that shows the structural differences between straight fatty acids chains (PubChem CID/Compound Identifier: 985), *iso*-BCFAs (PubChem CID: 164860) and *anteiso*-BCFAs (PubChem CID: 22207) of the palmitic acid (16:0). All structures were drawn using ChemDraw[40] and brought together using Inkscape[41]. Adapted from Taormina *et al.*[42].

The complexity of the *S. aureus* membrane increases because these fatty acids are combined with other lipids such as Lysyl Phosphatidylglycerol (LPG, 7.95-18.4%), Phosphatidylglycerol (PG, 22.5-54.5%), Diacylglycerol (DAG, 10.1-23.6%) and Cardiolipin (CL, 0-20.4%). The exact contribution of each lipid species to the membrane depends on the study and experimental setup. Furthermore, lipids in the *S. aureus* membrane have mainly 15:15 or 17:15 number of carbons in their isoprenoid chains, in a 1/3 ratio between the latter two[39,43–46].

Respiration occurs in the *S. aureus* membrane. Respiratory proteins perform the chemiosmotic process, established in 1961 by Peter Mitchell[47]. He postulated that the transmembrane differences in charge concentration are what enhances energy synthesis (ATP). In other words, electron transportation, occasionally allied to the proton pumping to the intermembrane space (P side) against their concentration gradient, creates an electrochemical potential. Then, the electrochemical potential is dissipated by ATP Synthase, releasing ATP in the matrix (N side)[47–53]. This energy, created in the respiratory membrane, is essential to the bacterium survival.

S. aureus possesses a plethora of respiratory proteins:

(i) NDH-2 (Type II NAD(P)H Dehydrogenase), which catalyses the same catalytic reaction as Complex I (EC 7.1.1.2, NAD(P)H: Quinone Oxidoreductase, alternatively Type I NADH Dehydrogenase, NDH-1 or CpI), but does not pump protons[54–56].

(ii) Other alternative entrance ways of electrons into the membrane (menaquinone reductases) such as Complex II or SDH (Succinate Dehydrogenase)[57,58].

(iii) Menaquinol oxidases such as Cyd (Cytochrome *bd* Menaquinol : O_2 Oxidoreductase or Cytochrome *bd* Oxidase), Qox (aa_3-type Menaquinol : Oxygen Oxidoreductase) and NarGH (Menaquinol : Nitrate Oxidoreductase), which catalyses the reducing of nitrate to nitrite, so that *S. aureus* can respirate anaerobically, if O_2 is not present[57,58].

As exhaustively enumerated, there is a great diversity of proteins in the *S. aureus* membrane[59–61]:

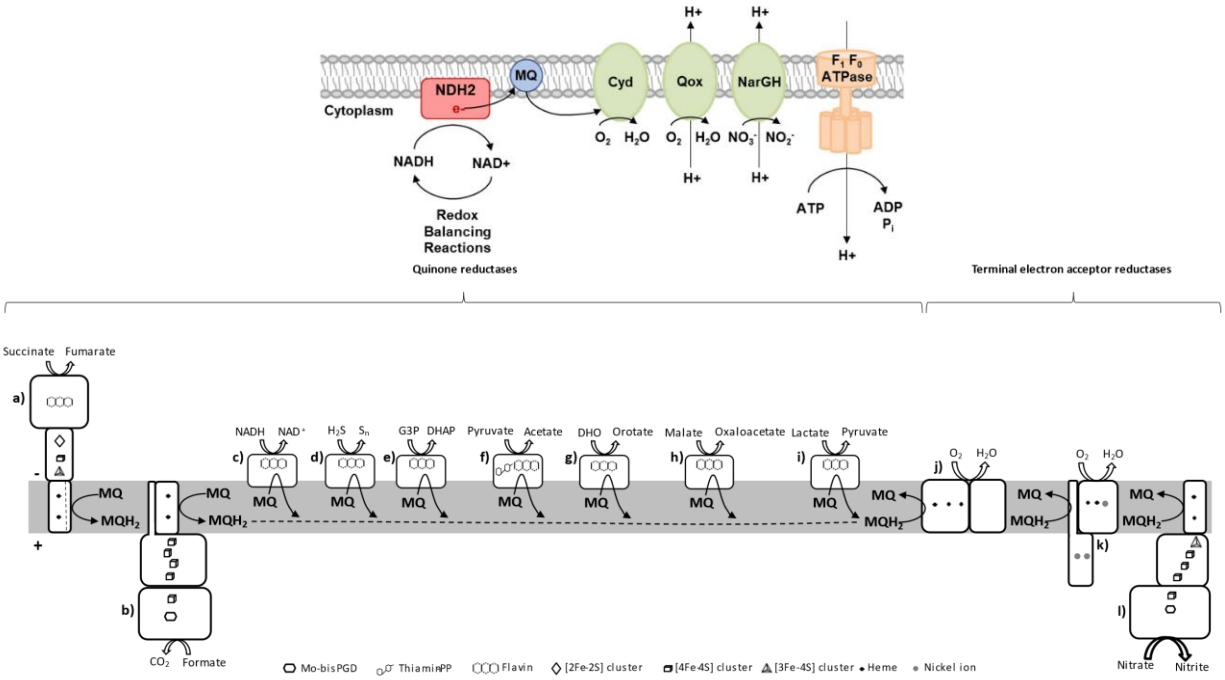


Figure 1.5 – Illustrations of the *S. aureus* respiratory complexes. At the top, we have a simplified version which does not show SDH, only showing the electron flux from NDH-2, retrieved from Potter's thesis[62]. At the bottom, we have a much more complex extended version, of *S. aureus* respiratory proteins, retrieved from Sousa's thesis[63]: a)-i) are Menaquinone reductases, j)-l) the final oxidases. NDH-2 is c). G3P: Glyceraldehyde 3-phosphate; DHAP: Dihydroxyacetone phosphate; DHO: Dihydroorotate. The rest if the abbreviation definitions are given in the text.

A particular aspect of respiration, electron transport, only occurs because respiratory membranes possess electron donors (such as NADH or NADPH) and electron carriers (such as quinones). Without them, organisms that rely on respiration to obtain energy couldn't perform it.

Nucleotides (such as NADH or NADPH) are ubiquitous and fundamental elements of life. They are i) the building blocks of DNA and RNA, ii) transporters of energy as high-potential electrons (NADH/NADPH) or high-energy phosphates (ADP/ATP), iii) cofactors of many enzymes (as, for instance, FAD), and iv) important signalling metabolites (for example, cyclic AMP)[64–70].

Additionally, there is another group of molecules that complements the energy-related roles of nucleotides. They are called quinones. They are important electron transporters in the respiratory chain[71,72]. Contrary to nucleotides, quinones are lipophilic, which enables the transfer of electrons and protons across the hydrophobic region of bacterial, mitochondrial and chloroplast membranes[73,74]. Some of these quinones, such as menaquinones, plastoquinones and ubiquinones, have further important functions in bacterial metabolism[75], making them potential therapeutic targets[76]. In the case of *S. aureus* membrane, the main quinone is menaquinone-8 or MNQ-8[77,78].

According to the IUPAC, quinones are "compounds having a fully conjugated cyclic dione structure, such as that of benzoquinones, derived from aromatic compounds by conversion of an even number of –CH= groups into –C(=O)– groups with any necessary rearrangement of double bonds (polycyclic and heterocyclic analogues are included)"[79].

In the respiratory membrane, these quinones are reduced to quinols (in *S. aureus*, menaquinone to menaquinol, MNQOL), involving two protons and two electrons, as shown in the following reaction scheme (**Figure 1.6**):

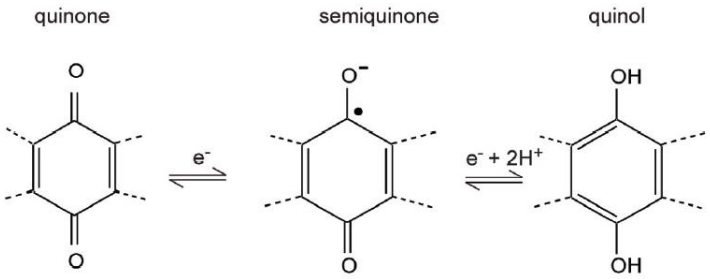


Figure 1.6 – Illustration that shows the general reaction scheme of a quinone being reduced to a quinol in a two-step mechanism. Adapted from the Bioenergetic book[48].

Quinones can be organised into two major groups (**Figure 1.7**): benzoquinones and naphthoquinones. Ubiquinones (UBQN), Rhodoquinones (RQ) and Plastoquinones (PQ) belong to the former group, whereas Demethylmenaquinones (DMQ), Menaquinones (MNQ), 2,4-dihydroxynaphtoic acid (DHNA) and Menadione (MND) belong to the latter. UBQNs are dispersed among prokaryotes and eukaryotes (including in the human mitochondrial respiratory membrane), PQs can be found only in cyanobacteria and plants, RQs in very few bacterial, unicellular eukaryotic and animal species and DMQs and MNQs are present mainly in respiratory membranes of Gram-positive bacteria and *archaea*; MNQ-2 can also be found in mammals (vitamin K_2), but not as a part of their respiratory membranes[75,80].

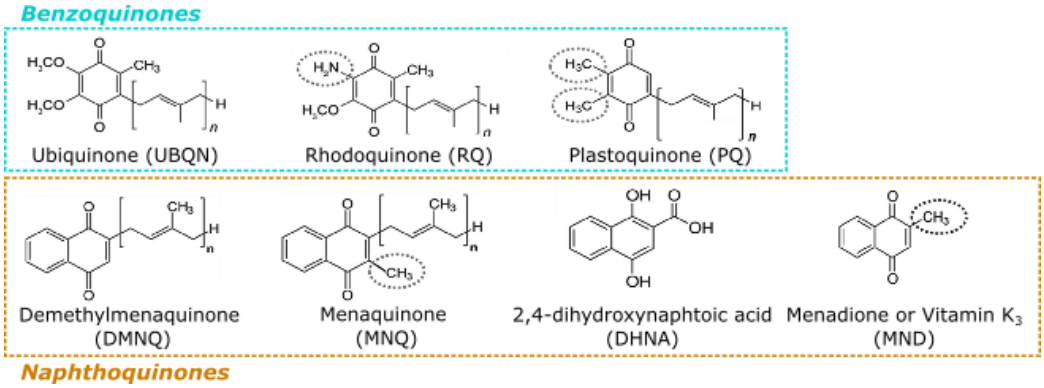


Figure 1.7 – Illustration of the different chemical structures of quinones divided into their two groups: benzoquinones and naphthoquinones. Some chemical changes between the different quinone structures are emphasised with circles. Adapted from Franza & Gaudu[75] using Inkscape[41]. Abbreviation meanings are in the text above.

Nowadays, these quinones are considered more than hydrophobic electron shuttles since they are essential for many other functions such as activation of proteins involved in blood clotting and bone metabolism (MNQ-2 in humans), potentiation of haem toxicity, enhancement of biofilm production and control of regulatory systems (MNQ-8) in *S. aureus,* whilst increasing its virulence[75,76,81–84]. Additionally, menaquinone reduces the ability of colonisation by *S. aureus* and, consequently, its pathogenicity[85].

NDH-2 is a protein that reduces MNQ-8 to menaquinol and uses NADH or NADPH as electron donors. This protein is not found in humans and does not establish any electrochemical potential in the *S. aureus* membrane; thus, this protein can be seen as a therapeutic target.

1.3.2 NDH-2: biochemical characteristics

Type II NAD(P)H Dehydrogenase, Type II NAD(P)H : Menaquinone Oxidoreductase from *S. aureus* or simply NDH-2 (EC 1.6.5.9[86,87], Protein Data Bank or PDB entry: 5NA1[88]) is a non-pumping, non-electrogenic homodimer with 45 kDa per subunit. NDH-2 is a monotopic protein which catalyses the oxidation of NADH/NADPH with the concomitant reduction of menaquinone to menaquinol (MNQOL), which can be used as a substrate by the terminal oxidases. It is non-electrogenic because monotopic proteins never do any work (in a thermodynamic sense), and thus, do not generate any electrochemical potential (charges do not cross the membrane against their chemical or electric gradient through NDH-2). Furthermore, this redox chemical reaction does not lead to the formation of radical species[59–61].

NDH-2 performs the same chemical reaction as CpI but does not pump protons. CpI is not present in the *S. aureus* membrane; therefore NDH-2 has a preponderant role in NADH metabolism, especially in the [NADH]/[NAD$^+$] ratio, which is crucial for the maintenance of some cell pathways by avoiding [NADH] excess and regenerating NAD$^+$ (**Figure 1.8**-left)[59–61].

NDH-2 does not exist in mammals but can be found in the three domains of life. It has a non-covalently bound Flavin Adenine Dinucleotide (FAD) as a prosthetic group[59–61]:

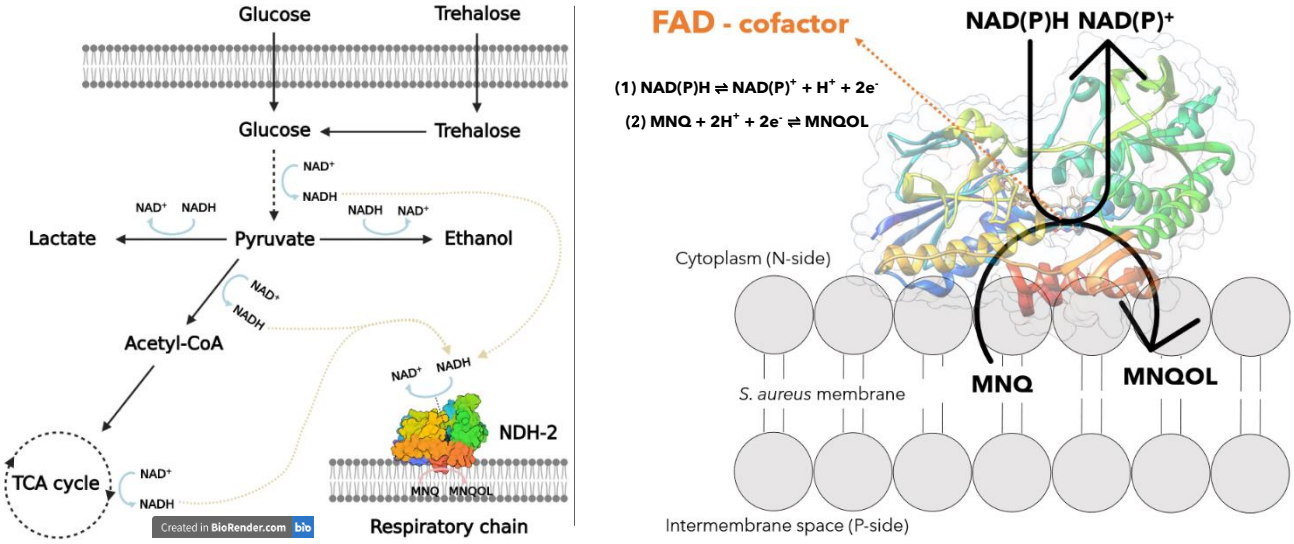


Figure 1.8 – On the left, an illustration that shows the NADH metabolism and how NDH-2 influences. Dark yellow dashed lines on the left indicate NADH feeding the respiratory chain by NDH-2. On the right, an illustration of NDH-2 and FAD structures, PDB: 5NA1[88], from *S. aureus* and the chemical reactions that occur in the binding pocket, is shown. The picture on the left was made using Biorender[27] and adapted from Sousa's thesis[63] whereas the one on the right was created using Chimera software[89].

NDH-2 was first discovered in 1967 by Bragg and Hou[90], who reported the presence of two distinct NADH Dehydrogenase activities in *Escherichia coli*. Also, this rotenone(Complex I inhibitor)-insensitive NADH Dehydrogenase from *S. aureus* is only present in the cytoplasmatic leaflet, as in all prokaryotes, but can be found in both leaflet sides in yeasts. In 2015, *S. aureus* NDH-2 was first purified and crystallised by Manuela Pereira's group in collaboration with Margarida Archer's group[60,91] at ITQB NOVA. In the same year, pre-steady-state kinetic studies showed the establishment of a ternary

complex mechanism. NADH donates its electrons and one proton to FAD (the other proton comes from the medium) in the so-called *re*-side, forming a charge-transfer complex (delocalised electron shared between NADH and FAD), dissociated by the menaquinone (NAD^+ is only released in the presence of the menaquinone). Afterwards, menaquinone is reduced to a menaquinol on the *si*-side, being the rate-limiting step. It was observed that NADH and menaquinone bind into two different sites[92,93].

This protein is composed of NAD(P)H and FAD binding domains, with identical structural Rossman folds, belonging to the two-dinucleotide binding domains flavoprotein (tDBDF) superfamily. In this situation, NADH/NADPH interact with one domain, whereas FAD interacts with the other (at the N-terminal). Regarding its secondary structure, this protein contains six parallel β-strands (an extended β-sheet formed by hydrogen bonds), the first three strands being connected by an α-helix. The tight loop between the first β-strand and the α-helix is characterised by having a specific consensus sequence: G-X-G-X-X-G, where X can be any amino acid residue. Furthermore, this protein possesses two isozymes: NDH-2A and NDH-2B[61,85,92,94,95].

In NDH-2, some amino acid residues have been found to have central importance in catalytic mechanisms. Glutamate (E)-172 has a vital role in proton transfer in the quinone pocket so that quinone reduction may take place. On the other hand, Aspartate (D)-302 seems to be in charge of making hydrogen bonds with FAD, being decisive for the enzyme catalytic activity by guaranteeing FAD protonation or $FADH_2$ deprotonation[93,96,97]. The lack of NDH-2 in mammal mitochondria opens doors for the use of drugs that could be administrated to humans that are infected with *S. aureus* without harming any of our cells[98]. This immediately singles out NDH-2 as a promising drug target to reduce *S. aureus* virulence. This latter reason for targeting this enzyme as a therapeutic approach will be clarified in the next section.

1.3.2.1 NDH-2 as a drug target for Complex I-derived diseases

In addition to its bactericidal interest, it is known that some diseases, namely neurodegenerative such as Parkinson's and Alzheimer, are related to deficiencies in mitochondrial Complex I. This leads to a decrease in ATP synthesis and a rise in the $[NADH]/[NAD^+]$ ratio, with catastrophic repercussions in people that face these dysfunctionalities[99,100].

Although NDH-2 does not pump protons, it catalyses the same chemical reaction as NDH-1. As a result, it was theorised the possibility of inserting NDH-2 into mammalian mitochondria to smooth the lack of Complex I consequences. In 1998, it was observed that in Complex-I deficient Chinese hamster cells, with the transfection of the yeast gene that codifies NDH-2, the NADH oxidase activity was restored. This right away indicated that NDH-2 could be used in gene therapy to treat diseases inherently associated with Complex I abnormalities[101]. In 1999 and 2000, the transfection to human embryonic kidney cells and nonproliferating human cells was successful[59,102,103].

In 2006, in mice administrated with a parkinsonian neurotoxin, NDH-2 showed protective effects against neurodegeneration because it increased the resistance to the neuronal injury-inducing neurotoxin[60,92,104]. More recently, in 2019, transfections of NDH-2 yeast to Complex-I deficient plants showed that mitochondrial activity was restored. Nevertheless, if NDH-1 was not inhibited, a competition between NDH-1 and NDH-2 arose that might be deleterious to potential patients. As a conclusion, this gene therapy may be only applicable if electron transfer from NADH to Complex I is impaired[105].

This enzyme also seems to be involved in a low production of reactive oxygen species (ROS), delaying ageing. Yet, there are still no potent nanomolar inhibitors reported, despite the potential of NDH-2 as a drug target[59,60,106,107]. Actually, there are more reasons supporting the selection of this protein as a therapeutic target to diminish *S. aureus* infections, but they are still speculation[85].

As has been done with the Complex I[108], we used Molecular Dynamics (MD) simulations to study the interactions occurring in the NDH-2 binding site.

1.4 Biological main knowledge gaps

It is known that NDH-2 can be a therapeutic target to decrease the number of *S. aureus* infections. However, there is sparse knowledge on how NDH-2 binding site can be druggable. This thesis aims at modelling and studying the dynamic binding of several substrates to NDH-2 binding site by using Coarse-Grain Molecular Dynamic simulations since it is known that the time- and size-scales of substrate-protein interactions are within reach of those simulations.

2 INTRODUCTION TO METHODOLOGY

Molecular modelling consists of the study of molecular structure and function using physical or theoretical models to explain or predict a seen behaviour. There are several levels of modelling in computational biochemistry, which are chosen according to the degrees of freedom we are studying and the timescale of the processes or properties we wish to see[109]. The most significant methods are Quantum Mechanics (QM), Molecular Mechanics (MM) and Continuum Mechanics (CM)[110].

The first is usually applied to studying chemical reactions and their mechanisms since QM deals with the electrons in the system. MM can be applied to proteins to check their structure. It ignores electronic movements and calculates the energy of a system, only using nuclear positions. CM methods are normally used to examine supramolecular properties such as viscosity and fluidity[109–111].

2.1 Molecular Mechanics

Reality is quantic. The motion of electrons in atoms is described by Schrödinger's equation, which predicts probabilistic distributions for the subatomic particles. Although Classic or Newtonian laws fail to describe a proper behaviour in the microscopic interactions field, electromagnetism, and optics realms[112], they use fewer computing resources[113–115], and equations that describe the Classic approach are used in MM.

We can use Newton's laws to describe nuclei movement and this movement can be separated from electrons' movement (Born-Oppenheimer approximation)[115] since they possess a large mass difference.

In MM, electrons are treated as single point/static charges collapsed in their nucleus atom centre and nuclei as point masses, whose motions are the ones to be considered[116–119]. This already causes a limitation, where MM cannot deal with phenomena such as charge transfer, photoisomerisation reactions or making or breaking of bonds[109,111,112,120–123].

Moreover, in an infinitely small time scale, the movements between two particles are considered to be independent of others, and ultimately, trajectories are obtained from a finite-difference approach, using discrete time steps since solving a set of second-order differential equations (of Newton's law) would be computationally expensive[120,123].

Furthermore, MM introduces a concept inherent to it: additivity. It indicates that molecular energy can be expressed as a sum of potentials derived from simple physical forces of the Classical framework (see Section 2.3).

2.2 MD simulations

We employed MD simulations since the size and time scales of biochemical processes involving nucleotides and quinones — typically within the tens of nanometres and up to the microsecond scale — are within reach of molecular dynamic (MD) simulation methods. Even though MD simulations are generally limited to non-reactive interactions (in that models do not allow for bond-breaking), they have been able to provide important molecular-level insight into the interactions of nucleotides and quinones with relevant biomolecules[108].

Molecular Dynamics is "the science of simulating the motions of a system of particles"[124] using Newton's equations of motion over time to get the motion of the particles in a system and, consequently, quantitative and qualitative information about the particle behaviour in the system[123].

From the integration of Newton's equations of motion, we can predict the forces, accelerations, velocities, and the new positions of the particles from their potentials. Then, integrations of Newton's equations occur, and atoms move. Afterwards, new potentials are obtained, and after that, new forces, velocities, and positions are calculated. The cycle repeats itself as many times as the user states (**Figure 2.1**).

First, there is a relation between force and potential that can be written as[123]:

$$\mathbf{F} = -\nabla(\mathbf{r})\mathrm{V}(\mathbf{r}) = -\left(\frac{\partial \mathrm{V}(\mathbf{r})}{\partial \mathrm{x}} + \frac{\partial \mathrm{V}(\mathbf{r})}{\partial \mathrm{y}} + \frac{\partial \mathrm{V}(\mathbf{r})}{\partial \mathrm{z}}\right) \quad (2.1)$$

Where **F** is the Force in Newtons (N) and the potential energy function, V(**r**), is in Joules/mole ($Jmol^{-1}$), $\mathbf{r} = \mathrm{x}\hat{\mathbf{i}} + \mathrm{y}\hat{\mathbf{j}} + \mathrm{z}\hat{\mathbf{k}}$ and $\nabla(\mathbf{r}) = \frac{\partial \mathrm{r}}{\partial \mathrm{x}} + \frac{\partial \mathrm{r}}{\partial \mathrm{y}} + \frac{\partial \mathrm{r}}{\partial \mathrm{z}}$.

Therefore, from the potential energy, we can predict which forces will be applied to each particle since the force is the symmetric of the first derivative of potential for each coordinate. To know the velocities and positions of each particle, we can use Newton's second law, which can be written as:

$$\mathbf{F} = \mathrm{m}\mathbf{a} \Leftrightarrow \mathbf{F} = \mathrm{m}\frac{\mathrm{d}\mathbf{v}}{\mathrm{dt}} \Leftrightarrow \mathbf{F} = \frac{\mathrm{d}\mathbf{p}}{\mathrm{dt}} \Leftrightarrow \mathbf{F} = \mathrm{m}\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{dt}^2} \quad (2.2)$$

Where force (**F**, N) and acceleration (**a**, m/s^2) are correlated. One should bear in mind that the acceleration is the first derivative of velocity (**v**, m/s) and the second derivative of positions (**r**, meters), both with respect to time (t, s). Also, the linear momentum (**p**, $kg.m.s^{-1}$) is the product of a particle mass (m, in kg) and its velocity.

Hence, from **Equation 2.1** and **Equation 2.2**, the velocities and positions for each particle can be calculated. This cycle is described as follows:

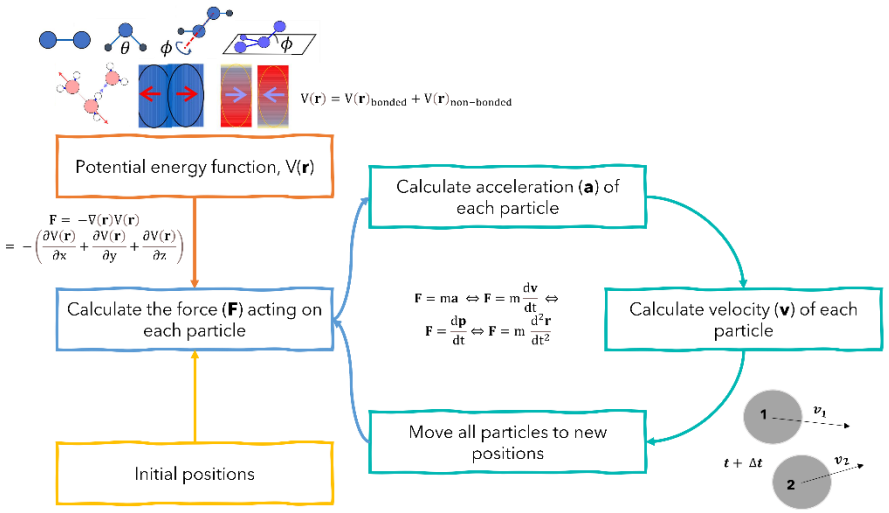
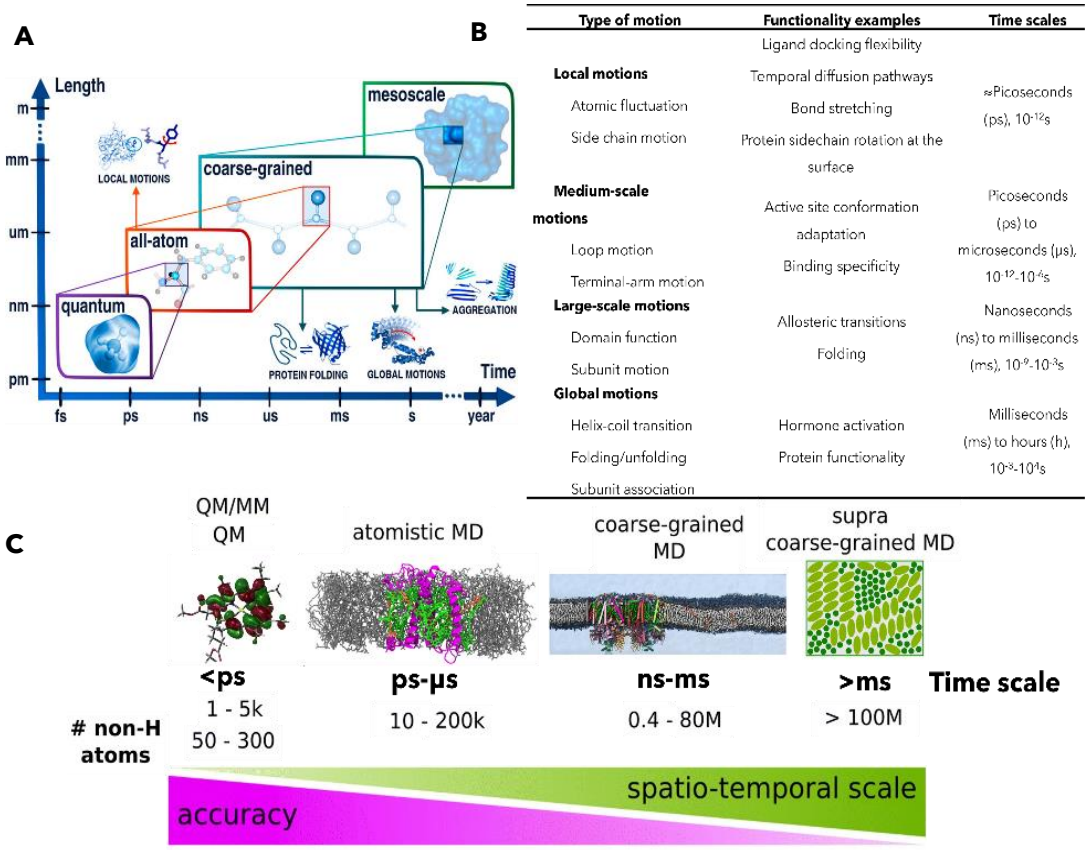


Figure 2.1 – Illustration of the Molecular Dynamics calculations cycle based on Equation 2.1, Equation 2.2, Equation 2.3 (addressed later), Verma *et al.*[123], Force fields and interactions website[125] and Bunker & Róg[126].

We can have different MD simulation resolutions that employ this latest cycle. These resolutions are chosen according to what we wish to simulate and which properties we want to study. For instance, atomistic detail is required in situations such as if chirality must be addressed or if we want to check the existence of hydrogen bonds. However, sometimes when atomistic detail is not needed and the computational velocity carries a more significant advantage, such as studying the interactions of a protein with a membrane, we may use a Coarse-Grain (CG) approach.

Here is an illustration that sums up several simulation resolutions and their time- and space-scales:



Type of motion	Functionality examples	Time scales
	Ligand docking flexibility	
Local motions	Temporal diffusion pathways	≈Picoseconds (ps), 10^{-12}s
Atomic fluctuation	Bond stretching	
Side chain motion	Protein sidechain rotation at the surface	
Medium-scale motions	Active site conformation adaptation	Picoseconds (ps) to microseconds (µs), 10^{-12}-10^{-6}s
Loop motion	Binding specificity	
Terminal-arm motion		
Large-scale motions	Allosteric transitions	Nanoseconds (ns) to milliseconds (ms), 10^{-9}-10^{-3}s
Domain function	Folding	
Subunit motion		
Global motions		Milliseconds (ms) to hours (h), 10^{-3}-10^{4}s
Helix-coil transition	Hormone activation	
Folding/unfolding	Protein functionality	
Subunit association		

Figure 2.2 – Illustration that shows different simulation resolutions, their space- and time- scales (A and C) and which types of motion are better characterised according to those scales (B). Please note that some types of motions (and examples) and some time scales may not be congruent between A, B and C since they are retrieved from different sources. Picture A was retrieved from Kmiecik *et al*.[127], Table B was elaborated according to the Dynamics of Proteins and Nucleic Acids' book[128] and Computational Biochemistry and Biophysics' book[120] and Picture C was adapted from Liguori *et al*.[129] and shows a chlorophyll, a single protein, protein supercomplexes and an assemblage of supercomplexes, in a spatio-temporal scale ascending order.

As seen in **Figure 2.2**-A/C, there are several different simulation resolutions. Quantum Mechanics (QM) and QM associated with Molecular Mechanics (MM) are the most accurate simulations and involve electrons' movement and the study of electrons' density. Atomistic (AA) simulations require an atomistic detail. These simulations are already MD simulations therefore, the electron movements are neglected, and the electron cloud is collapsed into the nucleus[127,129]. This detail is required in some cases, such as the studying of which hydrogen bonds are involved in an enzymatic reaction that occurs in a binding site. With all-atom (AA) simulations we have a fully atomic description of the system, but sometimes molecules can be accurately simulated without having to explicitly model hydrogen positions. In these circumstances, we may use a level of simplicity between AA and CG: united-atom (UA) simulations. In UA simulations, non-polar hydrogen atoms are not explicitly treated[130,131]. Simpler than UA is CG.

2.2.1 Coarse-Grain

Processes involving molecular diffusion within membranes or involving multiple molecular interaction partners — such as those involving multiple proteins — may have typical timescales in the millisecond range[132]. These systems may, therefore, be too costly to address using atomistic-resolution (AA) simulations, and coarse-grain (CG) MD methods — where some degree of system detail is sacrificed in exchange for faster simulation times — have become a common alternative[133–135].

In CG (**Figure 2.2-A/C**), a certain chemical group of some atoms is mapped into a single bead. Also, here electrons and atoms are collapsed in the centre of the bead[116–119]. This collapses many degrees of freedom, and they are used when no atomistic detail is necessary, like for the study of lipids in a membrane. This also allows the study of many more molecules at larger time scales and much faster simulations[136].

The following figure sums up the main structural differences between AA, UA and CG:

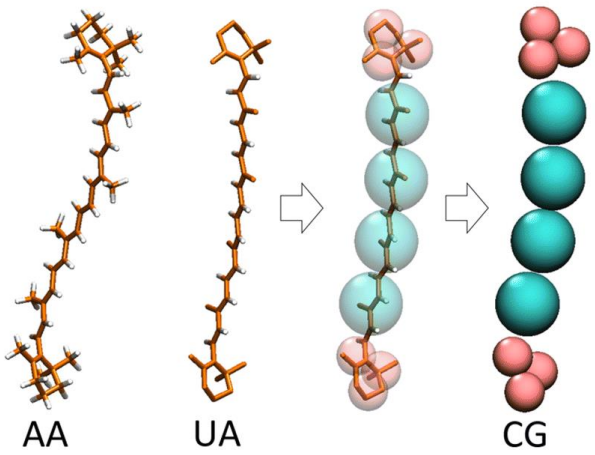


Figure 2.3 – Illustration of a β-carotene molecular structure represented at different levels of resolution: AA, UA and CG. Retrieved from Liguori *et al*.[129].

I will now focus on several MD aspects that are quite different in AA, UA and CG, highlighting the main features of the latter MD resolution.

2.2.2 Degrees of freedom (DOFs)

As **Figure 2.3** shows, in CG models, there is a reduction in the number of particles. There is a reduction of 1/3 to 1/5 in relation to UA and 1/10 in relation to AA[136]. For example, one water bead (W) corresponds to four atomistic water molecules, meaning that there is a one-fourth less computational cost only regarding the solvent.

2.2.3 Time step (Δt)

The loss of degrees of freedom leads to a smoother conformational landscape. This allows the increasing of the time step without leading the system to crash such as it is in Coarse-Grain. This allows achieving higher time scales to be simulated. In CG we have a bigger Δt, so simulations take less time to run than atomistic/UA simulations (with a smaller Δt)[123]. However, greater time steps diminish the accuracy, in spite of decreasing the computational cost.

2.2.4 Cut-off radius distance (r_{cut})

For the typically used Lennard-Jones and Coulomb potentials, two particles are constantly interacting. That means the potential approximates 0 when the distance between the two particles goes toward infinity, meaning that their contribution to the system energy becomes residual[123]. Since the calculation of potentials is computationally expensive, a cut-off radius (r_{cut}) is used, meaning that only

particles that are within the radius of another particle are considered to be interacting with each other. Thus, the energy contribution of that particle with others only counts when they are in a distance below or equal to r_{cut}[123].

A smaller r_{cut} reduces the simulation accuracy but improves the computing efficiency. Using the same MD simulations mentioned earlier, we can conclude that CG (minor r_{cut}) is faster than UA/AA simulations (higher r_{cut}) since the cut-off radius is lower thus, fewer far interactions are considered and lower the computational burden[123].

2.3 Potentials

As previously mentioned, in each iteration of the integration of Newton's Laws, forces, accelerations, velocities and positions are calculated from the potential energy function. The potential energy function, V(**r**), or force field is no more than a function of particles' positions. The force field mainly used in this thesis was the CG Martini 3[137], so some of the potentials that will be addressed are specific to this force field. The force field or the potential energy function can be split into two different components according to the type of interactions: bonded (internal) and nonbonded (external/non-covalent):

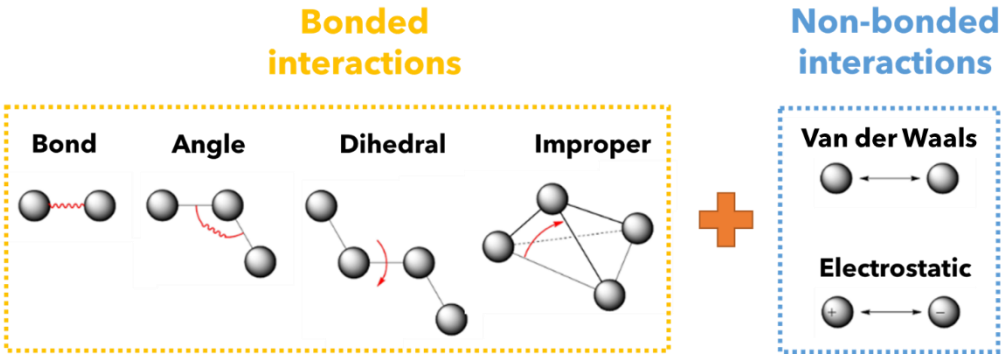


Figure 2.4 – Illustration that shows that the potential energy function can be divided according to the type of interactions between the two particles. In yellow, we have the bonded interactions (bonds, angles, proper dihedrals, and improper dihedrals), and in blue, the nonbonded interactions (Van der Waals and electrostatics). Adapted from Chang *et al.*[138].

It is possible to express **Figure 2.4** in a general equation[129]:

$$\mathrm{V}(\mathbf{r}) = \mathrm{V}(\mathbf{r})_{\mathrm{bonded}} + \mathrm{V}(\mathbf{r})_{\mathrm{nonbonded}} \tag{2.3}$$

Where the Potential Energy function, V(**r**), is in $Jmol^{-1}$ and where **r** is the unit vector (**ijk**).

From **Equation 2.3** and **Figure 2.4**, we can write the bonded potential terms as:

$$\mathrm{V(r)}_{\mathrm{bonded}} = \mathrm{V}_{\mathrm{bond}}(\mathrm{d}_{\mathrm{ij}}) + \mathrm{V}_{\mathrm{angle}}(\vartheta_{\mathrm{ijk}}) + \mathrm{V}_{\mathrm{proper\ dihedral}}(\phi_{\mathrm{ijkl}}) + \mathrm{V}_{\mathrm{improper\ dihedral}}(\varphi_{\mathrm{ijkl}}) \tag{2.4}$$

Where it is stated that the potential energy function in $Jmol^{-1}$ of bonded parameters depends on the potential energy of bonds, angles, proper dihedrals, and improper dihedrals, and where **r** is the unit vector (**ijk**).

The parameters of this equation get clearer with the following figure:

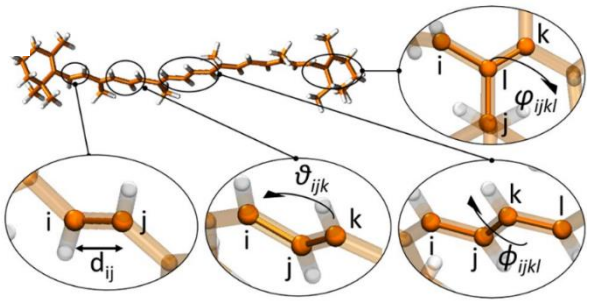


Figure 2.5 – Illustration that shows the addressed bonded potential terms in the example of a β-carotene molecule. Retrieved from Liguori *et al.*[129].

Lastly, we can unfold the nonbonded term of **Equation 2.3** into two terms:

$$\mathrm{V(r)_{non-bonded}} = \mathrm{V_{LJ}}(\ell) + \mathrm{V_{Coulomb}}(\ell) \tag{2.5}$$

This equation of the potential energy function (in Jmol^{-1}) of nonbonded parameters depends on two terms: Van der Waals (modelled by an Lennard-Jones potential) and electrostatics (modelled by Coulomb formula). **r** is the unit vector (**ijk**) and ℓ is the distance between two particles (SI unit: m, often used: nm).

Each term will now be addressed and explained:

Bonded potentials:

$$\mathrm{V_{bond}}(\mathrm{d_{ij}}) = \frac{1}{2}\mathrm{K}_b(\mathrm{d_{ij}} - \mathrm{d_b})^2 \tag{2.6}$$

The covalent stretching bond potential (in Jmol^{-1}, often used: kJmol^{-1}) depends on the distances (d units: m, often used: nm) and the force constant (K_b) in kJmol^{-1} units. In spite of a true bond-stretching potential not being truly harmonic, as suggested by **Equation 2.6**, this approximation is made since it is accurate near equilibrium (small deviations from the reference bond length $\mathrm{d_b}$). The $\mathrm{d_b}$ parameter is the "reference bond length" or "natural bond length", the distance adopted at the minimum of the potential energy surface[109,120,122]. On the other hand, K_b is the force constant of the bond, which tends to be a high value because large amounts of energy are required to stretch or compress a chemical bond, and $\mathrm{d_{ij}}$ is the actual bond length[127,129,138,139].

Then, we have a cosine harmonic potential (very similar to the former but used more often in CG):

$$\mathrm{V_{angle}}(\vartheta_{\mathrm{ijk}}) = \frac{1}{2}\mathrm{K}_a(\cos\vartheta_{\mathrm{ijk}} - \cos\vartheta_{\mathrm{a}})^2 \tag{2.7}$$

Equation 2.7 shows the terms of the cosine quadratic angle bending potential (in Jmol^{-1}, often used: kJmol^{-1}) where the angle units are °(degrees) and the force constant (K_a) is in kJmol^{-1} units. Here $\cos\vartheta_{\mathrm{ijk}} = \frac{\mathbf{r}_{\mathrm{ij}}\cdot\mathbf{r}_{\mathrm{kj}}}{\mathrm{r_{ij}r_{kj}}}$ where the numerator is the dot product, the denominator is the multiplication of the vectors' magnitudes, and $\mathbf{r}_{\mathrm{ij}}$ and $\mathbf{r}_{\mathrm{kj}}$ denote the bond directions[129,140].

This potential is derived from another angle potential, where $\mathrm{K}_c = \mathrm{K}_a\sin^2(\vartheta_{\mathrm{a}})$[129,140]:

$$V_{angle}(\vartheta_{ijk}) = \frac{1}{2} K_c (\vartheta_{ijk} - \vartheta_a)^2 \quad (2.8)$$

Equation 2.8 shows the bending angle potential (in Jmol^{-1}, often used: kJmol^{-1}), where the angle units are ˚(degrees) and the force constant (K_c) is in kJmol^{-1} units. Here the angle bending is also a harmonic potential, following Hooke's law[109,120,122,139]. The angle force constant, K_c, tends to be smaller than the bond force constant since it usually takes less energy for a bond angle (ϑ_{ijk}) to deviate from its reference value (ϑ_a). Likewise, more parameters can be included to increase the accuracy, but in common force fields, it is usually not necessary[127,129,138].

A dihedral is an angle between two planes, defined by four particles. If three of those four become collinear, the dihedral becomes undefined. In this situation, forces tend to infinity and the system explodes[141]. This way, a restricted bending potential (ReB), a potential used more often in CG, was created[141] to prevent one of the bond angles from reaching 180˚:

$$V_{ReB}(\vartheta_{ijk}) = \frac{1}{2} K_e \frac{(\cos\vartheta_{ijk} - \cos\vartheta_a)^2}{\sin^2(\vartheta_{ijk})} \quad (2.9)$$

In **Equation 2.9** the restricted bending potential (ReB) is defined (in Jmol^{-1}, often used: kJmol^{-1}) with degrees units (˚) and the force constant (K_e) is in kJmol^{-1} units. So, ReB prevents three consecutive particles from becoming contiguous since the wall of the ReB potential near 180º is very repulsive. K_e such as in the other potentials, controls the opening of the parabola[140–142]. This potential is not normally needed in AA simulations since collinearity is rare due to strong restraints imposed in bond angles by their force fields and is easily solved by re-starting the simulation. Conversely, in CG, molecules have lower number of degrees of freedom (lower number of particles to represent the same molecule), thus more flexibility and possibility of collinearity, therefore this can happen more frequently[142].

Notwithstanding, ReB needs to be used with caution (with low force constants) since the potential tends to be infinity (to extremely high forces). It is best utilised with other bond angle potentials (such as the cosine quadratic angle potential) if we want to place an angle near 180º. This is illustrated below:

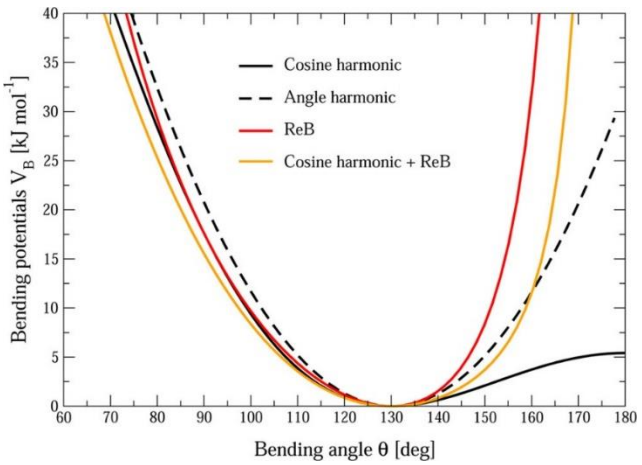


Figure 2.6 – Illustration that shows three different angle potentials. The cosine harmonic potential is represented in a solid black line and has a plateau when it reaches 180º. The harmonic angle potential is represented in a dashed black line and tends to infinity near 180º but softly. ReB is plotted in red and tends to infinity quickly near 180º. All of these potentials have a force constant of 85 kJmol^{-1}. In yellow is the cosine harmonic potential, which, allied to ReB, allows the potential to tend to infinite and to target angle values nearer 180º. In the latter, K_a = 50 kJmol^{-1} and K_e = 25 kJmol^{-1} were used. All of these potentials used 130º as the reference angle value. Adapted from Bulacu *et al.*[141].

As this figure shows, the use of the cosine harmonic allied to ReB allows the definition of certain angles' behaviour when they are near 180º. Likewise, there is also a restricted torsion (dihedral) potential with the same formula.

Returning to bonded potentials, the proper dihedral potential can be defined as:

$$V_{\text{proper dihedral}}(\phi_{ijkl}) = K_d(1 + \cos(n\phi_{ijkl} - \phi_d)) \quad (2.10)$$

Equation 2.10 defines the proper dihedral potential or torsional term (Jmol^{-1}, often used: kJmol^{-1}) in degrees (˚). Here the force constant (K_d) is in kJmol^{-1} units and n is the multiplicity. The proper dihedral is the angle (ϕ_{ijkl}) between two planes defined by four covalently bound particles (**Figure 2.5**). K_d, like other force constant equations, reproduces the stiffness of the angle between two planes in a certain molecule, being related to the force needed to rotate[109,120,122,139]. Here n is the periodicity of the potential (multiplicity), which gives the number of minimum points in the function in a 360˚ rotation about the central bond. Torsional potentials are expressed as a cosine function and ϕ_d is the phase factor from which it is possible to determine where the torsion angle passes through its minimum values. This factor is related to the bond geometry (linear, triangular, tetrahedral, etc)[127,129,138,139].

$$V_{\text{improper dihedral}}(\varphi_{ijkl}) = \frac{1}{2}K_{id}(\varphi_{ijkl} - \varphi_{id})^2 \quad (2.11)$$

Equation 2.11 defines, in degrees (˚), the improper dihedral potential (in Jmol^{-1}, often used kJmol^{-1}). The force constant is K_{id} and is in kJmol^{-1} units. Sometimes, an improper dihedral (φ_{ijkl}) potential is used to maintain the planarity of a molecule, such as in rigid planar structures [109,120,122]. Once again, **Equation 2.11** shows a harmonic potential that defines the out-of-plane motions (or improper torsions), whose force is described by K_{id} from a reference torsion (φ_{id})[127,129,138].

Nonbonded potentials:

Van der Waals interactions are of electrostatic nature and can be split into three types according to the molecular polarity: (i) (permanent) dipole-(permanent) dipole between polar molecules, (ii) permanent dipole-induced dipole between a polar and an apolar molecule and (iii) instantaneous/temporary dipole-induced dipole or London dispersion forces between apolar molecules. The latter occurs when a momentary asymmetry in the electron cloud in one of the molecules (temporary dipole) induces the generation of a dipole in the other molecule (induced dipole)[143]. That is what the attractive/dispersive LJ term (power of 6) tends to mimic, the London dispersive forces. Yet, when two particles are attracted by a force, and they reach an equilibrium distance (null derivative, null force), they start to feel a repulsive force because of the overlapping of their orbitals. That is what the LJ repulsive term (power of 12) aims to describe: the inviolability of the Pauli exclusion principle – 2 electrons shall not have the same quantum numbers[143,144].

Equation 2.12 defines the Lennard-Jones 12-6 potential between a pair of particles i,j (Jmol^{-1}, often used kJmol^{-1}):

$$V_{L-J}(\ell) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\ell}\right)^{12} - \left(\frac{\sigma_{ij}}{\ell}\right)^{6}\right] \quad (2.12)$$

Here, ℓ is the distance between the centre of two particles (SI unit: m, often used: nm), ε_{ij} is in J mol^{-1} (often used: kJ mol^{-1}) and σ_{ij} is in m (often used: nm). Here, ε_{ij} is the depth of the potential well, in other words, how strong is the attraction between two particles and σ_{ij} is the distance when the potential is 0. It reflects the size of particles.

In CG, the repulsion term has been shown to be too repulsive[130,145,146], so a 6-9 potential, for instance, may be more accurate. Although the use of a 6-9 potential could be more accurate, it would

be more costly[109,120,122,139,144]. The 12-6, is widely used, since the repulsive term is basically the attractive term squared, and so computational burden is reduced[127,129,138].

$$V_{Coulomb}(\ell) = \frac{Q_i Q_j}{4\pi\varepsilon_0\varepsilon_\ell \ell} \tag{2.13}$$

Equation 2.13 describes the Coulomb potential (electrostatic properties) between a point charge i and a point charge j (Jmol^{-1}, often used kJmol^{-1}). ℓ (SI unit: m, often used: nm) is the distance between the centre of the two particles, ε_0 is the vacuum permittivity (the ability of a force field to permeate in vacuum, ≈8.85X10^{-12} s^4 A^2 kg^{-1} m^{-3}) and ε_ℓ the relative dielectric constant/relative permittivity (how easily a medium can be polarised by an electric field). Partial charges (Q) are in C (Coulomb) or A.s (SI, ampere second). If Q_i and Q_j have the same sign, they repel each other, whereas if one is negative and the other is positive, they attract each other.

Note that the medium can facilitate or hamper the encounter of the two point charges. For instance, the potential energy of two separated charges in water is reduced by almost two orders of magnitude in comparison to those separated in vacuum[147]. Looking at **Equation 2.13**, and bearing in mind that the water relative permittivity is very high, the Coulomb potential is lower. Thus, water hampers charges from encountering each other.

In some CG models, we use a water bead with no charge, so the encounter between two charges is not hampered. Thus, these interactions are much stronger. To tone down this use, we use a relative permittivity constant that states that all of these interactions should be 15 times lower than they actually are[148].

To sum up, we can look at the shapes of some equations:

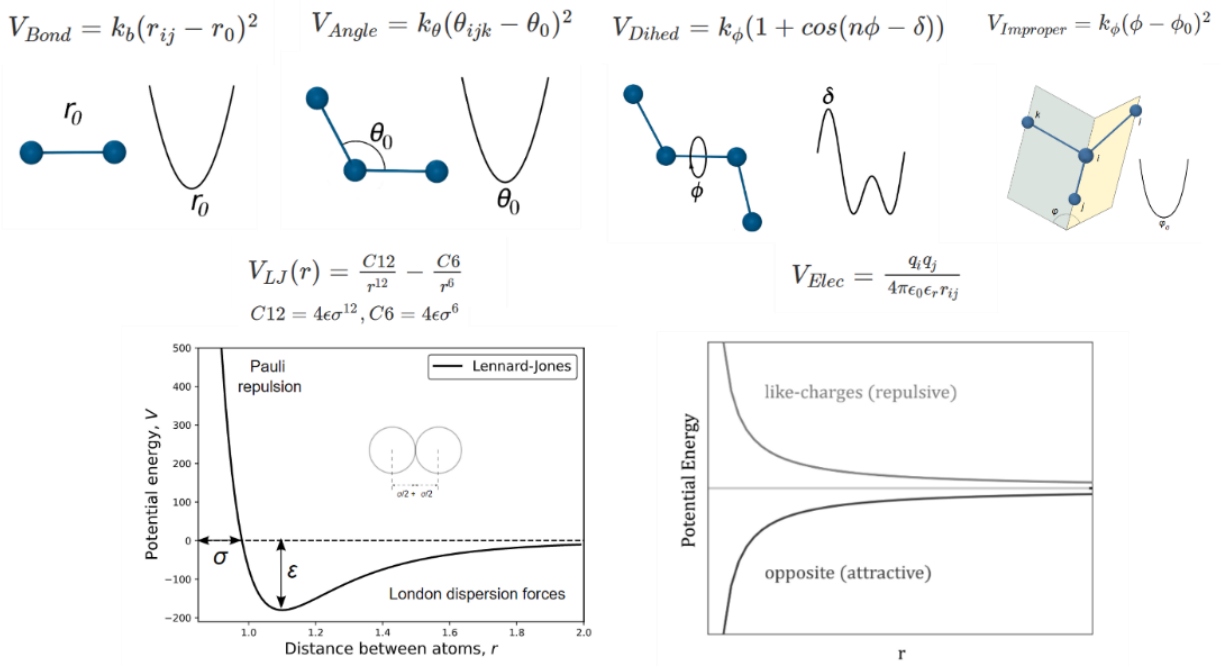


Figure 2.7 – Illustration that shows the shape of each potential interaction based on its formula (examples). Here r is the distance between the centre of two particles, and δ is the phase factor. Harmonic potentials are shaped as a parabola with its reference value as the minimum of the plot. For simplicity, general potentials are shown. Adapted from Force fields and Interactions website[125].

These potential terms build up the potential energy function (force field) necessary for performing MD simulations.

2.4 CG Martini 3 force field

Within CG frameworks, Martini has become a widely popular model[135], with a number of successful applications involving, among others[133,134,149], membranes[150], nucleic acids[151], and energy metabolism[152–155].

Twenty years ago, the first prototype Martini model (version 1.0) was released to study lipid vesicle formation and fusion[135,156]. It simplifies molecular models by typically representing (mapping) 4 non-hydrogen atoms as single particles or beads and can achieve up to 1000-fold speed-ups over AA simulations. This force field relies on a building-block approach, meaning that equal moieties, even in different molecules, are usually parameterised with the same bonded and nonbonded interactions parameters[153].

In 2021, the third version of Martini was released[137]. It addresses several drawbacks[157] of the previous iteration[157]. Of those, it is worth pointing out: (i) need for a greater variety of bead types arranged by polarity to get an accurate recovered chemical behaviour of the group being mapped, (ii) being tied up to about four atoms to be mapped into one bead, which would not allow to accurately distinguish different chemical group sizes. This was all thoroughly accounted for in this latest version of Martini.

Martini 3[137] discriminates the chemical spectrum into 29 different bead types (*vs*. Martini 2 18 types[135]) and is now explicitly parameterised to also allow 2-to-1 and 3-to-1 mappings — the use of mappings finer than 4-to-1 had been a source of incorrect and artefactual Martini 2 behaviour[157], and thus a limitation to the applicability of Martini. This was only possible because in Martini 3 we now have three different bead sizes, contrarily to Martini 2 which only had one. These three different bead sizes are tiny (T), small (S) and regular (R), in ascending order. These bead sizes represent, in general, the mapping of the different numbers of non-hydrogen atoms, where R beads map four atoms to one bead, or simply 4-1, S maps 3-1 and T maps 2-1. Thus, these beads can represent fewer than four atoms. This depends on the underlying planarity, branching, atomic radii and aromaticity[137] as shown in the next figure:

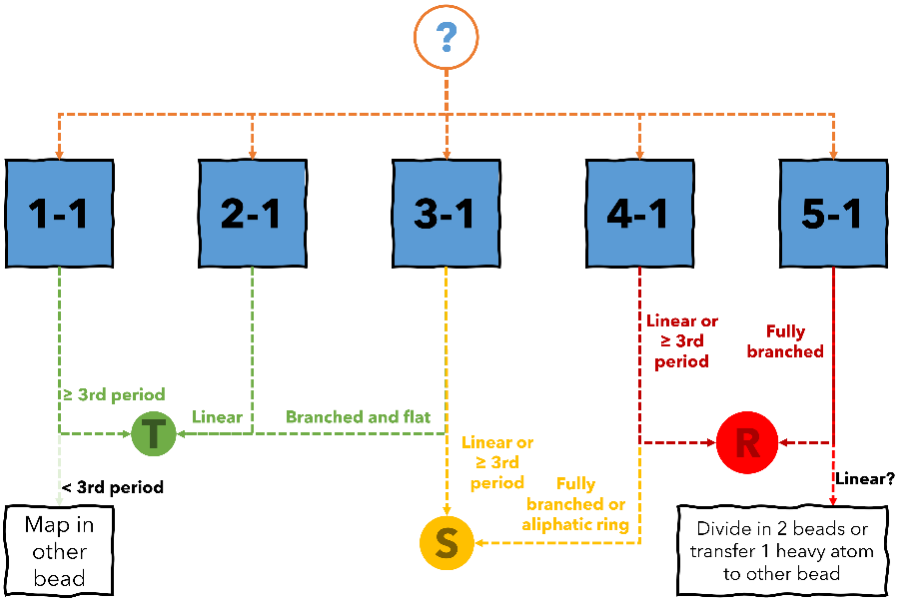


Figure 2.8 – Illustration that shows how to choose the bead type according to the number of atoms being mapped and the geometric characteristics of the chemical group. T stands for "Tiny", S for "Small", and R for "Regular beads". Adapted from Souza *et al.*[137].

With these improvements, coupled with other bead type variations, Martini 3 now boasts 843 bead type-size combinations compared to Martini 2 39[135].

Furthermore, and to mimic with accuracy the different chemical behaviours, six different bead types exist, which may be used for parameterisation: polar (P), intermediate/non-polar (N), apolar (C), halo-compounds (X), monovalent ions (Q) and divalent ions (D). Also, a bead exclusive for water (W) exists. Additionally, P, N and C beads have six subtypes concerning the degree of polarity (from 1, less polar, to 6, more polar). This way, for instance, a TN5 bead is less polar than a TN6 bead, which is less polar than a TP1 bead. There is only one D bead, but there are five subtypes in Q beads and four subtypes in X beads, with the same reasoning regarding polarity[137]. This is summed up in the next figure:

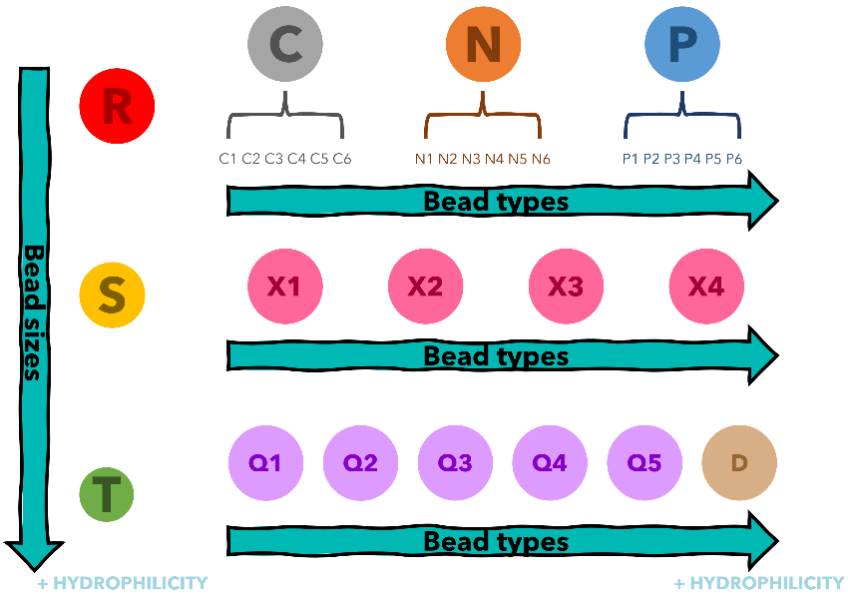


Figure 2.9 – Illustration that shows the different bead types and how they are successively more hydrophilic. According to their bead size, beads are successively more hydrophilic, in descending order (R or Regular, S or Small and T or Tiny). Polar (P), intermediate/non-polar (N), apolar (C), halo-compound (X), monovalent ion (Q) and divalent ion (D) beads are represented, where C < N < P, 1 < 2 < 3 < 4 < 5 < 6 and Q < D in terms of polarity. Adapted from Souza *et al*.[137].

To add even more specificity and increase the accuracy in mimicking the chemical behaviour, labels (specific subtypes) were added. Hydrogen bonding labels ("donor": d, "acceptor", a), if atoms in a chemical group are capable of establishing hydrogen bonds. Electron polarizability labels ("electron richer": e, "electron poor": v), whether chemical groups have a lesser or greater electron density. Others were also introduced to add a new level of accuracy in Martini models[137].

2.5 Computational main knowledge gaps

An important consequence of how the Martini 3 bead type expansion was implemented is that Martini 2 models are not directly usable with Martini 3. A bead type conversion/translation is possible but doing only that foregoes any other benefits of the new parameterisation possibilities. They do not make use of finer mappings and a chemical specificity that is found in Martini 3. Additionally, the parameterisation philosophy of Martini 3 now explicitly takes into account bead density distributions and excluded volumes of modelled molecules when deciding mappings, and these considerations cannot be readily retrofitted into pre-existing models. As such, molecular models must be re-developed to be fully usable with Martini 3, and a widespread effort is underway to expand the range of parameterised biologically relevant molecules[133,149]. In this work, we join this effort by developing Martini 3 parameters for nucleotides and quinones. We do so by revisiting the parameterisation what we did for Martini 2[70]— then focused on nucleotides only — and extending it with the quinones (and quinols, most of which had not yet been parameterised to Martini). Nucleotides and quinones parameterisation then allowed the performing of simulations with NDH-2. In addition, all these parameters are now available to all the Martini community to be employed by others in their research.

3 MAIN GOALS

Previously, the biological relevance of studying NDH-2 was shown, particularly the interactions of some substrates with this protein. The modelling of protein-substrates interactions is within the range of time- and space-scales of CG MD simulations, one of the reasons for employing them. Moreover, Martini 3 CG simulations are of great use to mimic *in vivo* proteins' behaviour, especially when they are placed into larger systems such as membranes.

The third version of the Martini force field can now address 2-to-1 and 3-to-1 mappings, which requires the previous parameterisations made in Martini 2 to be reparameterised, so that they account for the benefits brought by Martini 3. Therefore, a considerable amount of time was invested in this new parameterisation, parameterising bonded and nonbonded parameters of substrates used in NDH-2 simulations, such as FAD (cofactor), MNQ and NADH. This, allied with the transformation of NDH-2 AA structure into CG and *S. aureus* lipids parameterisation, was the initial driving force for the studying of several possible interactions occurring *in silico*, including in a membrane.

With this brief overview, the main aims of this thesis can be summarised as follows:

- Parameterisation of a set of molecules, some of which were used in the aqueous and membrane CG MD simulations of this thesis.
- Availability of nucleotides and quinones parameters to the Martini community to provide future research using these molecules since they possess countless and notorious biological functions.
- Validation of these parameters, including simulations of some of them with NDH-2.
- Modelling and study of NDH-2-substrates interactions.
- Analysis of substrate-enzyme binding.

The conclusions retrieved from this thesis may lead to future antimicrobial therapies against the pathogenicity of *S. aureus* in a way that the knowledge gap regards to protein-substrate interactions may be covered. This thesis may contribute by other studies of the Martini community on the health-related systems and expand the knowledge on NDH-2.

4 METHODS

In the Introduction to the Methodology Chapter, we saw the generic aspects of MD simulations. In this Chapter, some in-depth details will be given to better understand these simulations and how they work. Additionally, particularities on how the CG setup is made, namely the chosen simulation box, the addition of our molecules of interest into that box, solvating (and neutralising) it, minimisation, equilibration and production will be given. Some parts of this Chapter will be part of a future publication.

4.1 System creation

In this thesis we prepared the following CG systems:

(1) NDH-2 alone in water;
(2) NDH-2 with menadiones (menaquinone heads) in water;
(3) NDH-2 with other quinone/quinol headgroups in water;
(4) NDH-2 with menadiones and NADH in water;
(5) NDH-2 with menaquinones in *S. aureus* membrane;
(6) NDH-2 with menaquinones and NADH in *S. aureus* membrane.

Prior to the setup of these systems, the parameterisation of molecules was done.

4.1.1 Small molecules: from an AA structure to a CG structure

Small molecules structures such as FAD or MNQ were obtained using the SMILES[158] code, from either PubChem[159] or ChemDraw[40]. Then, they were converted to a PDB file using the Avogadro software[160].

These files were then converted into a GRO file format using *gmx editconf*. This format is similar to a PDB file, lacking some crystallographic details. The coordinates from the NDH-2 PDB file (GitHub script[161]: *FAD_coordinates.ipynb*) were added to the FAD GRO file, after adding the missing hydrogens.

To define the mapping, these structures were then uploaded to CG Builder[162], an online tool from which we can retrieve an NDX file that contains information on which atoms belong to each CG bead whose placement will be thoroughly explained in Chapter 5. NDX files are crucial to group atoms into particles. Later, using *gmx traj*, we got a coarse-grained structure, from each group centre-of-weight (COW).

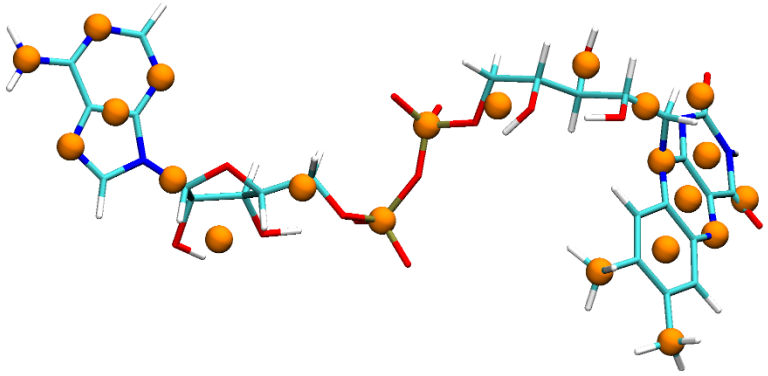


Figure 4.1 – Illustration that shows the FAD atomistic structure and the Coarse Gain beads. In cyan are represented the carbons, in dark blue the nitrogens, in white hydrogens, in red oxygens and in brown phosphates. The orange spheres are the CG beads. Picture rendered using VMD (Visual Molecular Dynamics)[163].

4.1.2 NDH-2: from an AA structure to a CG structure

The coarse-graining process is simplified for proteins. There is a tool that we can use to obtain a protein CG structure from its AA. This tool is called *Martinize2*[164], and it maintains tertiary and quaternary protein structures when combined with the use of elastic networks. Elastic networks are structural scaffolds because they are harmonic bonds that are added to keep the protein structure[165].

Martinize2 creates a topology file. Here, protein parameters are generated. As an input, we gave a PDB file of the NDH-2 AA monomer (5NA1[88]). Hydrogens were added to this PDB file using the online MolProbity tool[166] from Duke University. From the *Martinize2*[164] tool, we obtained a GRO file, but also ITP* and TOP files*.

FAD beads were added to this NDH-2 CG GRO file and FAD parameters to the NDH-2 CG ITP file. To maintain FAD stable inside the protein, the bead type and charge of the side chain of histidine-44 was changed from TN5a (with no charge) to TQ2p (with charge +1) mimicking a protonation. Although we did not check how prone this amino acid is to be protonated, we believe this happens due to its proximity with the phosphate group. This allowed this amino acid residue side chain to interact with the negative phosphate beads of FAD and, by that, keeping FAD in its *in vivo* place. Besides, we added a few protein-FAD bonds to make sure the latter would stay inside the protein since, in previous simulations without them, FAD would get out of the protein. These bonds were chosen after obtaining the CG structure from *Martinize2*. They were chosen due to their proximity to FAD with a distance below 0.5 nm in that *Martinize2* CG structure. The code that lists which bonds were below of a 5 Å distance is *ExclusionsAndBonds.ipynb* and can be found in my GitHub[161].

In addition, some FAD-protein beads were already too close, thus in a very repulsive regime, leading to clashes and consequently system crashing. Therefore, exclusions were added to these beads, meaning they would not possess any nonbonded interactions between each other. The same script was used (*ExclusionsAndBonds.ipynb*) to detect which distances were below the LJ's σ coefficient and that could be the reason for the system to explode.

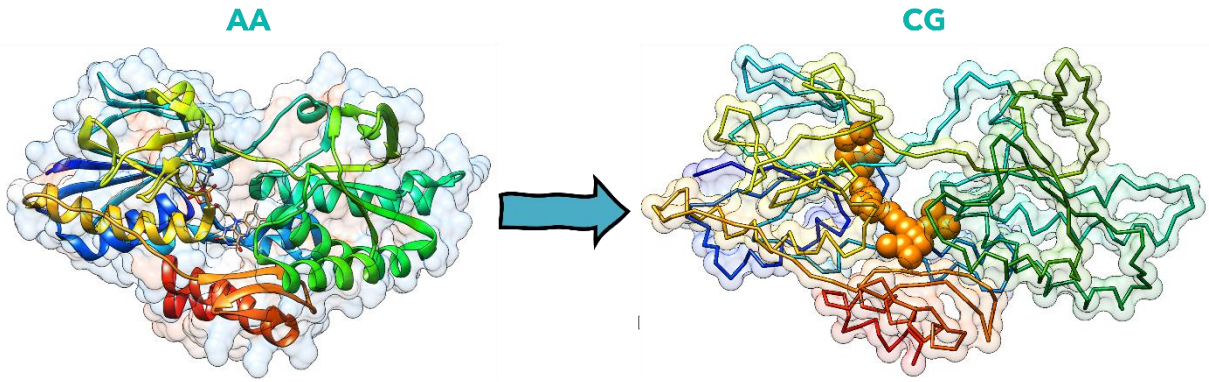


Figure 4.2 – Illustration that shows the outcomes of the processes described in sections 4.1.1 and 4.1.2: from the NDH-2 monomer AA structure (left) to NDH-2 CG monomer (right) and its cofactor (FAD). In the atomistic structure, the hydrophobicity surface is represented. In polar regions and in apolar regions. The orange spheres are the CG beads. Pictures were made using Chimera software[89].

*ITP stands for "include [in] topology" since these files are the pillars to generate potentials between particles. In ITP files, we can find information of the molecule parameters. Plus, the inputs of the force field, such as LJ coefficient values, are displayed in files of this format. The TOP files are topology files that possess information on the number of molecules we have in a system[140].

4.1.3 Preparing the simulation box

All the simulations in this thesis were made using explicit solvation and periodic boundary conditions (PBC). The former means that all molecules of a solvent were explicitly added *i.e.*, their movements and effects are considered within a given region around the solute molecules[167,168].

With regards to boundary conditions, PBC is a way of overcoming one of the main drawbacks of computational simulations: computational burden. It enables the calculation of macroscopic properties using a relatively low number of particles. In a PBC box, when the particle leaves one of the edges of the box, it enters again in the same box from the opposite side. The other boxes are just "periodic images" (8 in 2-dimensions/D, 26 in 3D) so that (i) we do not lose any particles and (ii) no forces from the sides of the box are applied. Note that if we had rigid edges in a box, the resulting forces applied to the particles in the centre of the box would be different from those applied to particles on the box edges. Thus, the system seems infinite but at the cost of periodicity effects[109,123]:

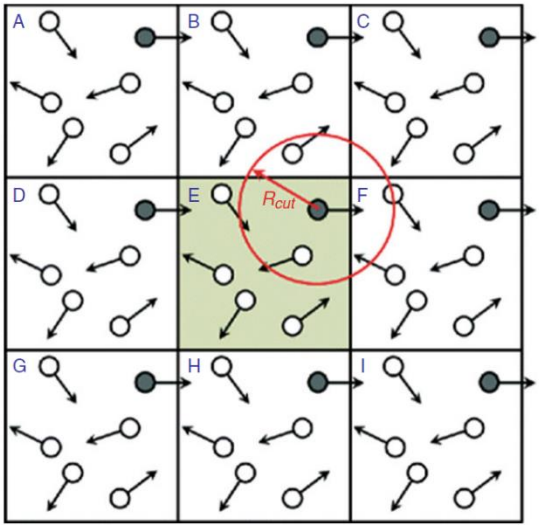


Figure 4.3 – Illustration of a two-dimensional box with PBC applied where the actual square being simulated is "E", but it can interact with all the others surrounding it, according to the r_{cut}. Retrieved from Zhao *et al.*[169].

These unit cells can have different shapes, such as cubic or dodecahedral. The former ends up having solvent molecules in the corners that are not needed, culminating in larger times of simulation[170], whereas the latter mimics best a sphere to save computational time, thus, this shape was chosen for all aqueous simulations present in this thesis. Ideally, we would use a sphere to not simulate edge areas out of the molecule radius cut-off. Yet, PBC requires a shape that will make a 3D crystal and spheres can't be packed to fill that space[111,170].

In practice, we first create a dodecahedral box with NDH-2 inside using *gmx editconf*, followed by inserting other molecules we want to simulate using *gmx insert-molecules*. Then, this box is solvated using *gmx solvate* and neutralised to a salt (NaCl) concentration of 0.15 M, where the number of ions to be added is calculated using the following formula:

$$\frac{4 \times \text{number of water beads} \times 0.15\ \text{M}}{55.5\ \text{M}} \tag{4.1}$$

Equation 4.1 gives us the number of sodium and chloride ion beads so that we get a final NaCl concentration of 0.15 M. These ion beads are calculated from their relative concentration in water. In this thesis, we applied an NPT ensemble, where the number of particles, pressure and temperature are kept constant by external baths and where the volume varies. Thus, calculating the ionic concentration

from a non-constant solvent volume would not make sense since the volume is always changing and, thus, also the ionic concentration. This way, these are calculated in relation to the number of water beads that do not change throughout the simulation. Since we are working with CG simulations where one water bead (W) corresponds to 4 atomistic waters, we had to multiply it by four. Finally, we divided by 55.5 M, which is the concentration of water in water. These ion beads are added to the system using *gmx genion*.

4.1.4 *Insane* method: NDH-2 into *S. aureus* membrane

We have covered how to prepare a system to be simulated in an aqueous environment. These systems were a target of multiple analyses, but so were membrane simulations. Membranes are generated using a script called *insane.py*[171], which creates a box and adds the lipids along a bilayer according to their total percentages, and solvates and neutralises the system if we want. It generates a structural (GRO) file and a topology (TOP) file and allows us to place a protein into the membrane. We use it to create membranes; in this case a membrane with the main *S. aureus* lipids. With *insane.py*, we placed NDH-2 1 nm above the membrane and then we reoriented it in space into the position it should be in the membrane using another script called *NDH2Rotation.ipynb* that can be found in my GitHub[161].

4.2 Minimisation

The potential energy surface (PES) of a system is an energy function of the particles' coordinates. Thus, PES is a multidimensional "surface" with changes in energy. On this surface, the interesting points are the minima, corresponding to more stable configurations. Of these, the one with the lowest energy is known as global energy minimum[109,172]:

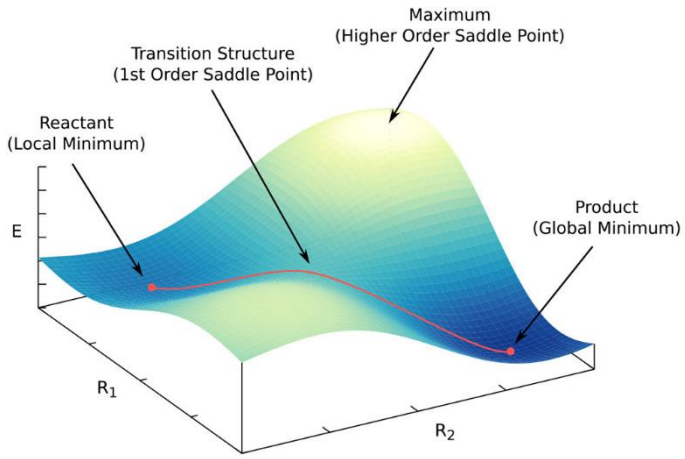


Figure 4.4 – Illustration of a system PES with the two coordinates R_1 and R_2. These minima can be connected by paths (red line) where rearrangements and reactions can occur (transition structures). These transition states are first-order saddle points, a maximum in one coordinate and a minimum in all others. The energy of the transition state is the minimum energy required to transition between two PES minima. Retrieved from Keith *et al.*[172].

The process of energy minimisation is needed for our simulation to start in a more stable state where forces are not too high which could lead to the system explosion. A minimisation algorithm is used for this purpose: finding a minimum of the potential energy surface from which we can start our simulation[109,123,140].

In this thesis, the algorithm used was the steepest descent algorithm (**Figure 4.5**). At each iteration, the direction chosen is the steepest. This cycle repeats itself until it finds a minimum that, in all directions, has a maximum absolute value of force (gradient) in all its components smaller than a specified threshold given by the user or after the running of a certain number of force evaluations specified by the user[109,123,140]:

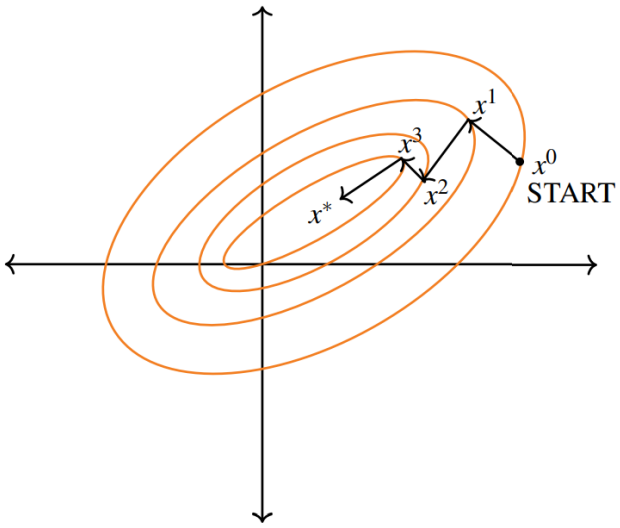


Figure 4.5 – Illustration of the Steepest Descent algorithm. The steepest direction is taken until we get to x*, the minimum point with the lowest energy. Note that this representation is in 2D. Retrieved from Mishra & Ram[173].

In CG, the minimisation process is faster since the energy landscape is smoother, in other words, because there is a smaller number of energy barriers between the different states[174]. This process generates the set of initial coordinates to be utilised for the MD simulation, the GRO minimised structure file*[129].

4.3 Equilibration

The process of equilibration, also known as relaxation, results in a relaxed structure suitable for the beginning of a simulation at given thermodynamic properties, namely temperature and pressure. In an analogous manner to minimisation, the TPR file is generated using the *gmx grompp* command with the GRO file resultant from the minimisation stage. The relaxed GRO structure is obtained from the *gmx mdrun* command[140,175,176].

The system is coupled to a thermostat and a barostat. These baths can adjust the motion of particles or alter their Newtonian equations of motion[109,119]. The thermostats and barostats used for coupling will now be addressed.

*To obtain a GRO file, we need the solvated GRO file, the TOP file and an MDP file. This MDP (Molecular Dynamics Parameters) file consists of the parameters that are going to be used during the MD process, containing information such as the energy minimisation algorithm, thermostat and barostat used, the time step (Δt), the number of steps, r_{cut}, the temperature to be maintained, among others, depending on whether we are performing minimisation, equilibration or production (the MD process itself). To perform minimisation, equilibration and production, we use the *gmx grompp* command to generate the final TPR file, a binary file which contains both coordinate and topology information. Then, from *gmx mdrun*, the GRO file with the positional coordinates is created[140].

4.3.1 Temperature coupling

The process of thermal or kinetic equilibration of all system particles to a thermal bath at a given temperature was ensured by a Velocity-Rescale temperature coupling thermostat[140,177] in all systems in this thesis.

This strong coupling thermostat is a derivation of the weak Berendsen[178] first-order decay coupling thermostat in which the temperature is not absolutely fixed, but an exponential relaxation of its instantaneous temperature is implemented instead[109,119,140,179]:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau_T} \Leftrightarrow T(t) = T_0 - Ce^{\frac{-t}{\tau_T}} \quad (4.2)$$

Equation 4.2 shows the first-order decay Berendsen coupling thermostat, where T(t) is the instantaneous temperature in Kelvin as a function of time (t, in seconds), T_0 is the target/bath temperature, C a constant and τ_T the coupling parameter (time constant) from which we can conclude how strongly the system is coupled with the "thermal bath" (the larger the τ_T, the weaker is the coupling and slower the decay, allowing more significant temperature fluctuations).

While the Berendsen thermostat modifies the velocities of particles in the direction of the desired temperature, with Velocity-Rescale (V-Rescale), these modifications are made to give the exact desired temperature at each step. The latter also has the advantage of ensuring a correct kinetic energy distribution[109,119,140,179].

4.3.2 Pressure coupling

In an analogous manner to temperature coupling, pressure can be regulated by a barostat, which varies the volume. Putting it in another way, the size of the system changes by scaling the coordinates of each particle. In this thesis, for equilibration steps, Berendsen's barostat was used, which follows the exact same equation as Berendsen's thermostat[178] seen in the previous section[109,119,140,180]:

$$\frac{dP(t)}{dt} = \frac{P_0 - P(t)}{\tau_P} \Leftrightarrow P(t) = P_0 - Ce^{\frac{-t}{\tau_P}} \quad (4.3)$$

Equation 4.3 defines the first-order decay Berendsen coupling barostat, where P(t) is the instantaneous pressure in bar as a function of time (t, in seconds), P_0 is the target/bath pressure, C a constant and τ_P the coupling parameter (time constant) from which we can conclude how strongly the system is coupled with the "pressure bath" (the larger the τ_P, the weaker is the coupling and slower the decay, allowing larger pressure fluctuations).

Note that the scaling of coordinates according to the pressure targeted may be done isotropically (meaning, the same scaling factor is applied in all three directions), semi-isotropically (the scaling is the same for x and y axes and different for the z axis) and anisotropically (independent scaling for the three directions). Regarding membrane simulations, a semi-isotropic coupling was applied to allow pressure adjustments in membrane and in water, independently. There, the system pressure is dominated by the membrane lateral pressure (x and y axes) and the solvent pressure (z axis). For aqueous environment simulations, isotropic coupling was employed, where pressure was adjusted equally in the three directions (x, y and z).

In equilibration, molecules already move by applying Newton's laws. Here, they move to allow temperature and pressure to stabilise.

4.4 Production

The production process in MD simulations is the most time-consuming phase, and that from which we retrieve our main conclusions. Like for equilibration and minimisation, *gmx grompp* command is used to generate the TPR file, using as well to the GRO file that comes from the equilibration step. Then, *gmx mdrun* command is used to start the simulation. At the end of the simulation an XTC file is written, which holds information on particles' trajectories.

4.4.1 Numerical algorithm integrator

We have seen that the particle positions and their velocities are updated by applying Newton's laws. Yet, solving a set of second-order differential equations for three or more particles is impossible, so trajectories are obtained from a finite-difference approach using discrete time steps. Here, molecular coordinates are obtained at $t + \Delta t$ and velocities are obtained at a time $t + \frac{1}{2}\Delta t$ from the molecular coordinates and velocities previously obtained (at an earlier time, t and $t - \frac{1}{2}\Delta t$, respectively). The equations are solved on a step-by-step basis. The Δt choice depends on the molecular system simulated potentials. The higher the force constants (AA > CG), the larger is the frequency[120,123,140]. To properly describe those frequencies, we need smaller time steps. Thus, in AA is smaller than in CG[136].

Throughout this thesis, the same software to perform MD simulations was used: GROMACS[140]. The GROMACS algorithm by default is the leap-frog algorithm, and that was the one employed:

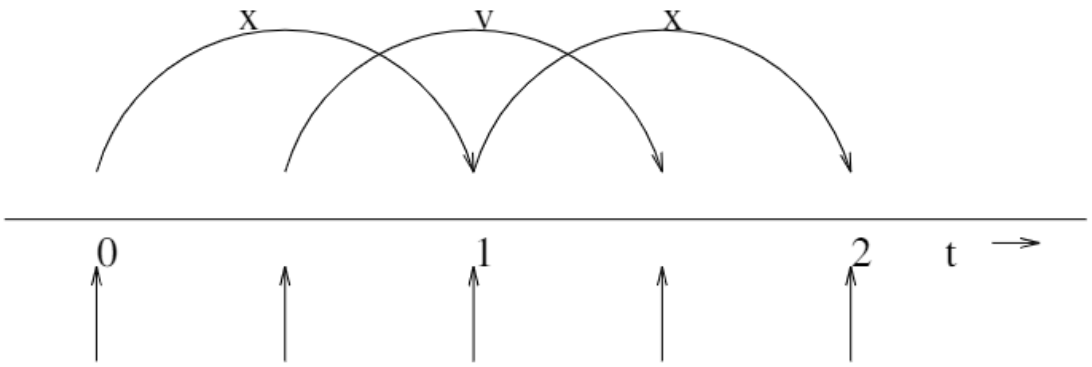


Figure 4.6 – Illustration of the Leap-Frog integration method. The name comes from the fact that **r** (in this picture as **X**) and **v** are leaping like frogs over each other backs. Retrieved from GROMACS manual[140].

So, according to this half-step algorithm, to calculate the positions at a step $n + 1$, the discrete form is[120]:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \mathbf{v}_{n+\frac{1}{2}}\Delta t \tag{4.4}$$

Equation 4.4 defines the positions at step $n + 1$, where **r** is the particle position vector and **v** is the particle velocity (in m/s) at a time $t + \frac{1}{2}\Delta t$ (in s).

Velocities are evaluated at the mid-point of the position. Thus, from **Figure 4.6**, the velocity at a time $t + \frac{1}{2}\Delta t$ can be crudely estimated as follows:

$$\mathbf{v}_{n+\frac{1}{2}} = \mathbf{v}_{n-\frac{1}{2}} + \frac{\mathbf{F}_n}{m}\Delta t \tag{4.5}$$

Equation 4.5 defines velocity (in m/s) at the step $n + 1$, where $\mathbf{v}$ is the particle velocity, $\mathbf{F}_n$ is the force of step n (in N), m is the particle mass (in kg) and Δt the time variation (in seconds).

With **Equation 4.5**, the integration algorithm is fully described.

4.4.2 Temperature and pressure coupling

In production, I made use of the same temperature coupling as in equilibration (described in section 4.3.1). Nonetheless, a dissimilar barostat was implemented in Production– the Parrinello-Rahman barostat[181]. In this method, the difference between the internal pressure and the external stress results in a second-order differential equation for the box vectors, enabling changes in the shape, size, and volume of the simulation cell[119,140,180,182,183].

4.5 System setup and simulation details

With all simulation steps acknowledged, we can summarise our steps in a final figure for a visual understanding on how a computational MD system is built up and simulated:

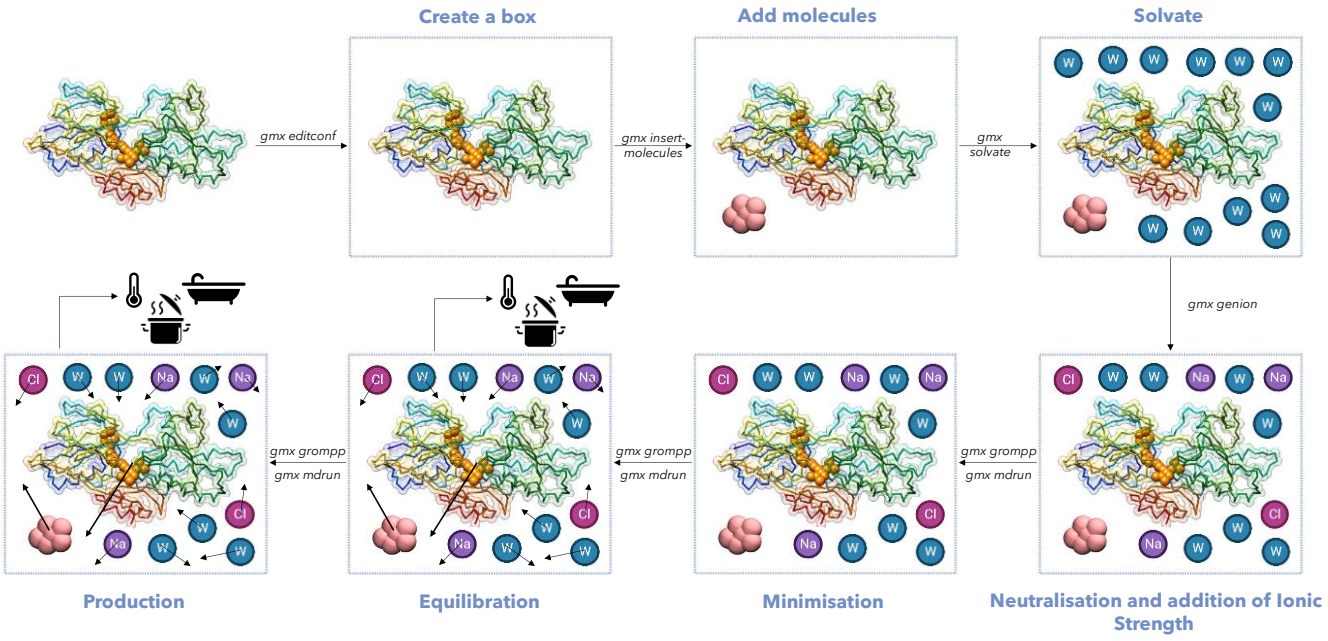


Figure 4.7 – Illustration that shows the overall steps of a MD CG simulation. In this case, it is represented by the simulation of NDH-2 monomer (with FAD) with menadiones (menaquinone heads, in pink) in water. To simplify, squares were used to represent the simulation boxes and for molecules represented by several beads, only the average velocity vector was reproduced in the equilibration and production states. Picture made using Chimera software[89], VMD[163] and BioRender[27].

We can now condense all the simulation details, whether they are UA, aqueous CG, or membrane CG simulations.

4.5.1 UA simulations

UA simulations were essential to obtain the best CG parameters and will be addressed in Chapter 5.

All UA simulations used the GROMOS54a7[184] force field. Each molecule was inserted into a 5 X 5 X 5 nm^3 dodecahedral box, which was then solvated with water. Then, all systems were minimised and equilibrated. For the production step, we used a Van der Waals cut-off of 1.4 nm and the Verlet list scheme[185] to update the particle neighbour list. Additionally, to maintain the system temperature at 298 K, we used the V-Rescale thermostat[177] with a 0.1 ps coupling constant. Furthermore, the pressure was coupled to 1.0 bar using a Parrinello-Rahman barostat[181] with an isotropic pressure coupling type and a 2.0 ps coupling time. Lastly, geran-8-yl geran (G8YLG) and methyl-geranylbenzoquinone (GBZ) UA simulations were run for 50 ns, whereas ubiquinone-8 (UBQN-8) run 100 ns (molecular structures in **Figure 4.8**).

To these united-atom trajectories, missing hydrogens were added to allow for proper excluded volume evaluation and matching*.

4.5.2 Aqueous CG simulations

For nucleotide molecules, we used a top-down approach (which will be addressed in Chapter 5) since we already had the larger molecules' united-atom (UA) trajectories from our previous paper[70]. This means that, after obtaining CG parameters for the larger molecules, we were able to fragment those same parameters for smaller molecules, whose moieties are the same as the larger ones. Putting it differently, the smaller molecules correspond to or possess the moieties presented in the larger ones, and those parameters are equal.

Nucleotides' mappings were performed using a centre-of-weight (COW) approach, where the beads placement depended on the relative weights of the composing atoms of that bead. We only used integer weights, which were established by typing their index as many times as the weighted we wanted to attribute in an NDX file, where atoms are grouped into beads, to allow for accurate accessible surface area matching between AA and CG structures (further referred to in Section 4.7).

For quinones, we applied a bottom-up approach (tackled in Chapter 5), where the behaviour of small moieties was required to build and define the parameters of larger molecules. This indicates that, from the two fragments addressed in the last section, we could obtain the CG parameters. The tail fragment (geran-8-yl geran) and the smaller tail linked to a menadione-like molecule (methyl-geranylbenzoquinone) UA topologies were gotten from the Automated Topology Builder (ATB)[186]. From the former, we found the CG parameters to mimic the chain behaviour, whereas from the latter, the CG parameters that better represent the head-tail movements. These parameters were adjusted by using the ubiquinone (UBQN-8) UA parameters retrieved from de Jong *et al.*[152]. Besides, for UBQN methoxy group behaviour to be successfully well represented, these parameters had to be employed. For HQNO, NDH-2 inhibitor, we got the chain parameters from Martini 3 octane topology[137]. This way, we effectively got CG parameters for plastoquinone-8 (PQ-8), menaquinone-8 (MNQ-8), ubiquinone-8 (UBQN-8) and 2-heptyl-4-quinolinol 1-oxide (HQNO, NDH-2 inhibitor[187]). For the former mapping, and since it lacks an adjacent methyl, when using UBQN molecule to tune the CG parameters, the methyl was disregarded. Quinones CG mapping was made by taking into account all atoms, as opposed to the nucleotides approach. The chemical structures of the mentioned molecules are represented in **Figure 4.8**.

* Please be aware that sometimes in this thesis we use UA and AA interchangeably to mention these UA molecules to which hydrogens were added.

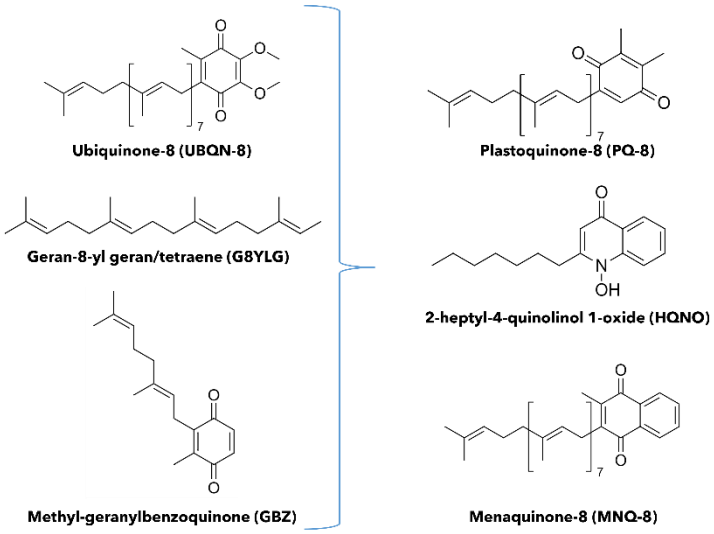


Figure 4.8 – Illustration that shows the starting points of quinones parameterisation. Two fragments from ATB[186] were used - G8YLG and GBZ– and UBQN-8 UA topology retrieved from de Jong *et al.*[152] (left) to accurately parameterise PQ-8, HQNO and MNQ-8, on the right. PubChem CID or SMILES (Simplified Molecular Input Line Entry System) codes can be found in **Table S1**.

All of these simulations were run employing the 2019.6 version of GROMACS[188] and the Martini 3 CG force-field model for biomolecular simulations[137]. Both nucleotides and quinones were solvated in water (W) in a 5 X 5 X 5 nm^3 dodecahedral box. These systems were then minimised and equilibrated. In production, a stochastic dynamics scheme[189] was used with a time step of 2 fs and a friction constant of 0.1 ps^{-1}. Temperature was coupled at 298 K. Furthermore, the pressure was coupled to 1.0 bar using a Parrinello-Rahman barostat[181] with an isotropic pressure coupling type and a 12.0 ps coupling time. Finally, all of the CG simulations needed to get the final distributions ran for 1 μs and the rest of aqueous simulations with NDH-2 and different substrates ran for at least 5 μs.

For nucleotides and smaller quinone moieties, we checked the probability distributions for the bonds, angles and dihedrals potentials. This way we can compare the preferential states of a molecule in UA to CG. In a trial-and-error fashion we set CG potentials, so that the UA behaviour is recovered.

4.5.3 Membrane CG simulations

Lipid membrane systems were created using the *insane.py* CG building tool[171], where lipids were placed in each leaflet in equal amounts. The proportion of each group of lipids added is shown in the next table:

Table 4.1 – Table that shows the different amounts (%) of each lipid (and for the different number of carbons of each chain) in the *S. aureus* membrane.

Lipids	Percentage (%)
LPG	
17:15	14.85
15:15	4.95
PG	
17:15	37.125
15:15	12.375
DAG	
17:15	18.5625
15:15	6.1875
CL	
17:15	4.95
MNQ	1.0

The structures of these lipids are similar to each other as shown below; therefore, a Martini building-block approach could be employed:

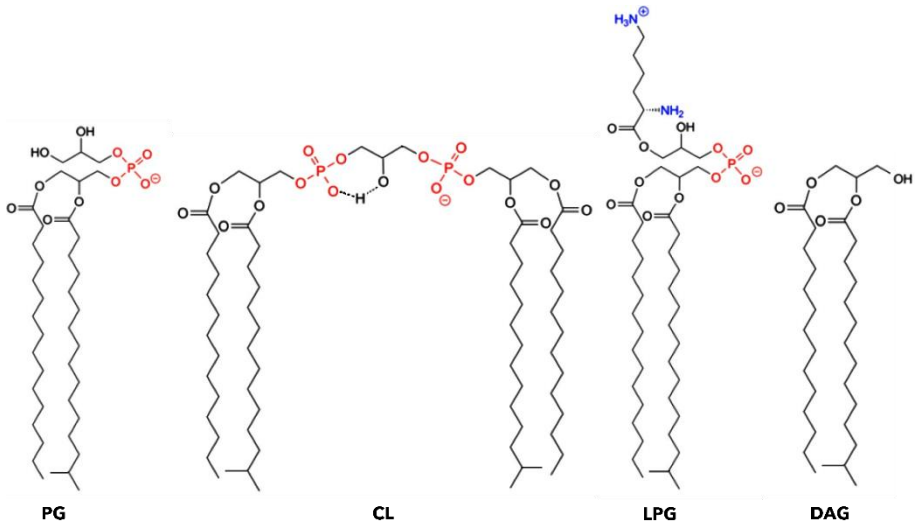


Figure 4.9 – Illustration that shows the four main components of the *S. aureus* lipidic membrane. The number of carbons here represented for each lipid does not represent the number of carbons employed in this work. It is just a picture to show the similarity between PG, CL, LPG and DAG. Adapted from Rehal *et al.*[190].

To mimic the unsaturations role in other organisms, *S. aureus* uses methyl branched groups. These methyls are crucial to keep its membrane fluidity. Thus, an angle kink of 150º was applied to the last bead of each lipid chain. When NADH was to be added, it was included in the exact same amount as MNQ. Afterwards, these membranes were solvated, neutralised with Na^+ beads so that the membrane system would be neutral and Na^+ and Cl^- beads were added to get a final concentration of 0.15 M. These simulations were run, making use of the 2019.6 version of GROMACS[188] and the Martini 3 CG force-field[137].

Nonbonded interactions were cut off at 1.1 nm, and Coulomb interactions were treated using reaction-field electrostatics[145] with a dielectric constant of 15. A Verlet list scheme was utilised to update the particle neighbour list[185]. A V-Rescale thermostat was employed[177] with a coupling time of 1.0 ps to

maintain the temperature stable at 293 K (gel-to-fluid transition of *S. aureus* membrane). Constant pressure was semi-isotropically coupled to 1.0 bar using the Parrinello-Rahman barostat[181] with a relaxation time of 12.0 ps. After the initial energy minimisation and equilibration steps, simulations were run with a 0.02 ps time step until about 10 μs.

4.6 Alchemical analysis: calculation of log P

For the most straightforward parameterisation of smaller molecules and to see whether our choices were accurately made, we availed ourselves of an alchemical analysis with the final purpose of calculating the octanol-water partition coefficients (log Ps or log K_Ps) to check our bead type choices. In this thesis, an alchemical analysis was employed according to the following thermodynamic cycle:

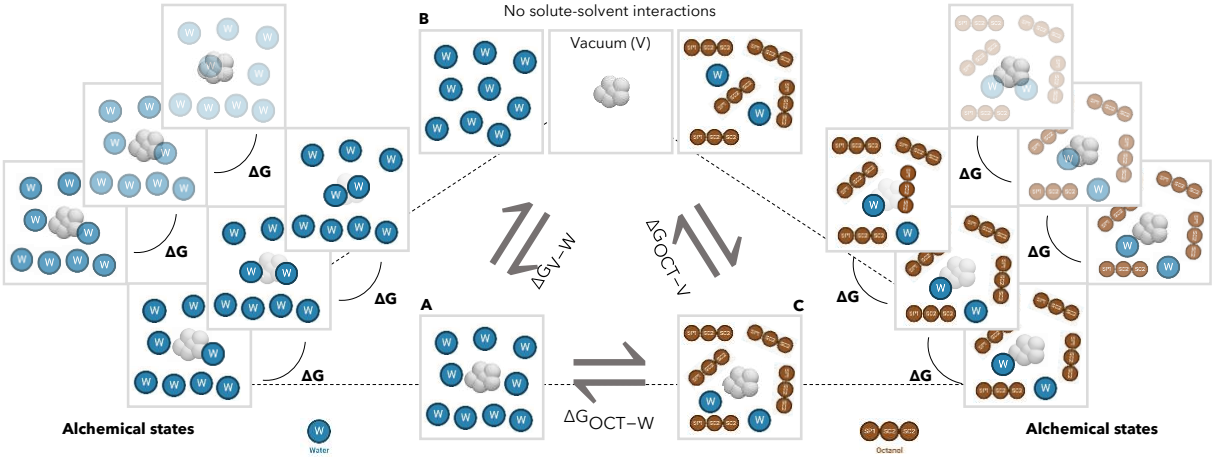


Figure 4.10 – Illustration that shows how an alchemical analysis works and the alchemical states. Octanol is denoted as OCT and represented with its Martini 3 beads, Water as the W bead in Martini and Vacuum (V) as pure blank. The solute (in grey) and solvents can be shown interacting according to different scales (more transparency - nonbonded interactions turn-off). The beads were made using Biorender[27].

In this thermodynamic cycle (**Figure 4.10**), we can summarise the three end states: (A) the molecule interacting with water in a water box, (B) the molecule not having any nonbonded interactions and (C) the molecule interacting with octanol hydrated in a box filled with octanol and water molecules. Since this is a thermodynamic cycle, we can write the following in equilibrium:

$$\Delta G_{OCT-W} = \Delta G_{V-W} + \Delta G_{OCT-V} = -\ln(10)\ R\,T\log(P) \qquad (4.6)$$

Equation 4.6 shows how ΔG (in J.mol^{-1}) and log P are related, where from the former, which results from the sum of the variation of Gibbs free energies between Water (W) and Vacuum (V) and Vacuum and Octanol (OCT) end states, we can get the latter (log P). R is the ideal gas constant (≈8.31445 J mol^{-1} K^{-1}), and T is the temperature in Kelvin (K).

It is common knowledge that Gibbs free energy is a state function. Hence, only the equilibrium endpoints are essential. The path we follow to get there does not necessarily need to be realistic, only the end states. Thus, we compute the so-named alchemical intermediate states, in which decoupling of the nonbonded interactions occurs gradually. This conveys that the sum of all the variations of Gibbs free energy between two alchemical states (in one side of the triangle) is equal to the variation of the Gibbs free energy between the two end states that delimit that side[191].

Here, we employed the Multistate Bennett Acceptance Ratio (MBAR) method[192]. This method calculates the ratios of how probable it is of having a certain configuration from one alchemical state in

another alchemical state, and vice-versa. In other words, evaluating the probability of encountering the exact same configuration structures from state A in a different alchemical state (B), then doing it the other way around (evaluating the energy of particles from state B in state A) and finally, getting their ratio[193].

Using **Figure 4.10** to exemplify this, let us suppose we evaluate state A (solute filled with water in a box with nonbonded interactions turned on) in state B (where nonbonded interactions are scaled to none, as if the solute and solvent could not see each other). In state B, we probably would never have a cavity between the water molecules for the solute to accommodate itself. It is highly improbable that in state B, waters to have such a cavity between them left by the solute and be so far away from one another. This would be non-spontaneous, thus rather highly endoenergetic. On the other way around, the same would happen since overlapping of particles would occur (note that water particles in the region of the solute cavity would overlap with the solute itself). The ratio, done by MBAR, is actually a function of frequencies, thus, when energy tends to infinity, the frequency tends to 0. If we would go for an evaluation of the ratio of state A evaluated in state B, and state B evaluated in state A, thus would result in an indeterminate form of 0/0. This explanation is the intuitive reasoning behind the impossibility of performing these calculations directly between end states. Besides, and looking at **Figure 4.10**, if we were to perform the same evaluation between states A and C and to know that these systems would never be similar since they have dissimilar bead arrangements, we would not be able to perform the ratio. In this situation, finding out three particles side by side (such as in octanol) in water would be almost impossible since water beads are loose and dispersed in the box.

Therefore, this process is done on a step-by-step decoupling basis, using the ratios between alchemical states, where there is a gradual turning off of the nonbonded interactions. These energetic alchemical states are more alike, so configurations from a state are possible in another state, resulting in lower energies, thus, in ratios that can be calculated. The MBAR method evaluates the ratio and does the sum of the ΔGs between the different alchemical states. Then, from **Equation 4.6**, from the ΔG calculated, we obtain the log P of the solute being studied[191].

In a nutshell, octanol-water Gibbs free energies were calculated from individual CG solvation free-energies into water and octanol-water, as thoroughly explained in Sousa *et al.*[70] and above. A stochastic dynamics scheme[189] was used with a time step of 0.02 ps and a friction constant of 1 ps^{-1}. The temperature was coupled to 298 K, whereas pressure was isotropically coupled to 1 bar using the Parinello-Rahman barostat[181]. Single solute molecules were decoupled from solvent boxes with either 898 Martini waters or with 431 octanol molecules hydrated with 38 Martini waters. The decoupling was employed in 11 steps of 25 ns each by scaling down solute-solvent LJ interactions (please note that Martini water and octanol are not explicitly charged, hence do not establish Coulombic interactions with solutes). Finally, the Multistate Bennett Acceptance Ratio (MBAR) method[191] was employed to calculate the partition free-energies, using the pymbar v3.03 Python package[133,192].

4.7 Solvent Accessible Surface Area (SASA)

Bead size choices were evaluated by the calculation of the AA-to-CG Solvent Accessible Surface Area (SASA) ratio. According to the accepted convention, an acceptable value ratio must be below 5 %.

SASA is, as the name suggests, the area of the surface that surrounds a molecule and is accessible by the solvent. This surface is built upon the centre of the solvent particle as it rolls over the Van der Waals surface of the molecule[194,195]. On the contrary, there is another definition intrinsically linked to SASA, the so-named Solvent Excluded Surface (SES) or Conolly surface[196].

SES is the surface that delimits the molecule region that cannot be accessed by the solvent, and is thereby excluded. In other words, the molecule surface that delimits up to where a solvent particle can penetrate[196,197]. These terms can be explicitly shown in the following figure:

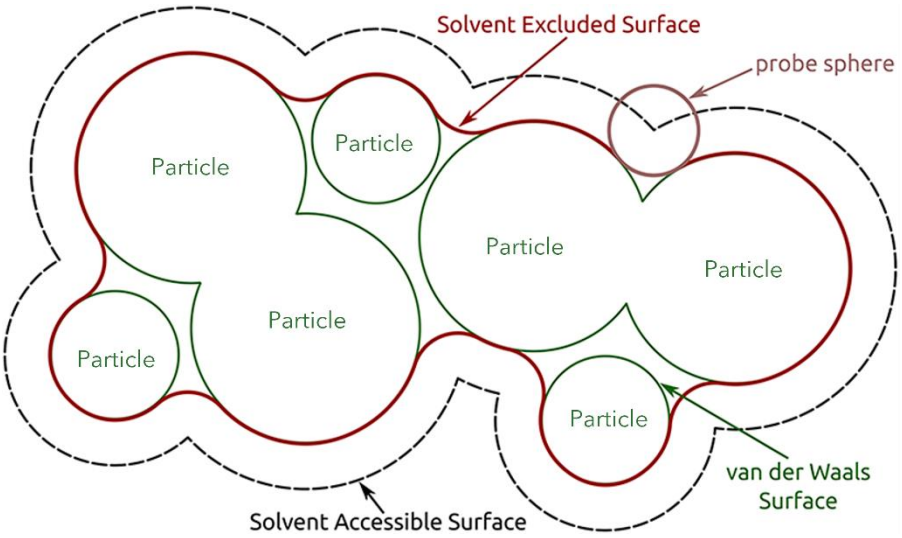


Figure 4.11 – Illustration that shows SAS (Solvent Accessible Surface), SES and Van der Waals surface definitions visually. The probe sphere can be understood as another particle interacting with this arbitrary molecule constituted by all the particles seen (for instance, the solvent). Adapted from Daberdaku[198].

Beads placement, as will be seen in the next Chapter, was done until we arrived at a satisfactory matching between the AA and CG SASAs. Yet, we can have a good CG-to-AA SASA ratio but a mismatch of the Connolly surfaces. They have to match as well, so that the region itself that interacts with the solvent beads is well recovered. This way, with the visualisation of Connolly surfaces, we rearranged some beads' positions not only to get a good CG-to-AA SASA ratio, but also to get a match between the CG and AA Connolly surfaces. Connolly surfaces were checked via VMD[163]. Their matching was accomplished by adjusting the weights or contribution of each atom to the bead placement.

For this thesis, SASA and Connolly surfaces were determined using the GROMACS[140] tool *gmx sasa*. A probe with 0.191 nm and 2000 dots was used for the calculation of SASAs, whilst only 150 dots were considered for Connolly surface pictures throughout this thesis. For Martini molecules, the Van der Waals radii that were employed were 0.264, 0.230 and 0.191 nm for R, S and T beads[137], respectively. For UA references, the Van der Waals radii were obtained from Rowland and Taylor's work[199] except for phosphates which were retrieved from Bondi's work[200]. Visualisation and rendering were done with VMD[163].

5 MODEL DEVELOPMENT

In this Chapter, it will be addressed how we created our models and how we refined them. It will focus on the parameterisation of several molecules, which will be part of a future article yet to be submitted.

5.1 Small molecules: quinones and nucleotides

Nucleotides and quinones were parameterised using the Martini 3 CG force field[137]. All the compounds we parameterised followed a building-block approach - a specification of Martini - whenever possible, meaning that the same moieties had the same bonded and nonbonded parameters. In parallel to our old paper[70], nucleotides were parameterised according to a top-down approach, where parameters used for larger molecules were used for their CG counterparts. On the other hand, quinones were parameterised following a bottom-up approach. Here the smaller moieties are parameterised beforehand. Then, those parameters are added to build the larger molecules' topologies.

Validation of our bonded parameters was targeted to get a Solvent Accessible Surface Area (SASA) CG-to-AA ratio smaller than 5%, whereas our nonbonded parameters were validated by calculating the octanol-water partition coefficient (log P) and compare with experimental values. Final validation involved simulations in biochemically relevant settings — FAD as the cofactor in an oxidoreductase and quinones in a membrane.

The result of our work is a set of parameters for 35 molecules (**Table S1**, in Supplementary Information) that are central to many biochemical interactions. Their availability within the state-of-the-art Martini 3 framework will enable a wide range of new applications in the field.

5.1.1 General parameterisation approach

Model parameters were developed, in line with the Martini philosophy, to i) capture the conformational space visited by each molecule and ii) reproduce emergent properties that reflect correct intermolecular interaction preferences. The first aspect was modelled with CG bonded parameters (bonds, angles, torsions, ...) that restrain the configuration space available to the CG molecules. These parameters were adjusted in a trial-and-error fashion to reproduce behaviour obtained by reference simulations at a finer (atomistic) resolution. The relatively rigid moieties involving resonant rings (adenine, flavin, nicotinamide, thiamine and quinone head groups) were further parameterised as rigid bodies using a combination of linear constraints and virtual sites, as exemplified in the development of the latest version of Martini 2 cholesterol[201].

The preference for intermolecular interactions was addressed via the chosen Martini particle types for a given molecule, which dictate the nonbonded Lennard-Jones particle-particle interaction potentials. Although Martini 3 provides a total of 843 possible particle types, a set of rules guides the type attribution process[137]. Suitable types are chosen according to the chemistry (polarity, charge) of the moieties being modelled and to their size (*i.e.*, how many non-hydrogen atoms constitute each bead). The growing body of Martini 3 models[133,149] provides examples that further assist in bead type choice, in that new molecules being parameterised should follow the bead types of existing, chemically similar moieties. Finally, particle choice was validated in a more consequential setting by comparing simulated and experimental octanol-water partition coefficients[70,137,202].

5.1.2 Composition strategies: bottom-up and top-down

Nucleotides were parameterised by a top-down approach in which reference atomistic simulations were used only for the largest nucleotides of each type. CG parameters for those nucleotides were then developed against these references. Subsequently, models for smaller nucleotides and constituent moieties (flavin, riboflavin, adenosine, nicotinamide and thiamine) were obtained by fragmenting and adapting the parameters of the larger CG models[203]. Conversely, quinones were parameterised by a bottom-up approach, which is common in the development of Martini lipids[202]. In this case, headgroups and tails were parameterised independently and then combined following the characterisation of their linked properties[203].

A sum of both these approaches can be shown as follows:

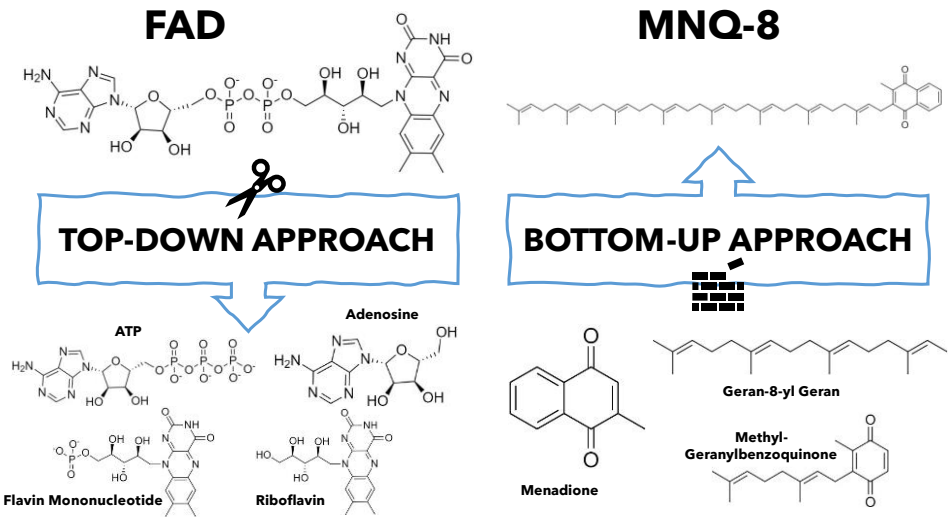


Figure 5.1 – Illustration that shows the nucleotides top-down approach, where parameters of larger molecules can be reproduced in smaller molecules according to their similar moieties (left) and the quinones bottom-up approach, where parameters of smaller fragments are added together to build a topology for larger molecules (right). Structures drawn using ChemDraw[40].

Figure 5.1 summarises both approaches to parameterise all molecules included in this thesis. Now, the choice of the beads placement will be addressed.

5.1.3 Beads placement in general

One particular parameterisation aspect introduced by Martini 3 is the matching of a molecule solvent-excluded volume to avoid inaccurate interactions due to over- or under-mapping situations (where CG bead density is either too high or too low, respectively). To this end, the solvent accessible surface area (SASA) of the resulting model — employing Martini 3-specific Van der Waals radii[137,149,157] — can be compared as a parameterisation metric to that of the atomistic reference. Shape matching can be further evaluated by comparison of Connolly surfaces[196]. As a rule of thumb, a centre-of-geometry mapping — where a bead is placed at the non-mass-weighted centre of constituting atoms, including hydrogens — yields better results in these regards than the traditional Martini 2 non-hydrogen centre-of-mass method[137,149]. Still, the simplistic centre-of-geometry approach often fails to preserve the shape and size of the molecular excluded volume, prompting an additional fine-tuning of CG bead particle positions. In order for such a fine-tuned mapping to be consistently applicable to different atomistic configurations (such as the different structures a nucleotide adopts along a reference atomistic simulation), we define each beads position as the centre-of-weights (COW) of underlying atoms, where

the tuning process adjusts the constituting atoms' relative weights. This process was used for all sets of nucleotides, apart from the thiamine set, in which, alongside with quinones, a simple centre-of-geometry (COG) process was employed.

5.1.4 Nucleotides: a top-down approach

We revisited the set of adenine-, nicotinamide-, flavin- and thiamine-based nucleotides that we recently parameterised for Martini 2[70], now within the scope of Martini 3[137]. We followed here a top-down approach, where FAD, ATP, NADH, NAD^+, NADPH, $NADP^+$ and Thiamine Pyrophosphate (TPP) were used as the starting points for parameterising all the other smaller nucleotide derivatives (**Table S1**). From FAD and ATP, we could parameterise adenosine (ADOS), adenosine monophosphate (AMP), adenosine diphosphate (ADP), flavin mononucleotide (FMN), riboflavin (RBFL) and $FADH_2$ (FAD reduced form). By using NADPH, NADH, NAD^+ and $NADP^+$, we also adjusted the FAD parameters (so that similar moieties would have the same parameters) and the counterpart moieties. TPP allowed parameterising Thiamine Monophosphate (TMP) and Thiamine (THI). The different moieties are shown in the next figure:

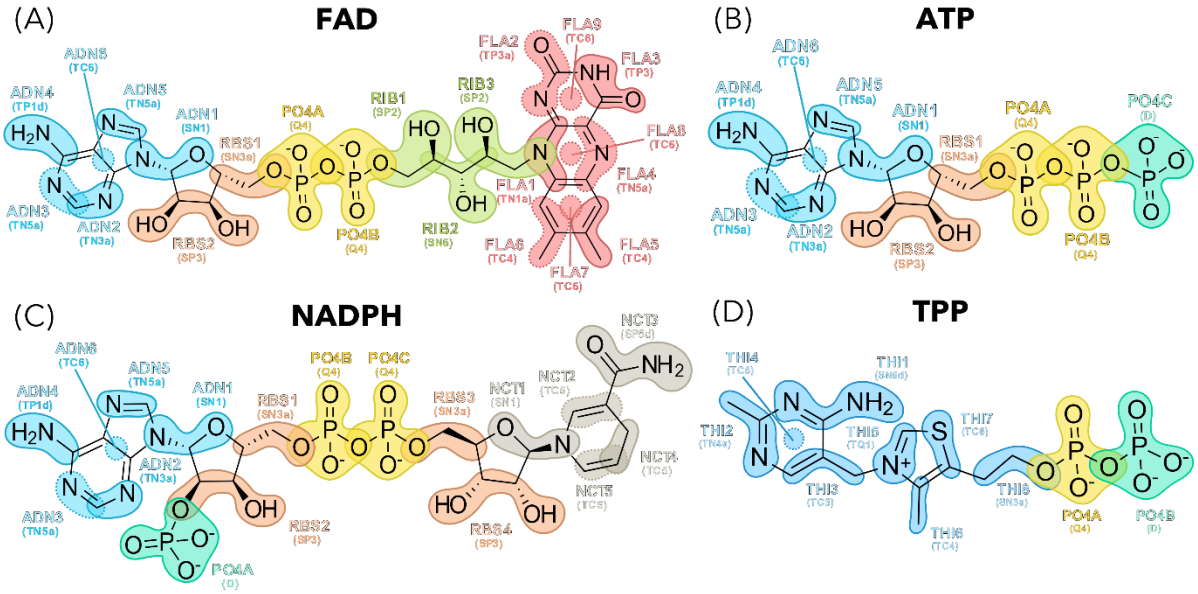


Figure 5.2 – Illustration that shows all the nucleotide moieties that were parameterised in different molecules and their representative compounds: (A) FAD (flavin set), (B) ATP (adenine set), (C) NADPH (nicotinamide set) and (D) TPP (thiamine set). Moieties with the same colour have the exact same parameters (bonded and nonbonded) in different molecules following the Martini building-block approach. The different moieties found in nucleotides are split into: ADN (adenine moiety, in light blue), RBS (ribose moiety, in orange), PO_4 (phosphate moiety, in turquoise or yellow depending on its protonation state), RIB (ribitol moiety, in green), FLA (flavin moiety, in red), NCT (nicotinamide moiety, in grey) and THI (thiamine moiety, in dark blue). Virtual bead sites are represented with dashed lines. Below each bead name, the bead type is displayed (in parenthesis). Structures were drawn using ChemDraw[40], and beads were depicted using Inkscape[41].

Also, these molecules had some virtual sites added, *i.e.*, some beads were placed relative to others using constraints. This approach allows the maintenance of the molecules symmetry avoiding excessive constraints or too stiff bonds[201]. The same reference GROMOS 54A7 united-atom[184] trajectories were used, with added hydrogens to allow for proper excluded volume evaluation and matching.

With the release of Martini 3, parameters were made available for the adenine, cytosine, guanine, thymine, and uracil nucleobases, as well as ribose[137]. In parameterising adenosine — the ribonucleoside formed by adenine bound to ribose that is a core component of most of the nucleotides we address here — we strove to retain as much compatibility as possible with those existing parameters for overall Martini

3 consistency and interoperability. Some differences, detailed below, were nonetheless introduced in order to better reproduce the atomistic behaviour.

5.1.4.1 Beads placement

The Supplementary Material in Google Drive includes all the molecule index files in GROMACS format, that detail each bead constructing atoms and their respective weights.

After discussion with P. C. T. Souza, a mapping was implemented for adenine that, when applied to our atomistic reference structures, mostly reproduced the intramolecular CG bead distances in the released Martini 3 topology[137] (differences, in the order of tenths of Å, could be attributed to atomistic model mismatch). Ribose bead placement followed the same aim, though with less fidelity since the model released with Martini 3 has three beads only and the conjugations required to form adenosine, NADH/NADPH/NAD^+/$NADP^+$ or FAD already change some of those bead placements.

In adenosine, a 7 C: 3 N ratio was given to the ADN1 bead, where the remaining were placed directly on each nitrogen, except for the middle bead (ADN6), which was set in the middle of the ADN2 and ADN5 beads. In ribose, only the oxygens were considered for the placement of bead RBS2.

Additionally, PO_4 beads were placed right on the phosphate, and ribitol beads were placed asymmetrically since doing it symmetrically would result in a large CG-to-AA SASA ratio.

Nitrogen beads were placed right on the nitrogen, whereas amide groups were placed taking into account two oxygens in order to avoid a poor spatial occupation. The methyl beads were placed on the carbon methyl, and the middle beads considered all the atoms in each ring.

In NADH, NAD^+, NADPH, and $NADP^+$, if we set beads on each nitrogen, it would result in an unsatisfactory CG-AA SASA match. Therefore, the NCT1 was placed like ADN1 in FAD and NCT2 was given an 11:11:7:11 (C:C:C:H) ratio. The amide group (NCT3) only took into consideration one oxygen and one hydrogen connected to the nitrogen. For NCT4 and NCT5, a 2:1:1 (C:H:H) and a 1:1:1 (C:C:H) ratios were contemplated. Again, it is important to recall that a COW mapping approach was employed, and sometimes, we assigned arbitrary ratios so that they would satisfy the spatial occupation.

5.1.4.2 Bonded parameters

The choice of our bonded parameters was a trial and error process where the main goal was to reproduce the molecules behaviour from the UA simulations from Sousa *et al.*[70]. The fits were checked by transforming the atomistic simulations into a CG-like trajectory and by getting afterwards the respective distributions. We started this approach in the larger molecules such as FAD, ATP $NADP^+$, NADPH, NAD^+ and NADH. Then, the resulting parameters were used for smaller molecules, such as riboflavin, FMN, adenosine, AMP and ADP, in a top-down approach, since these can be derived from the splitting of the larger molecules.

In adenine, we used the bonded parameters obtained from Martini 3 ITP file from Souza *et al.*[137]. The only adjustments made were regarding the connection of this nucleobase to other moieties, such as ribose. A mismatch between AA and CG distributions was found, probably due to a greater freedom in our CG models. The adenine was initially made for Martini 3 CG simulations of Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA), where it has much less freedom.

The bond of connection was a target of a thorough analysis since its parameters were also tried to fit in nicotinamide derivatives. Thus, simulation refinements were done so that it would fit not only to the adenine-ribose link (ADN1-RBS1) but also to the ribose-nicotinamide connection (RBS3-NCT1):

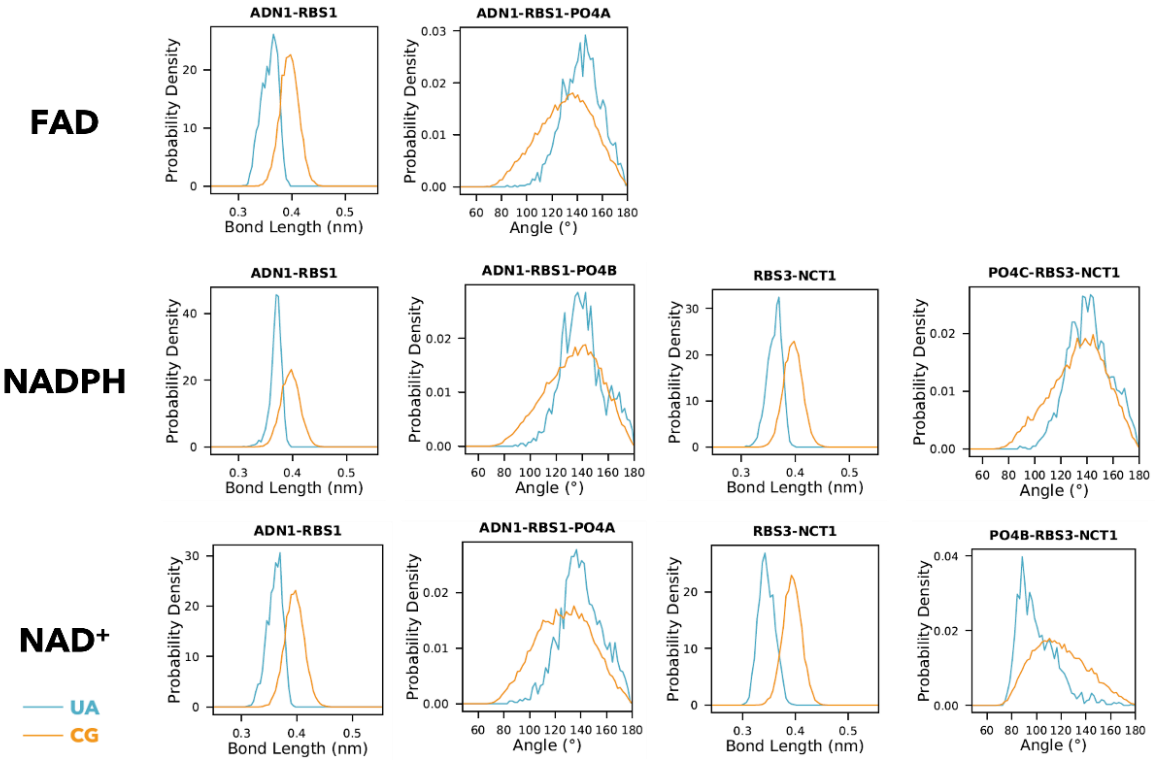


Figure 5.3 – Illustration that shows the plot distributions of distances and angles for ADN1-RBS1 and ADN1-RBS1-PO4A/B (depending on whether there is an extra phosphate group in the molecule, see **Figure 5.2**), RBS3-NCT1 and PO4B/C-RBS3-NCT1 for FAD, NADPH and NAD^+. Distributions in blue are UA behaviour, whereas CG is in orange. Plots were done using a script written by my colleague Gonçalo Vieira.

If we look at **Figure 5.3**, we can easily spot that the CG distance for the adenine-ribose (ADN1-RBS1) bead is a bit longer than UA. In fact, they are a bit longer than the mapped values, but this was done so that the published Martini 3[137] distances would be kept. In spite of the different UA behaviour for FAD, NADPH and NAD^+ when it comes to the ADN1-RBS1 bond and ADN1-RBS1-PO4A/B, the exact CG parameters were used, and they can acceptably mimic the UA behaviour in all three cases. The ADN1-RBS1 bond parameters were the same as RBS3-NCT1, and they reasonably mimic well the UA distributions. The only change made regarded the oxidised and reduced forms of nicotinamide derivatives, which explains the different CG behaviour in the last column of plots. NAD^+ and $NADP^+$ had the same angle (and dihedral) parameters, but they differ from NADPH and NADH, possibly due to the oxidation state being different and, thus, possibly dissimilar conformations and flexibility, which require distinct parameters.

Ribose constraints in Martini 3 topology[137] were changed to bonds since the behaviour in our larger molecules showed much more freedom for ribose. This makes sense since the previous topology was made for the ribose present in DNA or RNA, where the ring is kept still by binding to nucleobases and phosphates. For instance, in DNA or RNA, the RBS2 bead is usually phosphorylated, which does not happen in our molecules, thus contributing to a possible greater stretch of the hydroxyl groups. We tried to keep as much of the original bonded parameters as possible, but some angles and dihedrals were added, and constraints were swapped into bonds.

In nicotinamide and thiamine, virtual sites were added to mimic the distances we can find in these rings accurately. These distances were, most of the time, constraints since they are rigid moieties of a molecule, meaning that these bonds do not oscillate much, so constraints could easily mimic UA behaviour. Furthermore, it is of the utmost importance to call the reader's attention to the fact that the ribose linked to the nicotinamide is not equivalent to the one that is connected to the adenine. Although the same parameters were used for both to follow the building-block approach, not all the angles, distances, and dihedrals were entirely spot-on. For instance, some of the constraints in the nicotinamide distributions are a bit off. This might be explained by the fact that we had to find parameters that would

fit both moieties, perhaps because virtual sites were made using a single frame and maybe also by the fact that we have forced the amide to be symmetrical, which is not represented in our mapping. Besides, there is a slight wobble in the UA model of the nicotinamide ring itself that may cause some deviations in its distributions.

Whenever a system crashed, we would inspect which distances could be below LJ's σ value. In these situations, these potentials had to be turned off, and after seeing the intermediate distribution plots (and checking which distance the bond should have), we could turn it on again but excluding the particles from each other, meaning that they would not have any nonbonded interactions. This hampered our work in trying to represent the molecule UA behaviour. Other times, we had problems with particles becoming collinear and the dihedral becoming undefined. This happened, for instance, with particles RBS4-NCT1-NCT2. Here we had to introduce a ReB angle to avoid collinearity. The reasoning behind these choices was thoroughly explained in Chapter 2.

5.1.4.3 Nonbonded parameters and mapping

The choice of bead types, some already shown in **Figure 5.2**, was based on Riccardo Alessandri's building block table[149], which was kept consistent among all molecules. This table covers the correspondence between a certain chemical group to specific bead types in such a way that the same chemical groups have got the same bead types if they are placed in the same environment (for instance, in rings). Again, and according to the approach we followed, the bead types in larger nucleotides were the same as the ones used in the smaller moieties, in general. After parameterisation, the nucleotides' bead types were validated by calculating the octanol-water partition coefficients (log Ps) as was done in Sousa *et al*.'s paper[70]. According to these results, bead types could be readjusted, if necessary, to better match the experimental log P values.

To clarify the choices made and ease the understanding of the underlying rationale, the next figure was elaborated with the rest of the nucleotides divided into their sets:

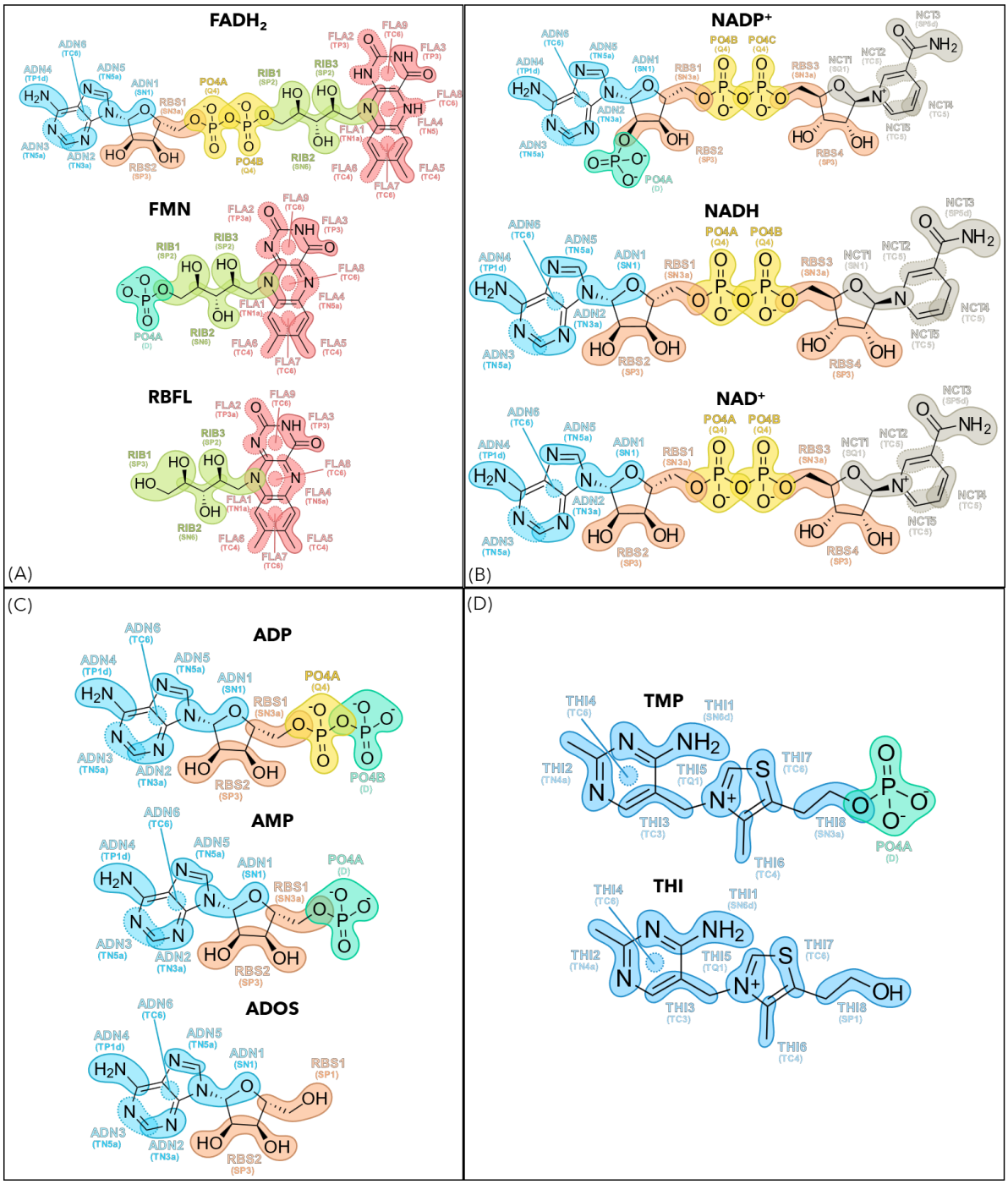


Figure 5.4 – Illustration of the rest of the molecules in each set: (A) flavin, (B) nicotinamide, (C) adenine and (D) thiamine, which are shown in **Figure 5.2**. In (A) $FADH_2$, FMN and RBFL are shown. In (B), we can see $NADP^+$, NADH and NAD^+. In (C) were drawn ADP, AMP and ADOS. In (D), TMP and THI are displayed. Caption details are the same as in **Figure 5.2**.

The bead choices used for adenine and ribose were already made from the Martini 3 models for DNA and RNA[137]. The only beads that were changed were those that linked different moieties (*i.e.*, the connections between ribose and other moieties). We maintained the diol bead type (RBS2 - SP3), and we named the RBS1 bead type SN3a when linked to another molecule. It is an SP1 when the hydroxyl is free, in the case of adenosine. The adenosine part was kept the same for all the molecules where we can find it (NAD^+, NADH, NADPH, $NADP^+$, ATP, ADP, AMP and FAD).

In all molecules except for the ones containing only one PO_4 moiety, the bead type chosen to represent an individual phosphate is Q4, differing from the typical Martini 3[137] parameterisation (Q5).

This change arose from the need to better represent the slightly less polar behaviour of each PO_4 group, originated from the shared anionic character of the P-O-P bonds. This change was motivated by the fact that we were seeing exceptionally hydrophilic values in molecules that were phosphorylated, whereas for molecules without phosphate groups, that behaviour was not being seen. Divalent phosphate beads were defined as D following typical Martini 3 parameterisation.

Then, the ribitol part in FAD was parameterised as being itself alcohol (SP2), where the bead in the middle (RIB2) is slightly more hydrophobic (SN6) since it only involves one polar atom (oxygen). The first bead (RIB1), when it is free, is more hydrophilic, being an SP3, like in riboflavin.

The beads used for the flavin ring are similar to the ones used for adenine, being the beads in the middle of the rings labelled as TC6. The methyl groups in rings are TC4, according to the building blocks table available[149]. In the ALO1 bead (with the first nitrogen), the N is placed in a ring, being less polar. The amide ring groups were designated as TP3, being ALO2 a hydrogen bond acceptor (TP3a), and ALO4 the same as ADN3 (TN5a). All beads are T since they are placed in ring structures representing 3-to-1 mappings[137].

In nicotinamide, all beads were denominated as TC5, except for the amide outside of the ring, which was set as SP5d and the bead that connects the ribose to the nicotinamide, which depends on the oxidation state of the molecule. Regarding $NADP^+$ and NAD^+, the NCT1 bead is an SQ1 (since it is charged), whereas for the reduced molecules is an SN1, like in adenosine. The TC5 bead type comes from being a bead that is inside a ring (representing 2/3-to-1 mappings) and that mimics resonance.

Finally, in TPP, the primary amine in the ring was labelled as an SN6d, while THI2 was labelled as TN4a. THI3 is quite similar to a methyl group in a ring, but here this bead is placed between two rings ending up being more hydrophobic (TC3). The THI4 is a bead inside the ring, so it is a TC6. THI5 bead is a TQ1 to consider it being charged. THI6 is a similar case to the methyl groups that are ring-bound (TC4). THI7 is more hydrophilic because of the sulphur atom and similar to the SC6 situation in the building block table[149], so we named it TC6 because it mimics the same chemical behaviour as SC6 and it is smaller. THI8, THI9 and THI10 are equal to NADPH RBS1/RBS3 (SN3a), PO4B/PO4C (Q4) and PO4A (D), respectively.

5.1.5 Quinones: a bottom-up approach

We used a bottom-up approach to parameterise quinones, meaning that we performed united-atom simulations for smaller molecules to build the CG topologies, and then they were all added to obtain a single CG topology for larger molecules. The starting points were GBZ and G8YLG, and ubiquinone was relevant for representing methoxy behaviour and adjusting the final quinone parameters.

In Chapter 1, we have acknowledged menaquinone as the substrate of the enzyme we are studying, NDH-2. Yet, in experimental procedures, menaquinone is not soluble in water, thus, an analogue is used: 2,3-Dimethyl-1,4-naphthoquinone (DMNFQ), also parameterised. Also, for the selectivity analysis, plastoquinone and ubiquinone were considered to assess whether NDH-2 showed selectivity for its natural quinone substrate. Additionally, NDH-2 inhibitor was also parameterised. Moreover, and since another target of this thesis was to study the selectivity of NDH-2 for quinones over quinols, the latter were also parameterised. This way, many quinone/quinol molecules and derivatives were also parameterised (**Table S1**).

5.1.5.1 Beads placement

All atoms that constitute a bead were contemplated in the NDX file that would generate the CG GRO file. In other words, each bead was placed taking into account all the atoms of which it is composed and not using a COW approach, as is done in nucleotides.

5.1.5.2 Bonded parameters

As previously mentioned in the Methods Chapter, the geran-8-yl geran (a tetraene-like molecule) and the methyl-geranylbenzoquinone (represented in **Figure 5.5**) were employed as the starting points to get the chain (tail) and the ring-chain connection (head-tail) behaviour, respectively. Besides, for refinements of this parameterisation and to simulate the methoxys behaviour, we used the ubiquinone topology from de Jong *et al.*[152]. As for HQNO, we retrieved the chain parameters from Martini 3 octane topology. The UA and CG behaviours for the fragment (GBZ) are depicted in the following figure:

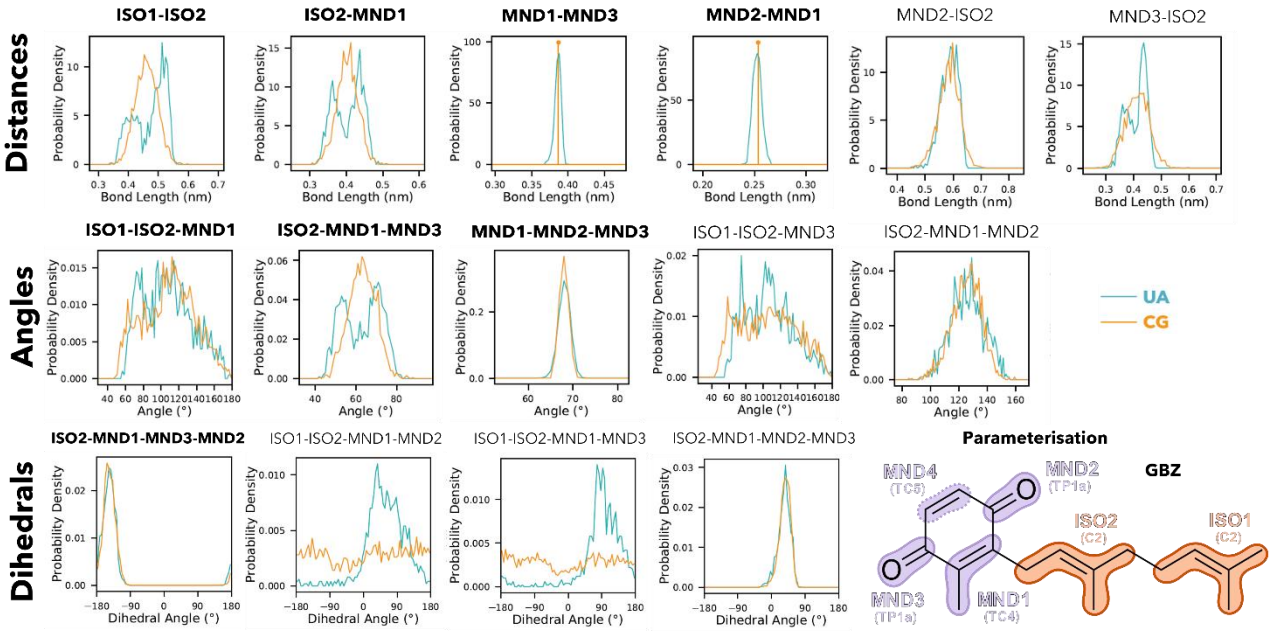


Figure 5.5 – Illustration that shows GBZ UA and CG behaviours for different distances, angles and dihedrals. Distributions in blue are UA behaviour, whereas CG is in orange. The structure with the different moieties and bead names is represented in the bottom right. In bold are potentials that are defined by us. Structure was drawn using ChemDraw[40] and beads in Inkscape[41]. In light orange, we have the ISO (isoprenoid) moiety, and in violet, the MND (menadione, menaquinone head) moiety. Virtual sites are depicted in dashed lines and bead types in smaller capitals in parentheses. Finally, lines with a final ball at the end mean those bonds are actually constraints. Plots were done using a script written by my colleague Gonçalo Vieira.

A great endeavour to try to represent best the UA behaviour was made, as seen in **Figure 5.5**. Some behaviours, such as ISO1-ISO2 (chain behaviour), ISO2-MND1 (tail-ring behaviour), and ISO2-MND1-MND3, were tough to recover, so an average behaviour was considered. Yet, the potentials we did not force and the imposed dihedral (ISO2-MND1-MND3-MND2), were pretty well established. These bonded parameters were also adjusted according to their SASA values.

With good accuracy, we were able to recover most of the UA behaviour for the chain and tail-ring bond.

Simulations were often crashing since the distance between ISO2 and MND3 was often getting below the LJ's σ value, thus an exclusion was included between these two beads.

5.1.5.3 Nonbonded parameters and mapping

Similarly to nucleotides, quinones bead types were chosen according to Riccardo's building block table[137], targeting the best possible log P values match between experimental and calculated values. Please recall that all atoms shown in each bead were taken into account for its placement. Possible re-adjustments were made regarding the log P value calculated. Here, in a bottom-up approach, the bead

types we used to characterise the nonbonded interactions in smaller molecules were the same as those used for the larger molecules. Our bead type choices are displayed in the next figure:

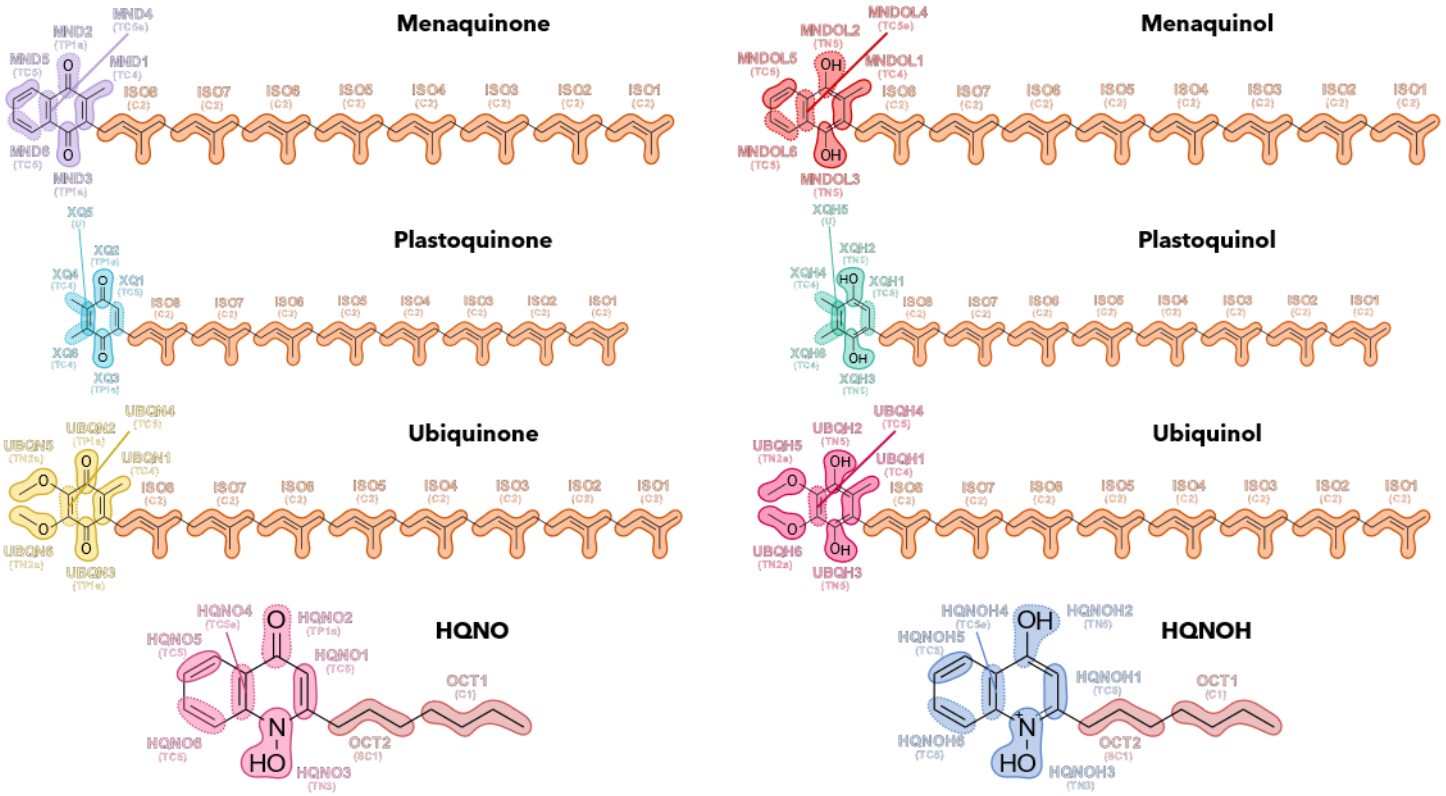


Figure 5.6 – Illustration that shows our bead type choices for quinones, quinols and NDH-2 inhibitor. Moieties with the same colour have the exact same parameters (bonded and nonbonded) in different molecules following Martini building-block approach. The different moieties found in quinones are split into: MND (menadione moiety, in violet), ISO (isoprenoid moiety, in orange), MNDOL (menadiol moiety, in red), XQ (xyloquinone moiety, in light blue), XQH (xylene diol moiety, in turquoise), UBQN (ubiquinone moiety, in yellow), UBQH (ubiquinol moiety in dark pink), HQNO (quinolinol moiety in light pink), OCT (octane moiety in dark red) and HQNOH (HQNO alcohol moiety, in dark blue). Virtual bead sites are represented with dashed lines. Below each bead name, the bead type is displayed (in parenthesis). Structures were drawn using ChemDraw[40] and beads were depicted using Inkscape[41].

As for nicotinamide, benzene-like ring beads were also tagged as TC5. The TC5e beads (for instance, in menaquinone) are beads which are placed between two ring structures, being electron richer.

The quinone beads chains which present a conjugated system, were labelled as C2 to overcome an issue we saw in membranes, later addressed. Quinolinol chain beads, on the other hand, were named C1 (like in octane chain) since they are simple bonds.

Then, carbonyl groups were labelled as TP1a. In contrast, the hydroxyl groups connected to the ring were named TN5 since the C=O bonds in quinones are more hydrophilic due to a strong dipole that is created between the negative partial charge in the oxygen and the positive partial charge in the carbon, which does not happen in quinols.

Like for the FAD methyl groups, here methyl beads were denominated as TC4 (for instance, PQ12 and PQ14). In plastoquinone, a U bead was added. This is a dummy bead that does not have any nonbonded interactions with the other beads. It was added to keep the molecule completely symmetrical forcing XQ2 and XQ3, the carbonyl beads, to be at the same distance from XQ1 and XQ5.

Lastly, the methoxy groups were labelled as TN2a as suggested in the building block table, and the ketone substituent in HQNO (HQNO5) was labelled as TN3 since it is a mix of a tertiary amine and a NO_2 group.

5.1.6 Solvent Accessible Surface Area (SASA) calculation – bead sizes

Almost all the CG SASA compound values (**Table S2**) were within a 95-105 % range of AA SASA since 5 % is the maximum difference recommended (green region in **Figure 5.7**)[137]. Smaller molecule/moiety SASAs are obtained from the larger molecules' trajectories but only by taking into account the atoms that are part of that moiety. Since we have an AA simulation trajectory we can compare to, we were able to obtain a confidence interval. Conversely, for larger quinones, that interval is not available since their construction followed a bottom-up approach and there is no single AA trajectory of the full molecule. The results from the four main molecules that were used as starting points for the nucleotides top-down approach can be displayed as follows:

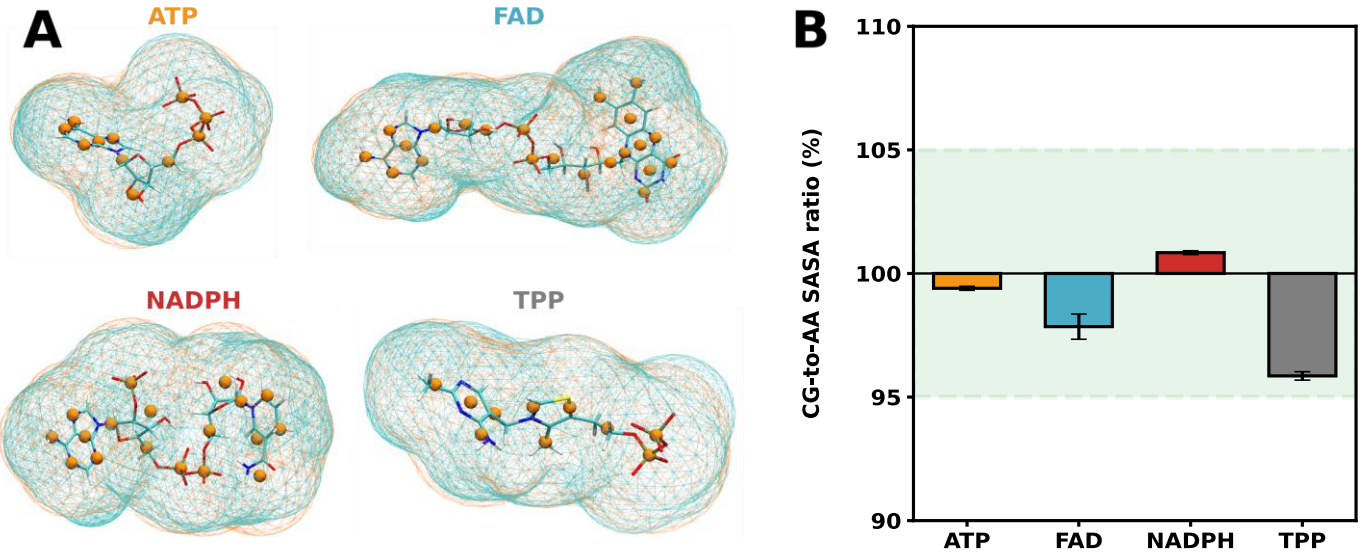


Figure 5.7 – Illustration shows (A) Connolly surfaces of ATP, FAD, NADPH and TPP, where atomistic structures and Coarse Gain beads are shown. In cyan are represented the carbons, in dark blue the nitrogens, in white hydrogens, in red oxygens, in brown phosphates and in yellow sulfurs. The orange spheres are the CG beads. CG Connolly is in light orange, and UA Connolly is in blue. Rendered using VMD[163] using only 150 dots. (B) Coloured bars represent the CG-to-AA ratios of ATP, FAD, NADPH and TPP, and the error bars represent how CG and AA structures along their trajectories adapt multiple and different conformations. Confidence intervals of 5 % were calculated using the *bootstrap* tool[204], and propagation error was calculated using the *uncertainties* python package[205] (all values are listed in **Table S2**). This plot script can be found in my GitHub[161] in *CGtoAARatioPlot.ipynb*.

In **Figure 5.7** are displayed the four main molecules of each set (adenine, flavin, nicotinamide and thiamine) and in all of them the CG-to-AA SASA ratio is below 5 % and Connolly surfaces matched. FAD is shown as having the larger error bar, because it adopts several different conformations in its trajectory.

Between all molecules (**Table S2**), ribitol was slightly above the 5% maximum (7.6 %). This is easily explained by the fact that this is a small molecule (3 beads) whose SASA values are being extrapolated from larger molecules. The CG mapping was improved as much as possible to mitigate the discrepancy between CG and AA SASAs, in such a way that beads in ribitol were placed in an unsymmetrical manner.

G8YLG and plastoquinone were also above the 5% target range (8.3 % and 5.9 %, respectively). However, little or nothing could be improved since we already used R beads for the tails and the best mapping possible to mitigate the large ratio.

Note that geran-8-yl geran and methyl-geranylbenzoquinone AA trajectories ran for 50 ns, and UBQN AA and CG trajectories ran for 100 ns. According to **Table S2**, Martini 3 successfully retrieved

the several states these molecules can adopt in a trajectory since the SASA dispersion for each molecule is of the same order of magnitude.

5.1.7 Octanol-water partition coefficients– bead types

Getting the octanol-water partition coefficients is a way of monitoring whether the bead type choices were adequate. These values were obtained through an Alchemical Analysis, explained in the Methods Chapter.

Overall, the results are good for non-charged species such as in riboflavin, RBFL. Adenosine and ribose are a bit off. Some of their parameters were not chosen by us (some were obtained from the DNA and RNA Martini 3 models). Furthermore, the log P values for phosphorylated compounds are off when compared to their experimental values. Even after an adjustment made to PO_4 bead type, from Q5 to Q4, by differentiating paired beads from individual monoprotonaded ones, these log Ps were still overly hydrophilic:

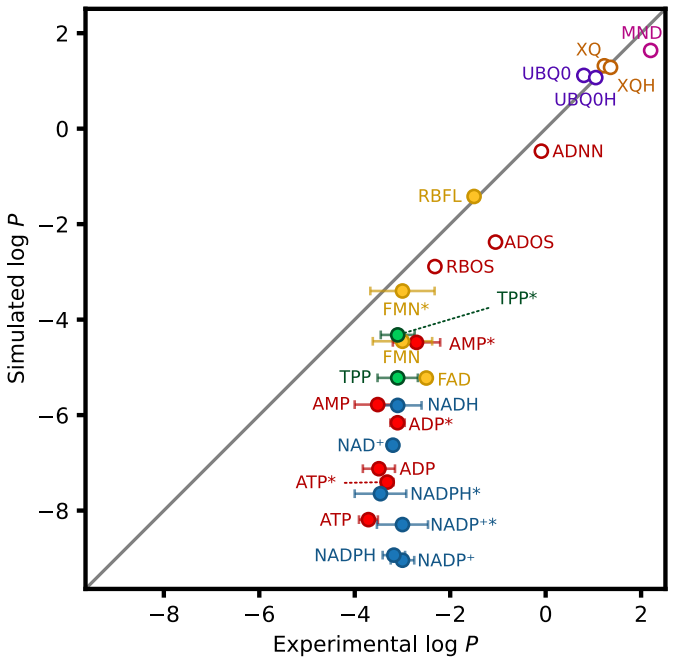


Figure 5.8 – Illustration that shows the calculated log P values as a function of the experimental results for several molecules. Each set has a different colour. Nicotinamide set is in blue, adenine set is in red, thiamine set is in green, flavin set is in yellow, ubiquinone set in purple, plastoquinone set in brown and menaquinone set in dark pink. Y = x function is drawn to guide the eye. Asterisks indicate protonated species. Compounds with experimentally measured log P from Hansch[206], Sangster[207] and Rich & Harper[208] are denoted as empty circles, and compounds whose measurement was done by Sousa *et al*.[70] are shown as filled circles. All log P values are shown in **Table S3**.

From **Figure 5.8**, we can see that the simulated phosphorylated compounds are very hydrophilic (log P more negative), showing that some refinements, namely regarding phosphate beads, need to be done. In the future, a way of bridging this issue would be the measurement of the second virial coefficient (B_{22}), which is experimentally measurable and can give us information related to the interaction between a solute and a solvent[133,209].

All calculated and experimental log P values are included in the Supporting Information in **Table S3**.

When we were doing the quinones and quinols parameterisation, we noticed that some quinols were getting a simulated log P much more hydrophilic than their experimental value. We were using a TN6 bead type, as suggested in Riccardo Alessandri's table[149], but we changed it to a TN5 to better recover quinols experimentally measured polarity. Besides, we were seeing that the corresponding quinones were, in the opposite way, yielding overly hydrophobic values using a TN6a bead type. We changed to a TP1a bead type, which more accurately would reproduce the experimental results. We believe this experimental hydrophilicity of quinones over quinols is due to a resonance stabilisation on quinone rings, which creates a strong dipole (positive partial charge on the carbon and negative partial charge in the oxygen), that is not found in quinol molecules. We extended this study to other alcohols to check whether this overly hydrophilicity problem was exclusive or not to quinol molecules and we built the following table:

Table 5.1 - Comparison of the variation of Gibbs free energy (in kJ.mol^{-1}) for methanol, ethanol, propanol and isopropanol for tabulated and experimental octanol-water (OCT-W) equilibria and hexadecane-water (HD-W) equilibria and which bead type would better represent the experimental value at a temperature of 298 K (in small capital letters).

Alcohol	Martini 3 OCT-W ΔG[a]	Experimental OCT-W ΔG[b]	Martini 3 HD-W ΔG[a]	Experimental HD-W ΔG[c]
Methanol*	-7.2 | -6.1 TP1 | TN6	-4.4 TN5	-13.7 | -11.9 TP1 | TN6	-16.1 TP2 | TP3
Ethanol	-5.2 SP1	-1.8 SN3	-13.1 SP1	-12.5 SP1
Propanol	-1.1 N6	1.4 N3	-9.8 N6	-8.3 N5
Isopropanol	-2.0 P1	0.3 N5	-11.1 P1	-9.5 N6

*Methanol is an SP2r in Martini 3 topology[a], which represents two molecules. We would consider it as a TP1 or TN6 if it was a single particle.

[a]Souza *et al.*[137].

[b]Hansch[206].

[c]Abraham *et al.*[210].

From **Table 5.1**, we can see that the bead types we are using for simpler alcohols in OCT-W equilibria, that are tabulated in Riccardo Alessandri's building-block table[149], are too hydrophilic as well when compared to experimental values (Gibbs free energy is much lower, thus the beads are more polar). Thus, this issue is not exclusive of quinols, but happens for all alcohols. This issue is also seen in HD-W equilibria, but it is more evident in OCT-W equilibria.

We can see that, for instance, ethanol is tabulated as an SP1 bead type for Martini 3, but by checking its experimental octanol-water Gibbs free energy, we know that the corresponding bead type should be an SN3. Yet, an SP1 parameterisation is best for the HD-W equilibrium. Therefore, this issue is not exclusive to quinols but also to other simpler alcohols. These issues could be mitigated by using more bead types specific to each chemical group.

5.1.8 Nicotinamide derivatives

As an improvement over our last paper with Martini 2[70], where nicotinamide derivatives (NADH, NAD^+, $NADP^+$ and NADPH) needed an alternative topology to represent the bent state (different from the one that would represent the stretched state), this time, we were able to get both states – stretched and bent (**Figure 5.9**-**A** and **Figure 5.9**-**B**, respectively) - using a single topology file. This pitfall of the previous Martini was now overcome with the most recent Martini 3, and it is a win not only for our group but for the Martini 3 community[137].

We were able to observe a bimodality if we look at the distances (**Figure 5.9**-**C**) between the first phosphate bead and the last bead of the ribose near the nicotinamide, where this bond adopts, without any enforcement, a distance of 0.5 nm fewer times (bent state) and a distance of 0.8 nm most of the time (stretched state). We initially had a pair, a bond where its beads are excluded from each other, between the first internal phosphate and the first nicotinamide bead (**Figure 5.9**-**D**), but we saw that without it, nicotinamide derivatives were still able to adopt both "closed" and "open" conformations, meaning that we did not need to force our molecules to assume any conformation, as seen next:

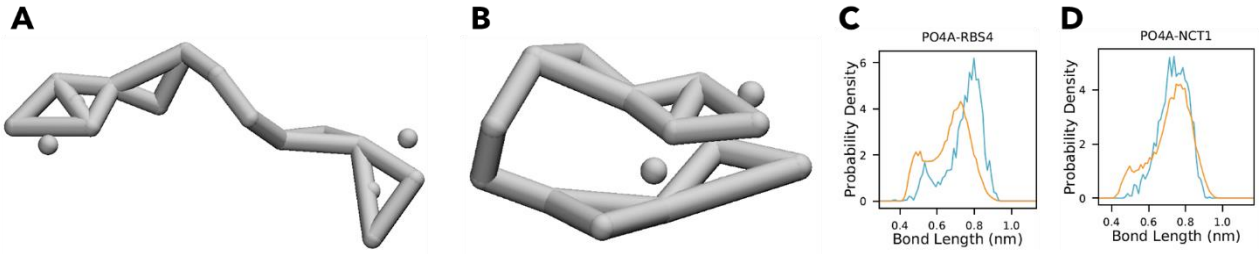


Figure 5.9 – Illustration of NADH (A) stretched state and (B) bent state and (C) UA (blue) and CG (orange) distance distributions between PO4A and RBS4 beads (since PO4A-RBS4 is not in **bold**, it means that we did not impose any potential on this bond). (D) Distance distributions between PO4A and NCT1 beads, which were initially forced. Pictures A and B were retrieved using VMD[163]. The plots in C and D were obtained from a script done by my colleague Gonçalo Vieira.

The bent state (**Figure 5.9**-**B**) is characterised by the "closing" of the molecule that can be due to: (1) a weak π-stacking in reduced molecules between adenine and nicotinamide moieties or (2) a salt bridge in the oxidised forms between phosphate groups and the last ribose and nicotinamide beads.

Even though only one topology file per molecule was used, we still had to use different parameters between oxidised and reduced forms. Hence, according to the oxidation state of each molecule, the dihedral RBS3-RBS4-NCT1-NCT2 behaves a bit different because of the dissimilar chemical behaviour. Thus, $NADP^+$ and NAD^+ have the same CG dihedral parameters but differ from the NADH and NADPH:

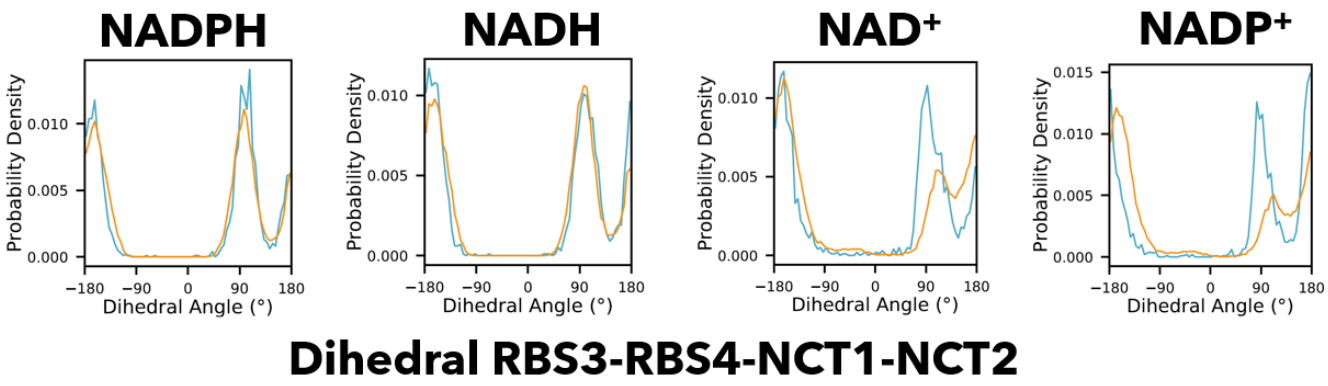


Figure 5.10 – Illustration that shows RBS3-RBS4-NCT1-NCT2 UA (blue) and CG (orange) dihedral distributions of NADPH, NADH, NAD^+ and $NADP^+$ (from left to right). These plots were drawn using a script done by my colleague Gonçalo Vieira.

As shown in **Figure 5.10**, the CG distributions between all molecules are a bit different. We tried as much as possible to use the same CG parameters between nicotinamide derivatives. However, and especially between $NADP^+$ and NADPH, when CG dihedral parameters from one would be used on the other, they would not recover accurately the UA dihedral parameters. Thus, the rationale behind this choice of differing nicotinamide derivative parameters according to their oxidation states relies mainly on the difference shown between NADPH and $NADP^+$. They show different conformational preferences.

5.1.9 Parameterisation of smaller molecules – main outcomes

In essence, the parameterisation of nucleotides and quinones was done successfully with overall SASA ratios below 5 % and, for non-charged species, a good calculated-experimental log P match. We, thus conclude that bead sizes and bead types were chosen with optimal accuracy.

Distributions that were a bit off were mostly due to collinearity issues, force-fields mismatch, parameters that we did not choose or equal potential parameters that were employed in different molecules.

SASAs above 5 % were tuned as much as possible, but ribitol, plastoquinone and geran-8-yl geran were still above that limit.

Octanol-water partition coefficients that were not near their experimental values were predominantly phosphorylated compounds, which was a bit mitigated by exchanging the bead types of phosphate beads from Q5 to Q4. Yet, overcoming this problem is still ongoing work. Alcohols were also seen to be too hydrophilic, mainly in octanol-water equilibria.

Moreover, nicotinamide derivatives were seen to adopt two different conformations (bent and stretched) using a single TOP file without forcing any potential, contrary to what was seen with Martini 2.

Without the accurate parameterisation of molecules such as FAD, NADH and MNQ, we would not be able to move forward to the larger simulations performed in this thesis. Thus, setting all these parameters was a vital step to perform the MD simulations.

5.2 NDH-2 and *S. aureus*' membrane

5.2.1 NDH-2 before minimisation

As mentioned in Section 4.1.2, we had some previous work before getting into the larger systems' simulations. Our protein of interest was not ready for simulations, and prior checks were necessary.

As a starting point and due to lack of time and information regarding the dimerization of NDH-2 in membranes, only monomer simulations were employed in this thesis. Yet, this is an ongoing work, and the simulation of the dimeric form is a future step.

After the use of the *Martinize2* tool to get the Coarse-Grain structure, initial simulations of our protein in water were established. Distances in the binding pockets of NADH and MND were checked using VMD tools[163] and compared to the distances obtained from the crystallographic structure. In the final NDH-2 structure, not a single bond had its bonded potential standard values that came from *Martinize2* altered.

Nonetheless, some FAD instability was observed. Thus, and to keep this cofactor in its crystallographic position, bonds were added to the NDH-2 CG ITP file. Moreover, Histidine-44 side chain bead was changed from TN5a to TQ2p ("p" is a label given to Q bead types that means "positive" and mean that they establish hydrogen bonds as donors[137]) to interact with Q4 beads of the negative phosphates.

Yet, when FAD was added, the system began to crash immediately. In a thorough inspection, we realised that some FAD-protein distances were below their LJ's σ value. This problem was solved by identifying the distances to be excluded using a script (*ExclusionsAndBonds.ipynb* that can be found in my GitHub[161]). In this script, by knowing the bead sizes of the two beads (R, S and T bead sizes values are tabulated) and using the coordinates in the GRO file of all beads, we were able to see which bead distances (calculated from their coordinates) were below their σ.

5.2.2 *S. aureus* membrane

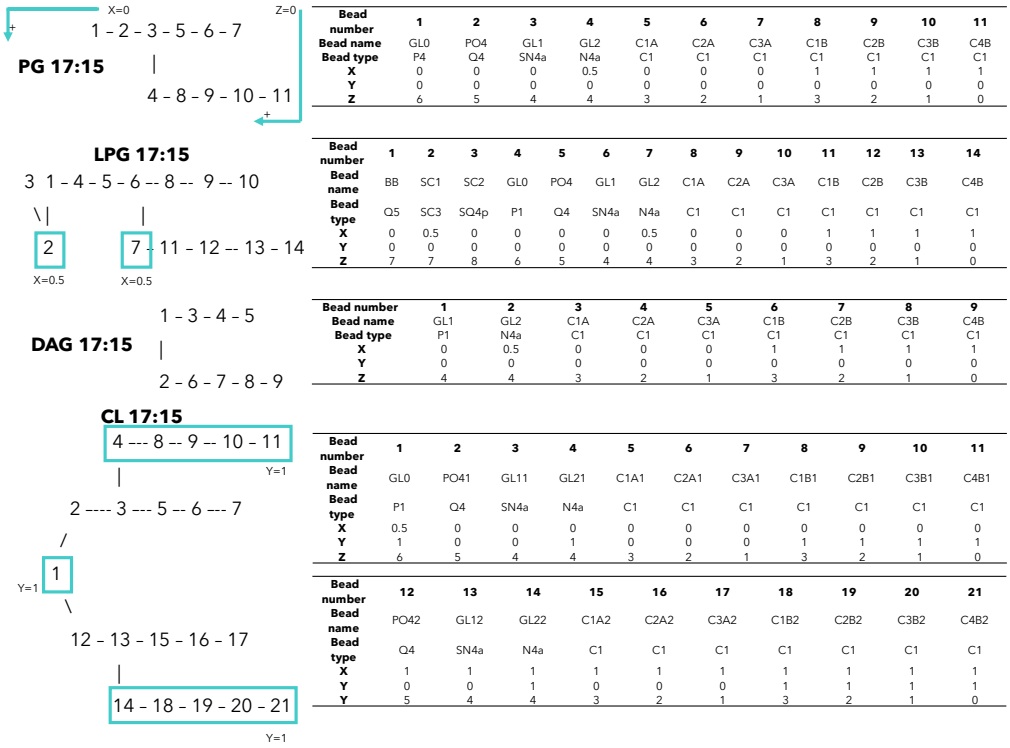
5.2.2.1 Lipids parameterisation

Lipids were parameterised following the intrinsic Martini rule, the building-block approach. We used an online Tutorial[211] held by our former colleague Dr. Luís Borges-Araújo and Dr. Manuel Melo on how to add new lipids into the *insane.py* (INSert membrANE) tool[171].

In this tutorial, the LPG topology is created by merging the pre-existing lysine and phosphatidylglycerol topologies for Martini 3, assuming that the bond potential between the lysine backbone and the glycerol PG has the same characteristics as the phosphate-glycerol bond.

Then, to add this lipid to *insane*, we defined pseudocoordinates for each of the CG beads from templates already available. In this case, all beads from the phosphate bead down should correspond to the default diacylglycerol template in the *insane* script. Thus, we got the pseudocoordinates for LPG 17:15, LPG 15:15, PG 17:15, PG 15:15, DAG 17:15, and DAG 15:15. For cardiolipin (CL 17:15), we used a pre-existing model template of an already built cardiolipin.

As an example, a figure with PG 17:15, LPG 17:15, DAG 17:15 and CL 17:15 templates and respective bead types and pseudocoordinates is shown:



Bead number	1	2	3	4	5	6	7	8	9	10	11
Bead name	GL0	PO4	GL1	GL2	C1A	C2A	C3A	C1B	C2B	C3B	C4B
Bead type	P4	Q4	SN4a	N4a	C1	C1	C1	C1	C1	C1	C1
X	0	0	0	0.5	0	0	0	1	1	1	1
Y	0	0	0	0	0	0	0	0	0	0	0
Z	6	5	4	4	3	2	1	3	2	1	0

Bead number	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Bead name	BB	SC1	SC2	GL0	PO4	GL1	GL2	C1A	C2A	C3A	C1B	C2B	C3B	C4B
Bead type	Q5	SC3	SQ4p	P1	Q4	SN4a	N4a	C1	C1	C1	C1	C1	C1	C1
X	0	0.5	0	0	0	0	0.5	0	0	0	1	1	1	1
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z	7	7	8	6	5	4	4	3	2	1	3	2	1	0

Bead number	1	2	3	4	5	6	7	8	9
Bead name	GL1	GL2	C1A	C2A	C3A	C1B	C2B	C3B	C4B
Bead type	P1	N4a	C1	C1	C1	C1	C1	C1	C1
X	0	0.5	0	0	0	1	1	1	1
Y	0	0	0	0	0	0	0	0	0
Z	4	4	3	2	1	3	2	1	0

Bead number	1	2	3	4	5	6	7	8	9	10	11
Bead name	GL0	PO41	GL11	GL21	C1A1	C2A1	C3A1	C1B1	C2B1	C3B1	C4B1
Bead type	P1	Q4	SN4a	N4a	C1	C1	C1	C1	C1	C1	C1
X	0.5	0	0	0	0	0	0	0	0	0	0
Y	1	0	0	1	0	0	0	1	1	1	1
Z	6	5	4	4	3	2	1	3	2	1	0

Bead number	12	13	14	15	16	17	18	19	20	21
Bead name	PO42	GL12	GL22	C1A2	C2A2	C3A2	C1B2	C2B2	C3B2	C4B2
Bead type	Q4	SN4a	N4a	C1	C1	C1	C1	C1	C1	C1
X	1	1	1	1	1	1	1	1	1	1
Y	0	0	1	0	0	0	1	1	1	1
Y	5	4	4	3	2	1	3	2	1	0

Figure 5.11 – Illustration of PG, LPG, DAG and CL templates and the respective pseudocoordinates given along with their bead types. On the left, each lipid is drawn with the respective beads, which correspond to the bead numbers presented in the tables on the right, the bead names, bead types and pseudocoordinates (x, y, z). Only 17:15 molecules are represented for the sake of simplicity.

In **Figure 5.11**, a scheme of how to obtain the pseudocoordinates is shown. The first chain of the molecule has an X = 0, then the first bead of the lateral chain and the first side chain of the amino acid residue, if they exist, have X = 0.5, and the rest has an X = 1. Then, from the last bead of the tail, we start adding one value in Z, meaning the last bead has Z = 1, the penultimate has Z = 2, and so on. The Y coordinate is perpendicular to the plane of the paper. Lateral chains and the glycerol beads have Y = 1 and the others Y = 0.

5.2.2.2 293 K: from gel phase to fluid phase

Membrane analysis would not be accurate if our membrane lipids were too compact or overly free. We tried to get the best temperature from which simulations would start in a gel phase (“frozen”) but would end up in a fluid phase, thus mimicking the role of branched lipids when the temperature decreases. The angle kink used for methyl branches in the last bead lipids to mimic that behaviour was found by trial-and-error.

Dissimilar combinations of different angle kink values with distinct temperatures were simulated. We tried angle kinks of 140º, 150º and 160º and temperatures varying between 283 K and 293 K. The combination of 150 º with a temperature of 293 K generated the desired results:

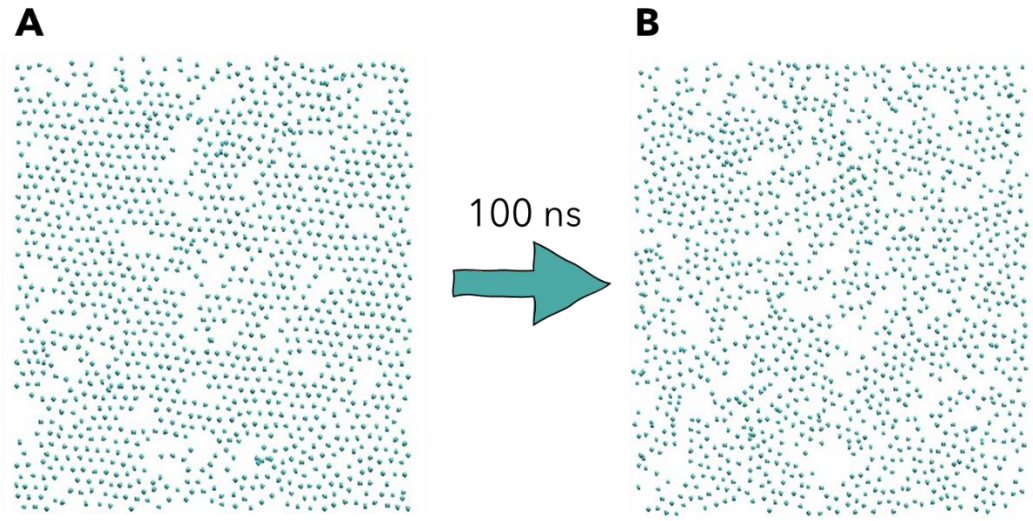


Figure 5.12 – Illustration of *S. aureus* CG membrane simulation using an angle kink of 150˚ at 293 K at different times. In A (0 s) – Gel phase - and B (100 ns) – Fluid phase - are shown the first bead of each lipid chain (a point of view of the membrane from above). These pictures were rendered using VMD[163].

This combination would allow sufficient membrane fluidity.

These simulations were performed before adding menaquinones, where aggregation problems started to appear.

5.2.2.3 Aggregation problems

Quinones were firstly parameterised using a C3 bead type, as suggested in Riccardo Alessandri's building-block table[149]. Menaquinones in a 2 % of the total amount of membrane lipids started to aggregate in the membrane and forming a bulk as shown:

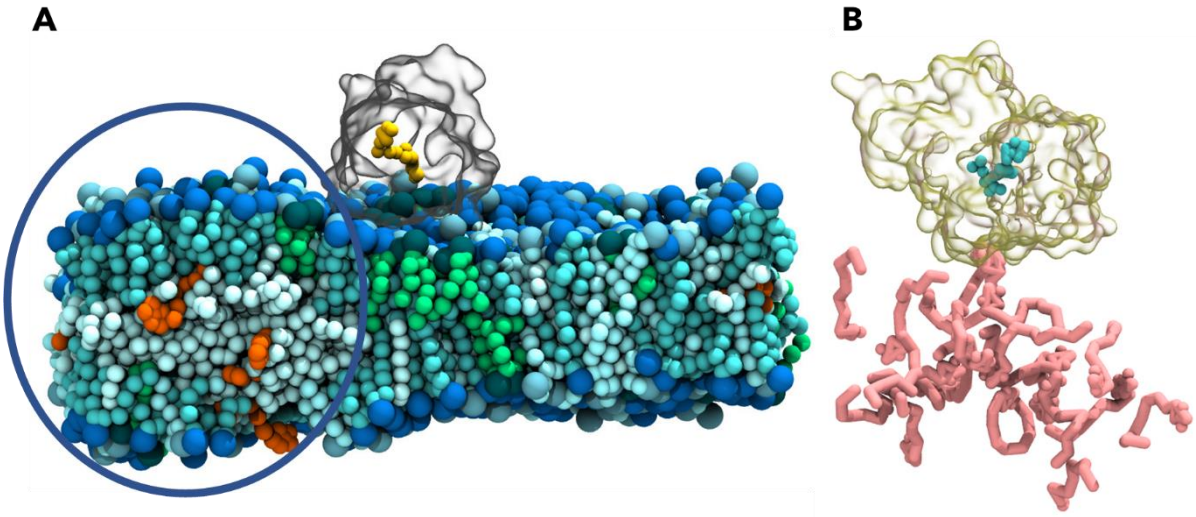


Figure 5.13 – Illustrations that show the aggregation problem in the *S. aureus* membrane. (A) A bulk in the membrane can be seen, with NDH-2 in grey and its FAD cofactor in yellow at top of the protein. (B) Shows NDH-2 in dark yellow in a membrane (not visible) and FAD in cyan, and many aggregated quinones in pink. Pictures were rendered using VMD[163].

Therefore, and judging by **Figure 5.13**-**A**, our membrane was not having a completely symmetrical behaviour due to bulks in some places as a consequence of menaquinone aggregation (**Figure 5.13**-**A**) with repercussions such as the greater probability of not seeing menaquinone-protein interactions (or seeing them as an artefact) or an uneven distribution of menaquinones throughout the membrane.

Menaquinones were seen to interact with themselves intermolecularly (mainly via the tails). To mitigate this, we increased the hydrophobicity of the chains from C3 to C2 and decreased their percentage in the membranes from 2 to 1 %. After running the membrane simulations again, we observed that these problems had been solved.

Having dealt with all these obstacles, we could proceed to the membrane simulations with NDH-2 and analysis.

5.2.3 NDH-2 and membrane model development – main outcomes

NDH-2 clashes were solved by finding which distances would be below the LJ's σ value. This, allied to the change of an amino acid residue side chain bead and the adding of bonds between NDH-2 and FAD contributed to the system stability and maintenance of FAD inside the protein.

Also, when moving on to membrane simulations, lipids were accurately parameterised and were seen to have just enough fluidity at a temperature of 293 K with the last bead kink of 150º, which allowed the next step of adding menaquinones. Menaquinones were added at 2 % of the total amount of lipids in the membrane with a chain bead type of (R)C3 and they were seen to aggregate. This would interfere with our analysis, so we increased the apolarity and decreased its concentration to C2 and 1 %, respectively. After this change, no aggregations were seen in systems with only lipids and menaquinones. Therefore, we could proceed to the final simulations and analysis.

6 MODEL APPLICATION

In this Chapter, I will report and discuss the results of the larger simulations, namely, NDH-2 with numerous substrates in water or in membrane. It is important that validation of all parameters added and modelling of interactions is done in a membrane environment, so that our simulations can replicate what happens *in vivo*.

6.1 NDH-2 MD simulations in water

We added to a system with NDH-2 different combinations of molecules, which will be highlighted in each section. To all systems were added either 20 quinone/quinol head molecules or 10 quinone/quinol molecules with 10 NADH molecules. The analysis of the former will be addressed first, followed by the analysis of the latter combination.

6.1.1 Contacts

Our first approach was to check whether menadione could interact with the prosthetic group, FAD. For that, a contact analysis approach was followed to study quinones/quinols-FAD interactions, thus NADHs were neglected and not added to these simulations. We immediately saw that interactions between MND and FAD were occurring, however, we wanted a thorough approach on which FAD moiety was interacting with MND and, if so, their difference in terms of contacts being established. Moreover, we wanted to extend this same study to a variety of other molecules, as well.

Thus, we generated two scripts with the same code, only changing which targeted Universe (adenine or flavin moieties) we were studying. These scripts were generated and based on the following assumptions:

- An interaction occurs if the distance between an adenine FAD bead or a flavin FAD bead with the substrate being studied is below or equal to 0.6 nm.
- To avoid overestimating the number of interactions, since each moiety has more than one bead, only one interaction per frame for each substrate could be considered. In other words, for each frame, only 0 or one contacts could be established.
- The percentage of contacts for each substrate per frame was calculated, taking into account the time of each simulation for each replicate (a total of 5 system replicates were simulated for each substrate).
- Each simulation of each replicate had at least 5 μs of simulation time so that it could be studied in an already considerable amount of time.

These scripts can be found in my GitHub[161] with the code not only for having those percentages but also for how the plot was created. I named the file codes *Plots_Contacts_Flavin_NDH2FAD.ipynb* and *Plots_Contacts_Adenosine_NDH2FAD.ipynb* for the studying of substrates-flavin and substrates-adenosine interactions, respectively. The results can be summarised in the next figure as follows:

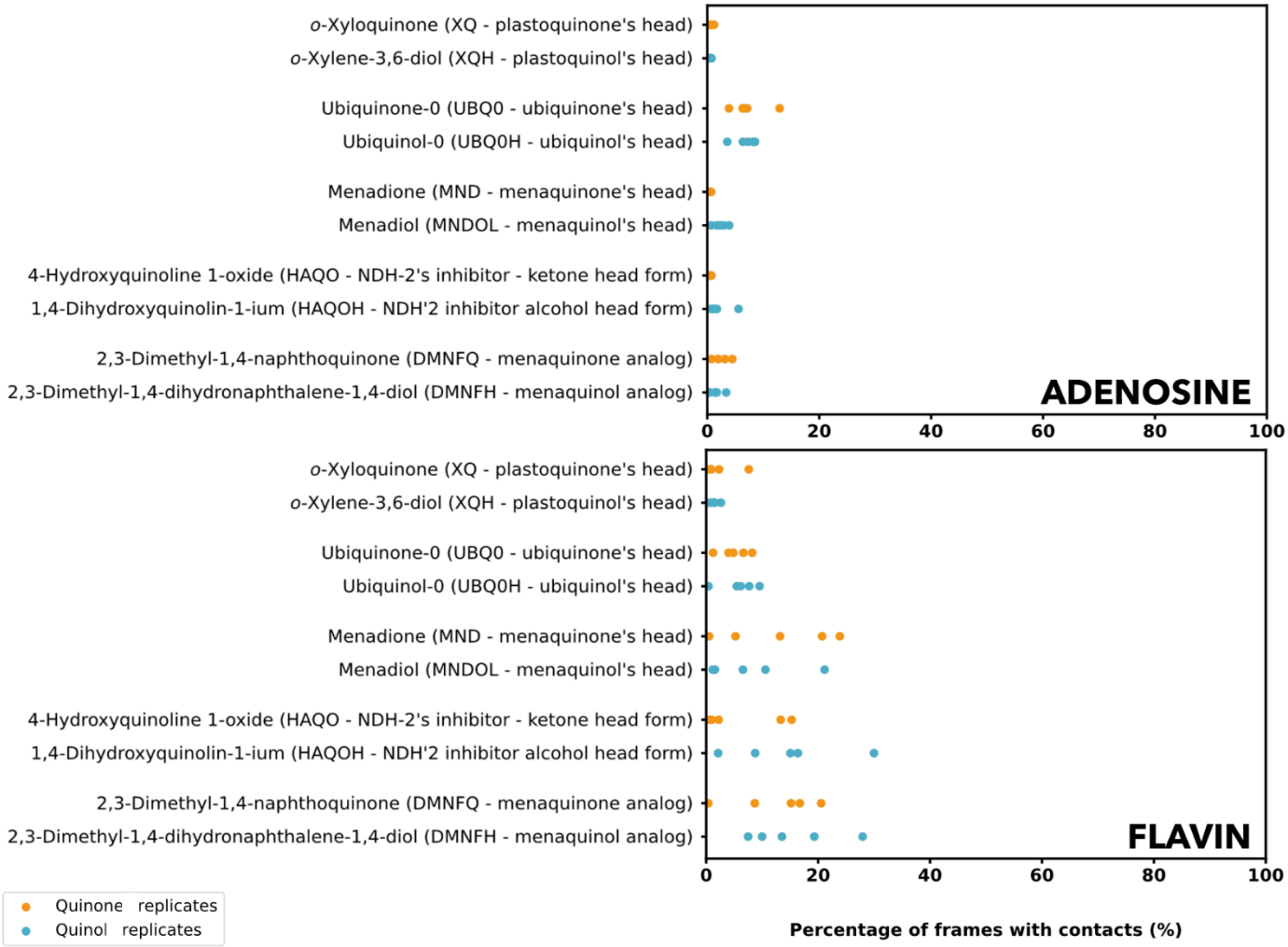


Figure 6.1 – Illustration that shows a graphic representation of the percentage of frames with contacts (%) for different substrates (on the left) and adenosine FAD moiety (top) and flavin FAD moiety (bottom). All five replicates are represented for each substrate-FAD system with at least 5 μs of simulation time. Each replicate has the exact same bonded and nonbonded parameters and number of molecules. Quinone replicates are represented with orange points, whereas quinol replicates are represented with blue points. These plots were created using scripts that can be found in my GitHub[161] (*Plots_Contacts_Adenosine_NDH2FAD.ipynb* – top - and *Plots_Contacts_Flavin_NDH2FAD.ipynb - bottom*).

From **Figure 6.1**, an evident result is shown by looking at both graphic representations: the preferential interaction binding site for quinones and quinols is the flavin moiety. There were almost no contacts with the adenosine region, except for the ubiquinone head, which seems to have a preference for this moiety. Nonetheless, this result comes out as a bit irrelevant since, in the membrane this moiety would not be visitable by these quinones. A plausible explanation for encountering a higher number of interactions here can be that this particular molecule is composed of the methoxy bead groups, which have some flexibility and can help this molecule to adopt particular spatial conformations that others cannot.

Coming back to the evident result, having a higher percentage of values for the flavin moiety is in agreement with the literature[92,97], which states that the core reaction from which menaquinone is reduced occurs in the flavin region. Thus, our results could validate what was already seen experimentally.

Then, when it comes to selectivity, our results also showed success in discriminating the preference for menadione and analogues for the NDH-2 flavin. When we look at **Figure 6.1-bottom** we can observe

that the percentage of frames with contacts between the flavin FAD and MND/DMNFQ are, at least, twice those for plastoquinone and ubiquinone heads. Once again, validation of experimental results is observed, since the literature is rather convincing in affirming that menaquinone is the quinone substrate of NDH-2[77,78,212,213]. Even the NDH-2 inhibitor[187] is found to have some replicates with a 2-fold increase in percentage values in comparison to PQ and UBQN heads.

Yet, this approach failed to successfully predict a possible preference of the prosthetic group for quinone or quinols, where the former was expected. Plus, substrate-FAD simulation replicates showed very different percentage of frames with contacts suggesting we should simulate longer to see if the values would converge. However, this approach seems good enough to mimic that molecules with a lower number of atoms (and fewer beads) such is plastoquinone, when compared to other quinones, probably interact less with NDH-2.

To validate that menaquinones prefer to interact with the flavin region, we depicted menadione occupancies around NDH-2. Besides, and to eliminate the different number of beads as a variable, we proceeded to the calculation of the so-called times of residence, where the amount of time a substrate interacts with the flavin moiety is the target.

6.1.2 Occupancies

Occupancy tests were done by using the VolMap tool in VMD[163], which draws surfaces of average occupancies during a specific time of simulation, that is, surfaces drawn by the regions where we can find a particular molecule most of the time. In this simulation, our protein was simulated, with 10 NADH and 10 MND. These tests allow the direct observation of where our substrates prefer to be during the the simulations. The results are in:

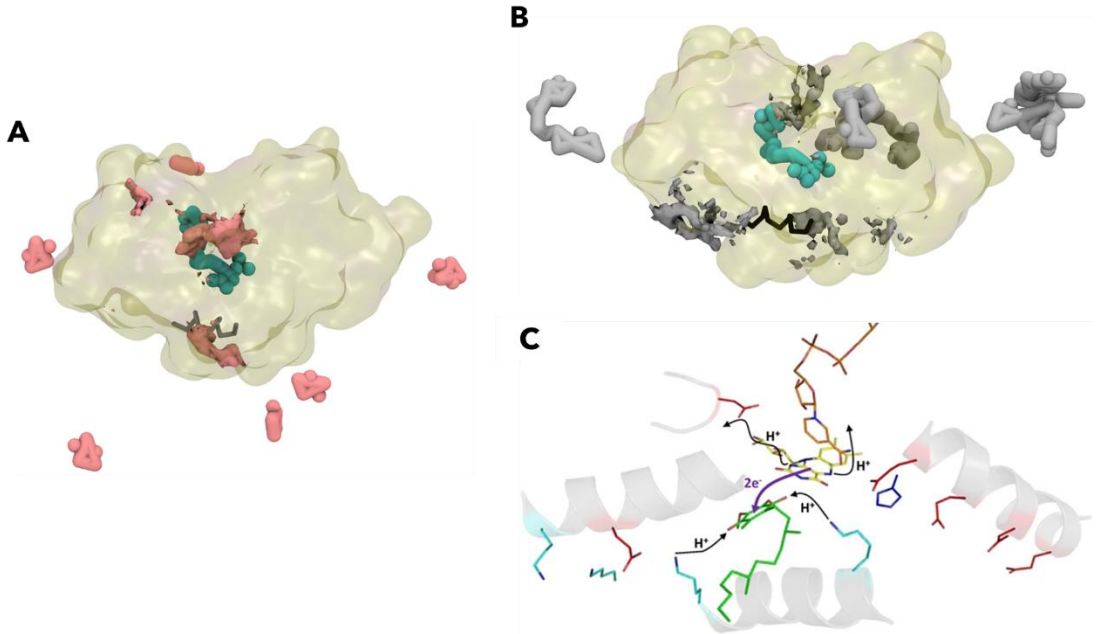


Figure 6.2 – Illustration of (A) MND molecules and occupancies in pink glued to the protein in yellow (FAD in turquoise), (B) NADH molecules and occupancies in grey stuck in NDH-2 in yellow (FAD in turquoise) and (C) Representation of the protonation of menaquinone (in green), where NADH was the proton donor (orange) using FAD (yellow) as intermediate. Oxygens are represented in red, nitrogens in blue and phosphates in brown. The α-helix represented in the bottom in grey, from where MNQ gets into the protein, is also represented in (A) and (B) in grey and black, respectively. Occupancies were calculated taking into account all frames in their simulation, but the molecule positions shown are from a single frame. (A) and (B) were rendered using VMD[163], and (C) was retrieved from Marreiros *et al*.[97].

According to the results shown in **Figure 6.2**-**A**, menadiones in an aqueous environment with NDH-2 tend to prefer its natural entry cavity (as shown in **Figure 6.2**-**C**) and NADH entry channel. The former validates that our results are congruent with experimental data, and the latter indicates that NADH entry channel is large enough to also accommodate menadiones. These latter results are not very relevant since in the membrane, the NADH entry channel would never be accessible to menaquinones; therefore, these results are just a consequence of the fact that these are aqueous simulations.

Similarly, and according to **Figure 6.2**-**B**, NADH visits its natural entry, although not penetrating as close to FAD as MND. The average measurements of some FAD beads and NADH showed that the latter was at a distance range of about 8-11 Å of the beads that correspond to the atoms that participate in the chemical reaction. This is a reasonable distance for electron transfer, being below 15 Å, the maximum distance for an electron transfer to occur, according to Marcus Theory[214,215]. Once again, our results showed a consequence of being aqueous simulations, where NADH may enter by another side (bottom), which in the membrane would not be accessible since NDH-2 would be on top of the membrane, so these should be ignored.

Thus, these results reinforce our previous ones regarding the place where substrates prefer to interact (in the region of the FAD flavin, the place frequently visitded by MND, and directly accessible through the membrane).

6.1.3 Residence times

Focusing now on the binding-unbinding time and hoping that these results would be more clarifying than those shown in Section 6.1.1, we performed an analysis based on binding events occurring when distances between substrates and flavin FAD would be lower than 6 Å. This analysis towards the study of a time extension of a binding event between substrates and flavin FAD was already focused on this region due to the conclusions retrieved in the previous Sections. Also, this study is an improvement in a way that accounts for the different number of beads of each substrate, thus allowing more than one interaction per frame. Moreover, this approach also excludes bindings that already happen in the first frame of simulation and situations where, at the end of the simulation, molecules are still bound. The reasoning behind this is the fact that we would never know how much time molecules were or will still be bound, which would unintentionally lead to misleading outcomes.

The results of the residence times for different substrates are the following:

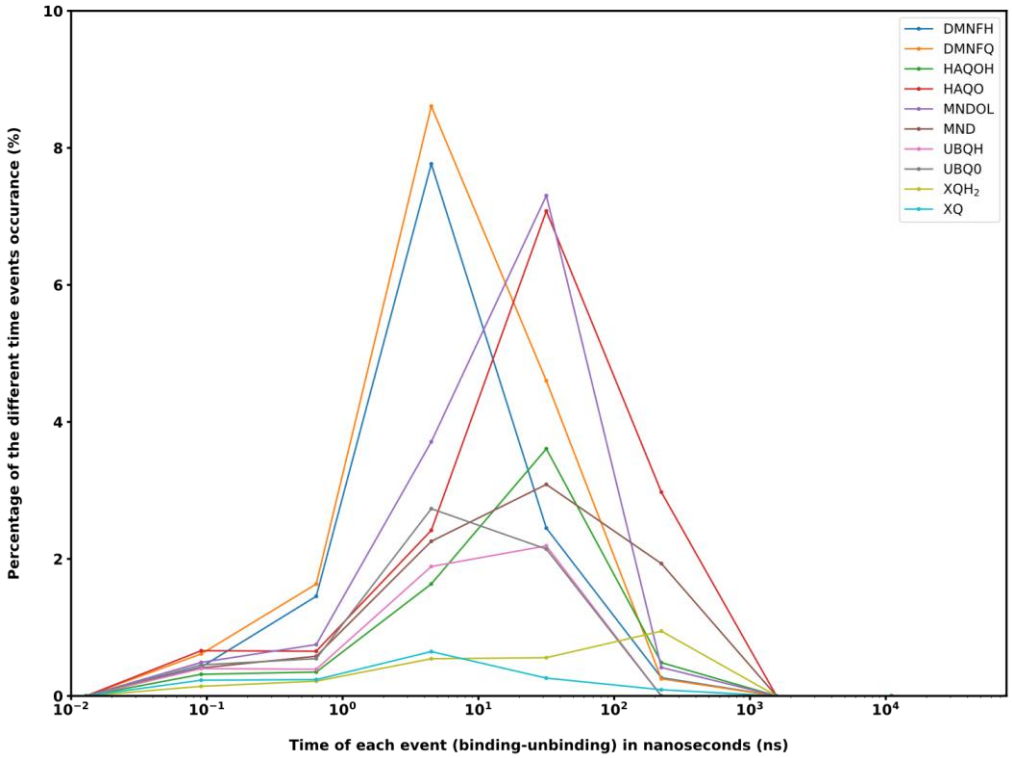


Figure 6.3 – Illustration of the percentage of occurrence (%) of different binding-unbinding time events (in nanoseconds, ns) in a logarithmic scale for several different substrates. The colour that represents each substrate can be found at the top right corner of the graphical representation. All simulations have at least 5 μs of simulation time, and five replicates were used. For each substrate, the sum of all replicates simulation times was used, thus accounting for all interactions occurring in all five replicates. This graphical representation results from a Python code that can be found in my GitHub[161] with the name of *Times_of_residence_NDH2FAD.ipynb*. Abbreviation definitions can be found in **Table S1**.

Looking at **Figure 6.3**, we can see that most of the binding events last about 10 to 100 ns. Furthermore, this approach also failed to discriminate between quinones and quinols, where even contradictory results might be seen. In some cases, some quinols are bound for more time than quinones and, in others, it is the other way around. Yet, these results showed that DMNFQ/DMNFH (menaquinone/menaquinol analogues), MND/MNDOL (menaquinone and menaquinol heads), and the NDH-2 inhibitors heads (HAQO/HAQOH, ketone/alcohol forms) were precisely the pairs that showed more binding events, in which menaquinone analogues seem to leave the binding pocket sooner. Likewise, the contacts approach had previously stated that these were the compounds with more frames with contacts, where NDH-2 could be quite selective and preferentially selected for its inhibitor, natural substrate and MNQ analogue. With this is mind, results seem to be congruent.

We should not retrieve strict conclusions from this graphical representation, but we may suppose that menaquinones will probably leave even sooner the binding pocket than the aforementioned molecules and will mostly have binding events with duration not longer than 10 ns. This reasoning is supported by the fact that DMNFQ is identical to MND but has an S bead instead of a T bead (with the same bead type). This is sufficient to encounter differences of one order of magnitude when it comes to binding-unbinding events. Probably the former, by being more hydrophobic, has less affinity to the binding pocket and stays there for less time. Menaquinone is much more hydrophobic, and thus, probably stays there for even less time or the same if the polarity of the chain does not influence the reaction occurrence (data that we cannot confirm with this study). The next step would be employing this exact same approach to menaquinone and seeing whether the results would confirm this theory.

6.1.4 Aspartate 302: a new role discovery?

An exciting observation was found in simulations with NADH and MND in water, and it lasted almost 1.3 μs. In the literature, the 302nd amino acid residue of this protein (an aspartate) is pointed out with the assumption of being essential to the protonation of FAD or deprotonation of $FADH_2$, ending up being vital for the enzyme catalytic activity[93,96,97]. Yet, the reason for being essential to the enzyme activity may be another, now addressed.

The two beads that compose this amino acid (one bead for the backbone and one bead to the side chain) were seen to "lock" NADH by its adenine in such a way that the nicotinamide NADH could directly interact with the flavin FAD, somehow helping NADH to bend:

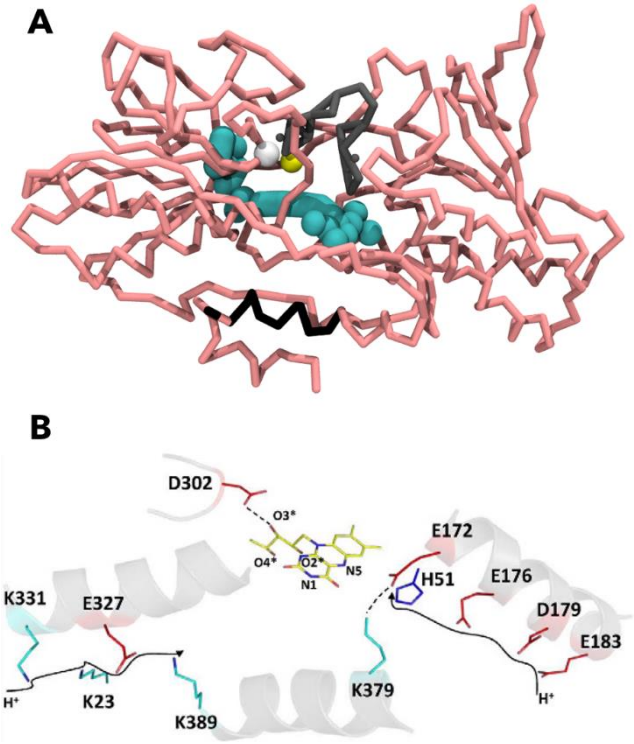


Figure 6.4 – Illustration that shows in (A) the NDH-2 backbone structure in pink, the FAD structure in turquoise, the NADH structure in grey and a yellow bead (D302 side chain bead) and a white bead (D302 backbone bead). In black, it is represented the α-helix delimited in (B) by the two lysines (K379-K389). D302 side chain is here shown in red, showing a hydrogen bond with oxygen (in red) in FAD (in yellow). Nitrogens are represented in blue. In (A) is seen the nicotinamide NADH moiety interacting with the flavin FAD moiety, and the adenine NADH moiety interacting with the D302 residue. Picture A was rendered using VMD[163], and Picture B was retrieved from Marreiros *et al*.[97].

This figure shows the direct interaction between the adenine NADH and D302. What even more exciting was the fact that the adenine bead seen to be near this aspartate was, in fact, no less than the amine bead, possibly indicating a hydrogen bond between both. Yet, the given average bond distance was around 4-5 Å (greater than a hydrogen bond distance), but we are also using CG. CG is not the best MD resolution to check the existence of hydrogen bonds because it lacks atomistic detail, so we cannot know if this hypothesis is true or not. Besides, the given distance is between the centre of the CG particles, not between atoms (hydrogens are sometimes far from the centre of the bead thus, a lower distance would be observed). The observation made should be clarified using atomistic simulations and then experimental procedures. An idea could be simply substituting the amine group in NADH for another similar non-bonding hydrogen-forming group and follow the NDH-2 catalytic enzyme activity.

Besides, nicotinamide amide bead and flavin beads average distances were seen to be around 5-7 Å when NADH would adopt the bent conformation, which is a distance within the range of possible electron transfer reactions to occur[214,215].

Hence, we hypothesise that D302 may be crucial for the enzyme activity because it interacts with adenine NADH in such a way that NADH can adopt its bent state and its nicotinamide moiety can, thus, interact with the flavin FAD, and the chemical reaction of electron transfer could take place. This hypothesis should be validated by using atomistic simulations and experimental procedures. A video of this observation (that lasted for about 1.3 μs) can be seen in the Supplementary Material Google Drive.

6.1.5 Main outcomes

From the NDH-2 aqueous simulations, validations of the experimental results were done. Menaquinone is the natural quinone substrate of NDH-2, and selectivity towards this molecule, its analogue and NDH-2 inhibitor were shown. Plus, it was theorised that binding events in menaquinones would last for 10 ns or less due to their greater hydrophobicity.

Occupancies and contacts approaches showed a preference of menaquinones for the flavin moiety of FAD and NADH preference to enter from the top of the protein, probably the same way it happens *in vivo*.

Finally, it was hypothesised that residue D302 is vital for NADH conformational stacking with the flavin FAD region so that electron transfer can take place *in vivo*. To prove this hypothesis, atomistic simulations and experimental studies should be undertaken.

6.2 NDH-2 MD CG membrane simulations

Larger membrane simulations ran for about 10 μs. These simulations were prepared using the *insane* method with the same number of NADH and MNQ molecules or without NADHs. These results complement the previous results shown in aqueous systems and also validate all of the smaller molecules parameters by showing that molecules could interact without clashing problems.

6.2.1 NDH-2 orientation onto the membrane

NDH-2 was oriented onto the membrane using, in part, our script (*NDH2Rotation.ipynb* in my GitHub[161]) and, on the other hand, the *insane* flags, specifically to place the protein in a certain position. First, with our script, we got a vector with the desired directionality so that the entrance of menaquinones would be near the membrane. Next, NDH-2 was positioned 1 nm above the membrane and became spontaneously adsorbed to the latter a short time after the simulation started. Simulations ran for about 10 μs. The next figure showcases some snapshots of our membrane:

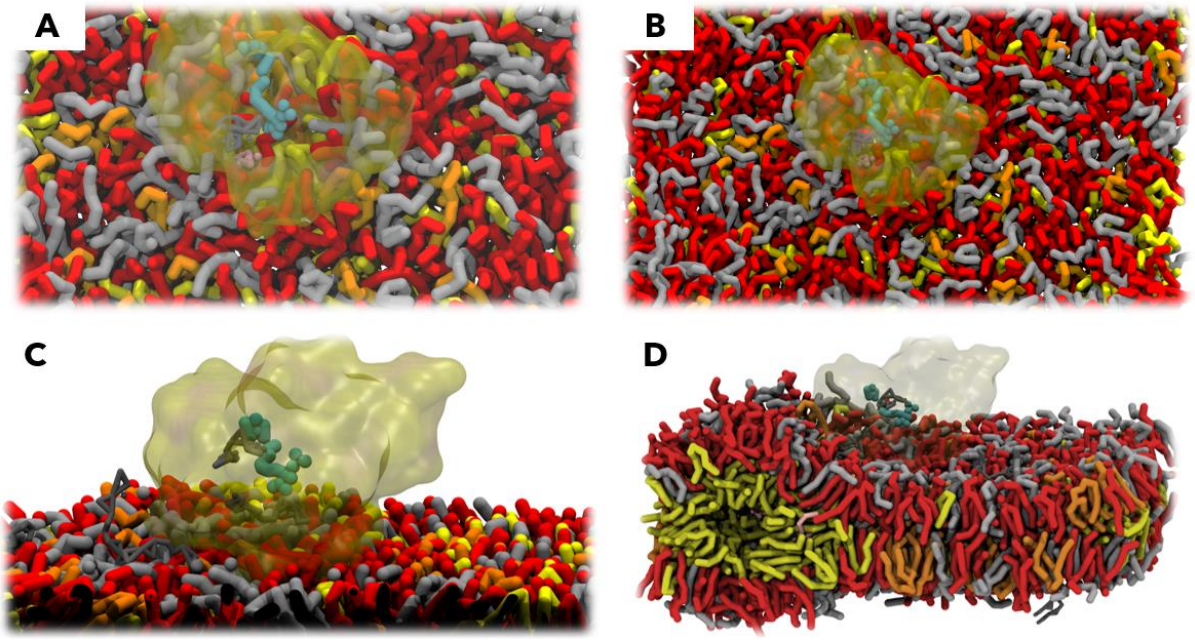


Figure 6.5 – Illustrations that show, from a top view, NDH-2 with FAD (yellow and turquoise, respectively) with an NADH in grey inside the protein and an MNQ trying to enter into NDH-2 in pink. In (C), we have a side view of our protein, and in (D), we have a broader image of all membrane. LPG17:15 and LPG15:15 are represented in silver, PG17:15 and PG15:15 are represented in red, DAG17:15 and DAG15:15 are represented in yellow and CL17:15 in orange. Pictures were rendered using VMD[163].

In **Figure 6.5**-**A** and **Figure 6.5**-**B**, we can see a top view of NDH-2 in the *S. aureus* membrane. Menaquinone is seen to be pulled out by our protein, replicating what happens *in vivo*, so that the redox reaction can take place. Yet, this conclusion would need a much more detailed analysis, such as seeing how much menaquinone can be above the phosphate lipids plane. These are my latest results, so it is still an ongoing work. To complement this analysis, in the future, we aim to perform Gibbs free energy calculations of the process of NDH-2 pulling plastoquinones, menaquinones, ubiquinones and HQNOs out of the membrane.

In **Figure 6.5**-**D**, in the simulation with NADH, we see a hole in the membrane caused by the repulsion of lipids to waters which ultimately led to aggregation of menaquinones. This problem is temporary. What happened was that the membrane structure that came from *insane* already had waters inside it, meaning that the cut-off radius used (0.21 nm) had to be increased. The structure that comes from *insane* has beads very close to each other and lipids end up being very flattened leading to more space for water beads to accommodate. This means that the hole is filled with water beads that, due to their hydrophilicity, repel lipids and attract menaquinones heads which end up aggregating.

Membrane simulations with only menaquinones were stable, and no bulk was observed (a small video is in the Google Drive, only showing the lipid phosphate beads, representing the membrane). Future analyses such as those aforementioned might complement our understanding of the MNQ-protein and NADH-protein interactions.

6.2.2 Occupancies

As already done in an aqueous environment, occupancy analyses were performed for menaquinone and NADH in membrane simulations:

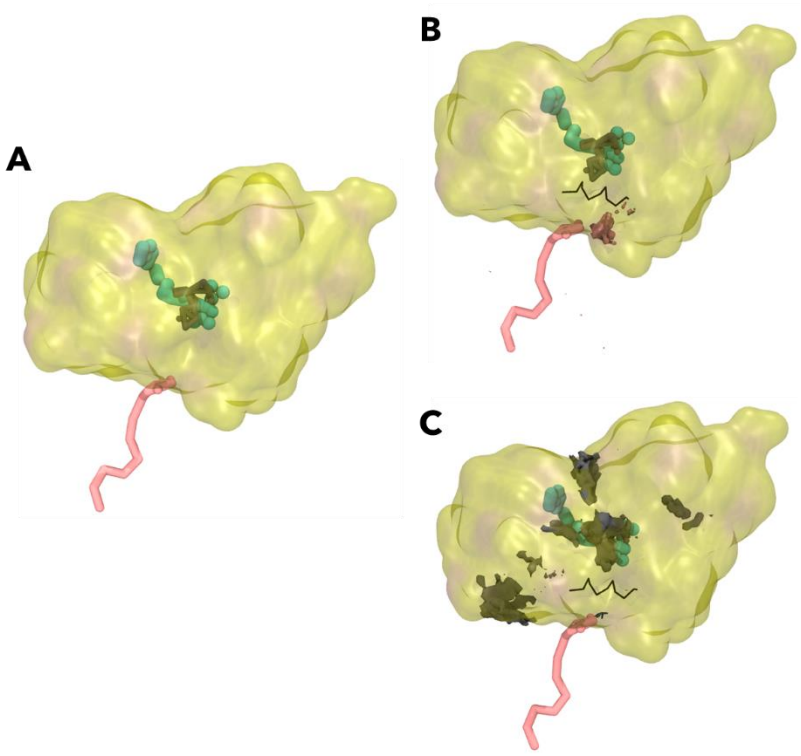


Figure 6.6 – Illustration of NDH-2 monomer (yellow), FAD (turquoise), MNQ (pink) and NADH (in grey). These results are from a membrane simulation, but the latter was hidden so that it would not be too confusing visually. In (A), neither the occupancy surfaces nor the α-helix (of MNQ entrance, in black) are shown, which are displayed in (B) and (C). In (B), the MNQ occupancy surface is shown in pink, and in (C), the NADH surfaces are in grey. Pictures rendered using VMD[163].

Figure 6.6 shows the occupancy analysis results, which are found to be within our expectations, being quite similar to the aqueous simulations but much more consistent with what most likely happens *in vivo*. In **Figure 6.6**-**B** it is shown the only entrance preference of menaquinone in the region of that α-helix in the bottom, whereas in **Figure 6.6**-**C**, we have multiple occupancy surfaces. Please note that an occupancy region is not equivalent to regions where that molecule enters. Therefore, having NADH occupancy regions at the protein bottom is insufficient to state that NADH can enter by that side. These occupancy regions are probably there because NDH-2 took some time to stabilise on top of the membrane, and so, NADH could interact with the NDH-2 bottom region. NADH most likely will interact with adenine FAD (protein top) as described in the literature[97]. To make sure about NADH entrance, other analyses would have to be performed or we could just analyse the occupancy regions right after NDH-2 being placed on top of the membrane.

6.2.3 Main outcomes

The main outcomes of the last two results are the successful orientation of NDH-2 onto the membrane (yet with a bulk in the membrane simulations with NADH) and the occupancy analyses that showed similar results to the aqueous ones. These ended up validating that in a membrane, menaquinones can only interact with our protein from its bottom. Although not entirely conclusive, NADH is most likely to interact by the top of the protein. Moreover, NDH-2 was seen to pull menaquinones out of the membrane frequently, showing their substrate-capturing capacity.

7 CONCLUSION

In this work, we managed to obtain CG Martini 3 parameters for 35 different molecules in a building-block approach, many with direct biological relevance as compounds involved in the energy metabolism. Several quinones and nucleotides were parameterised according to a bottom-up and a top-down approach, respectively. From the former approach, 20 molecules were parameterised, such as plastoquinone, ubiquinone and menaquinone, the three main quinones that constitute the quinone pools we can find in membranes. From the latter, 15 compounds were parameterised in a COW mapping approach, such as NADH, NADPH, FMN and FAD, which are known as electron donors to the respiratory membrane and/or can act as cofactors of the respiratory proteins. This parameterisation is of the utmost significance for future works involving these molecules within the scope of a better understanding of metabolic pathways, especially involving respiratory proteins. This way, all the Martini 3 community can now start using these parameters in their research, showing a major contribution from this work.

This work followed what had already been partially done in our group with Martini 2 with nucleotides, but now extending it to quinones. Since Martini 3 has a broader range of bead types and bead sizes that can more accurately mimic 2-to-1 and 3-to-1 mappings, parameters had to be redone from scratch. One of the primary achievements for our group was that, with these new parameters, we could see the nicotinamide derivatives adopting their two conformations (bent and stretched), only using a single topology file. This has a major impact since now we can accurately simulate their behaviour and eventually study the biological consequences of these nicotinamide derivatives opting for one of the states in different environments and seeing whether the adopted conformation would affect a specific chemical reaction and in a broader spectrum, the energy metabolism.

Nicotinamide derivative parameters were only dissimilar when their oxidation state was different, better reflecting the conformational behaviour according to their chemical one. Moreover, and now taking into account all the nucleotide sets, some distributions were slightly off when the same parameters were tried to replicate numerous molecules' behaviour.

Quinones were much easier to parameterise than nucleotides, and that was facilitated by the fact that their heads are rigid, hence almost no flexibility exists, and thus, constraints are more than enough to accurately represent their behaviour.

Plus, and intending to validate our parameterisation results, we calculated CG-to-AA SASAs ratios and the Connolly surfaces of our compounds to validate our bead size choices, and log Ps to validate our bead types. We found that phosphorylated compounds are too hydrophilic, and the most probable cause is that phosphate beads are too hydrophilic and that alcohols also seem to be parameterised as too polar, which might be mitigated with the addition of more bead type particles. SASAs ratios above 5% were tuned as best as we could, however, having ribitol with only three beads hampers a better result than this (even after deciding to map it unsymmetrically). Plastoquinone and G8YLG, we could not be improved since we had already used R beads for the bead chains, and the head SASAs were fine. Hence, and with all the other molecules ratios below 5 %, we moved on to other validations.

In octanol-water partition coefficients obtained by an alchemical analysis, we explained that we could, at least, try to quantify the polarity of phosphates by the calculation of a parameter called the second virial coefficient, from which we could retrieve information with regards to solute-solvent interactions. Yet, these overly polar results were mitigated when we made the decision to distinguish phosphate beads that would be alone or paired (internal paired phosphate beads, such as ours, which share their anionic character), thus decreasing the bead types' polarity from Q5 to Q4. Yet, only a slight improvement was observed.

The parameterisation of all these smaller molecules was vital to proceed to larger simulations and the construction of our NDH-2 CG structure and simulations, the main focus of this thesis. With these simulations, this protein can now be used for Martini 3 CG studies. This protein, a protein of the antibiotic multi-resistant *S. aureus*, was shown to be, in fact, a promising drug target. The fact that it can be found in many species, such as *S. aureus*, but not in humans, already piques our interest to the fact that we

can develop a specific inhibitor for this protein that will not be harmful to humans. Furthermore, and for future therapeutic genetics, this protein may be an alternative for defects in Complex I that lead to several diseases, such as Parkinson's. Moreover, this protein may also play a crucial role in *S. aureus* virulence. It is still a speculation, but it is thought that inhibition of this protein would inhibit the activation of several virulence factors of this bacterium, highlighting the importance of studying NDH-2[15,85].

The evident functions and roles of this protein in Health do not stop to surprise and persuade us that the study of this protein is of paramount relevance. With the choice of our protein more than justified, we engaged ourselves on this journey with NDH-2, and we first got its CG structure. When we were to add FAD to NDH-2, we noticed that the cofactor was leaving the protein so some bonds between them were applied. Beads between both that were in a too repulsive regime were excluded to avoid the system explosion. Then, all the process of creating and simulating the system was applied, and we could reach some conclusions regarding the region where the redox reaction takes place (in the flavin region) and the preference of the binding pocket for the menaquinone head, analogue and its inhibitor. The former conclusion was seen by having more contacts within the flavin region in detriment of the adenine region, and the latter is a major achievement, knowing that Martini 3 and simulations could accurately mimic the already seen experimental observations and show that menaquinone is, indeed, NDH-2 quinone substrate. Thus, our approaches revealed to be selective enough to represent the experimental results, even with the drawback of being aqueous simulations.

Then, by the calculation of the times of residence of several substrates, we not only reinforced the above mentioned conclusion, but we also hypothesised that menaquinone would be no more than ten ns in the binding pocket since, in the membrane, it is much harder for NDH-2 to pull out the menaquinone and because it is a larger and more hydrophobic molecule than the ones we have studied. Yet, this hypothesis would still need experimental validation. Another hypothesis concerned the alleged role that Aspartate 302 may have that is vital for the enzyme catalytic activity. Instead of directly being involved in the protonation/deprotonation of the cofactor, it is perhaps involved in helping NADH to adopt its bent state to interact with the flavin region and to allow the reaction to take place, as was shown in our simulations. Nonetheless, this behaviour would still have to be validated by using AA simulations and experimental procedures.

Continuing our work, but now in a membrane, we had to parameterise LPG, PG, CL and DAG, which were very alike from templates already available. We utilised the *insane* method. The membrane, which was found to have a gel-to-fluid transition of 293 K, was seen with a first issue regarding a significant aggregation of menaquinones. This was solved by lowering their concentration and chain beads' polarity. Then, the protein was placed 1 nm above the membrane and reorientated in such a way that the NDH-2 menaquinone entrance would fall onto the top of the membrane.

Occupancy tests showed the preference of menaquinones to enter from NDH-2 bottom region, precisely the same as seen in the literature[97]. This had been seen in the aqueous simulations, but there we were seeing several occupancy surfaces due to the inexistence of a membrane. We could also see that, in both aqueous and membrane simulations, NADH has several occupancy surfaces in the protein but it is most likely to penetrate it from its top. In membrane simulations where NADH was added, a lipidic bulk was seen, showing lipids and menaquinones aggregation that is due to the existence of water beads inside the membrane that already came from the *insane* structure. More importantly, these membrane simulations also showed NDH-2 pulling out menaquinones out of the membrane, successfully exhibiting its behaviour as a substrate-capturing enzyme. Yet, future analyses to validate this result, namely checking if menaquinones can be pulled out above the membrane plane, still have to be made.

This work is still an ongoing project, and some refinements should be done. For instance, the simulations with NDH-2 dimer are yet to be performed. Other approaches, such as the calculation of the Gibbs free energies of menaquinone, plastoquinone and ubiquinone being pulled out by NDH-2 still need to be done. This analysis can confirm the NDH-2 selectivity results. Besides, we still have to address some problems in the future such as the hydrophilicity of phosphates and alcohols.

We accomplished all the goals we proposed to in the beginning of this thesis. With this thesis, I hope that we can boost the search for a way to diminish *S. aureus* virulence using NDH-2 as a

therapeutic target and tackle the rising trend in cases (particularly antibiotic-resistant strains) in some countries.

8 REFERENCES

1. Kopinathan, A. *et al.* The next pandemic: antimicrobial resistance. *The golden age of antibiotics is under threat* https://www.science.org.au/curious/policy-features/next-pandemic-antimicrobial-resistance (2021).
2. CDC. What Exactly is Antibiotic Resistance? *Centers for Disease Control and Prevention* https://www.cdc.gov/drugresistance/about.html (2022).
3. WHO. Antimicrobial resistance. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance (2021).
4. Uddin, T. M. *et al.* Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *J. Infect. Public Health* **14**, 1750–1766 (2021).
5. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
6. WHO's first global report on antibiotic resistance reveals serious, worldwide threat to public health. https://www.who.int/news/item/30-04-2014-who-s-first-global-report-on-antibiotic-resistance-reveals-serious-worldwide-threat-to-public-health (2014).
7. Davies, O. L. & Bennett, S. WHO publishes list of bacteria for which new antibiotics are urgently needed. *WHO Newsletters* (2017).
8. Craft, K. M., Nguyen, J. M., Berg, L. J. & Townsend, S. D. Methicillin-resistant Staphylococcus aureus (MRSA): antibiotic-resistance and the biofilm phenotype. *Medchemcomm* **10**, 1231–1241 (2019).
9. Peres, D., Pina, E. & Fonseca Cardoso, M. Methicillin-Resistant Staphylococcus Aureus (MRSA) in a Portuguese hospital and its risk perception by health care professionals. *Revista Portuguesa de Saúde Pública* **29**, 132–139 (2011).
10. Petersen, A. *et al.* Increasing Incidences and Clonal Diversity of Methicillin-Resistant Staphylococcus aureus in the Nordic Countries - Results From the Nordic MRSA Surveillance. *Front. Microbiol.* **12**, 668900 (2021).
11. ECDC. Annual surveillance reports on antimicrobial resistance. *European Centre for Disease Prevention and Control* https://www.ecdc.europa.eu/en/antimicrobial-resistance/surveillance-and-disease-data/report (2022).
12. Kourtis, A. P. *et al.* Vital Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections - United States. *MMWR Morb. Mortal. Wkly. Rep.* **68**, 214–219 (2019).
13. Weiner-Lastinger, L. M. *et al.* The impact of coronavirus disease 2019 (COVID-19) on healthcare-associated infections in 2020: A summary of data reported to the National Healthcare Safety Network. *Infect. Control Hosp. Epidemiol.* **43**, 12–25 (2022).
14. Boswihi, S. S. & Udo, E. E. Methicillin-resistant Staphylococcus aureus: An update on the epidemiology, treatment options and infection control. *Current Medicine Research and Practice* **8**, 18–24 (2018).
15. Jenul, C. & Horswill, A. R. Regulation of Staphylococcus aureus Virulence. *Microbiol Spectr* **7**, (2019).
16. Jaradat, Z. W. *et al.* Methicillin Resistant Staphylococcus aureus and public fomites: a review. *Pathog. Glob. Health* **114**, 426–450 (2020).
17. Sakr, A., Brégeon, F., Mège, J.-L., Rolain, J.-M. & Blin, O. Staphylococcus aureus Nasal Colonization: An Update on Mechanisms, Epidemiology, Risk Factors, and Subsequent Infections. *Front. Microbiol.* **9**, 2419 (2018).
18. Gajdács, M. The Continuing Threat of Methicillin-Resistant Staphylococcus aureus. *Antibiotics (Basel)* **8**, (2019).
19. Ford, C. A., Hurford, I. M. & Cassat, J. E. Antivirulence Strategies for the Treatment of Staphylococcus aureus Infections: A Mini Review. *Front. Microbiol.* **11**, 632706 (2020).

20. Bush, L. M. & Vazquez-Pertejo, M. T. Staphylococcal infections. https://www.msdmanuals.com/professional/infectious-diseases/gram-positive-cocci/staphylococcal-infections (2019).
21. Kozajda, A., Jeżak, K. & Kapsa, A. Airborne Staphylococcus aureus in different environments-a review. *Environ. Sci. Pollut. Res. Int.* **26**, 34741–34753 (2019).
22. NHS. Endocarditis. *nhs.uk* https://www.nhs.uk/conditions/endocarditis/ (2022).
23. Mayo Clinic Staff. Osteomyelitis. *Mayo Clinic* https://www.mayoclinic.org/diseases-conditions/osteomyelitis/symptoms-causes/syc-20375913 (2020).
24. Pestana, E. Pneumonia: principais sintomas e como tratar - Lusíadas. *Lusíadas Saúde* https://www.lusiadas.pt/blog/doencas/pneumonia-principais-sintomas-como-tratar.
25. Better Health Channel. Gastroenteritis. *Better Health Channel* https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/gastroenteritis (2021).
26. O'Connell, K. & Cafasso, J. Septicemia. *Healthline* https://www.healthline.com/health/septicemia (2021).
27. BioRender. *BioRender* http://BioRender.com.
28. Donlan, R. M. Biofilms: microbial life on surfaces. *Emerg. Infect. Dis.* **8**, 881–890 (2002).
29. Oliveira, D., Borges, A. & Simões, M. Staphylococcus aureus Toxins and Their Molecular Activity in Infectious Diseases. *Toxins* **10**, (2018).
30. Algammal, A. M. *et al.* Methicillin-Resistant Staphylococcus aureus (MRSA): One Health Perspective Approach to the Bacterium Epidemiology, Virulence Factors, Antibiotic-Resistance, and Zoonotic Impact. *Infect. Drug Resist.* **13**, 3255–3265 (2020).
31. Kong, C., Neoh, H.-M. & Nathan, S. Targeting Staphylococcus aureus Toxins: A Potential form of Anti-Virulence Therapy. *Toxins* **8**, (2016).
32. Maillard, I., Luthi, F., Acha-Orbea, H. & Diggelmann, H. Role of the immune response induced by superantigens in the pathogenesis of microbial infections. *Parasitology* **115 Suppl**, S67-78 (1997).
33. de Moerloose, P., Casini, A. & Neerman-Arbez, M. Congenital fibrinogen disorders: an update. *Semin. Thromb. Hemost.* **39**, 585–595 (2013).
34. Wójcik-Bojek, U., Różalska, B. & Sadowska, B. Staphylococcus aureus-A Known Opponent against Host Defense Mechanisms and Vaccine Development-Do We Still Have a Chance to Win? *Int. J. Mol. Sci.* **23**, (2022).
35. 5.3: Components and structure - membrane fluidity. *Biology LibreTexts* https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_(Boundless)/05%3A_Structure_and_Function_of_Plasma_Membranes/5.03%3A_Components_and_Structure_-__Membrane_Fluidity (2018).
36. Kaneda, T. Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Rev.* **55**, 288–302 (1991).
37. Zhang, Y.-M. & Rock, C. O. Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.* **6**, 222–233 (2008).
38. Sen, S. *et al.* Growth-Environment Dependent Modulation of Staphylococcus aureus Branched-Chain to Straight-Chain Fatty Acid Ratio and Incorporation of Unsaturated Fatty Acids. *PLoS One* **11**, e0165300 (2016).
39. Freeman, C. *et al.* Revealing Fatty Acid Heterogeneity in Staphylococcal Lipids with Isotope Labeling and RPLC-IM-MS. *J. Am. Soc. Mass Spectrom.* **32**, 2376–2385 (2021).
40. Perkinelmer-Informatics. *ChemDraw*.
41. Inkscape Project. *Inkscape*. (2020).
42. Taormina, V. M., Unger, A. L., Schiksnis, M. R., Torres-Gonzalez, M. & Kraft, J. Branched-Chain Fatty Acids-An Underexplored Class of Dairy-Derived Fatty Acids. *Nutrients* **12**, (2020).
43. Short, S. A. & White, D. C. Metabolism of phosphatidylglycerol, lysylphosphatidylglycerol, and cardiolipin of Staphylococcus aureus. *J. Bacteriol.* **108**, 219–226 (1971).

REFERENCES

44. White, D. C. & Frerman, F. E. Extraction, characterization, and cellular localization of the lipids of Staphylococcus aureus. *J. Bacteriol.* **94**, 1854–1867 (1967).
45. Piggot, T. J., Holdbrook, D. A. & Khalid, S. Electroporation of the E. coli and S. Aureus membranes: molecular dynamics simulations of complex bacterial membranes. *J. Phys. Chem. B* **115**, 13381–13388 (2011).
46. Young, S. A., Desbois, A. P., Coote, P. J. & Smith, T. K. Characterisation of Staphylococcus aureus lipids by nanoelectrospray ionisation tandem mass spectrometry (nESI-MS/MS). *bioRxiv* 593483 (2019) doi:10.1101/593483.
47. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
48. Nicholls, D. G. & Ferguson, S. J. *Bioenergetics*. (Academic Press, 2013). doi:10.1016/c2010-0-64902-9.
49. Miesfeld, R. L. & McEvoy, M. M. *Biochemistry*. (W.W. Norton, 2016).
50. Garrett, R. H. & Grisham, C. M. *Biochemistry*. (Cengage Learning, 2017).
51. Cramer, W. A. & Knaff, D. B. *Energy Transduction in Biological Membranes*. (Springer New York). doi:10.1007/978-1-4612-3220-9.
52. Nelson, D. L. & Cox, M. M. *Principles of Biochemistry*. (W. H. Freeman, 2021).
53. Calisto, F., Sousa, F. M., Sena, F. V., Refojo, P. N. & Pereira, M. M. Mechanisms of Energy Transduction by Charge Translocating Membrane Proteins. *Chem. Rev.* **121**, 1804–1844 (2021).
54. Hirst, J. Mitochondrial complex I. *Annu. Rev. Biochem.* **82**, 551–575 (2013).
55. Sazanov, L. A. The mechanism of coupling between electron transfer and proton translocation in respiratory complex I. *J. Bioenerg. Biomembr.* **46**, 247–253 (2014).
56. Kampjut & Sazanov. Structure of respiratory complex I–An emerging blueprint for the mechanism. *Curr. Opin. Struct. Biol.* (2022).
57. Götz, F. & Mayer, S. Both terminal oxidases contribute to fitness and virulence during organ-specific Staphylococcus aureus colonization. *mBio* vol. 4 e00976-13 (2013).
58. Hammer, N. D. *et al.* Two heme-dependent terminal oxidases power Staphylococcus aureus organ-specific colonization of the vertebrate host. *MBio* **4**, e00241-13 (2013).
59. Melo, A. M. P., Bandeiras, T. M. & Teixeira, M. New insights into type II NAD(P)H:quinone oxidoreductases. *Microbiol. Mol. Biol. Rev.* **68**, 603–616 (2004).
60. Kerscher, S., Dröse, S., Zickermann, V. & Brandt, U. The three families of respiratory NADH dehydrogenases. *Results Probl. Cell Differ.* **45**, 185–222 (2008).
61. Marreiros, B. C., Sena, F. V., Sousa, F. M., Batista, A. P. & Pereira, M. M. Type II NADH:quinone oxidoreductase family: phylogenetic distribution, structural diversity and evolutionary divergences. *Environ. Microbiol.* **18**, 4697–4709 (2016).
62. Potter, A. The Skeletal Environment Mediates Staphylococcus aureus Virulence Responses and Metabolic Processes in Vivo. (Vanderbilt University, 2019).
63. Sousa, F. M. Exploring Monotopic quinone reductases. (Universidade Nova de Lisboa, 2022).
64. Furchgott, R. F. & Lee, K. S. High energy phosphates and the force of contraction of cardiac muscle. *Circulation* **24**, 416–432 (1961).
65. Cosgrove, M. Perinatal and infant nutrition. Nucleotides. *Nutrition* **14**, 748–751 (1998).
66. Sassone-Corsi, P. The cyclic AMP pathway. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).
67. Munro, A. W. & McLean, K. J. Electron Transfer Cofactors. in *Encyclopedia of Biophysics* (ed. Roberts, G. C. K.) 601–606 (Springer Berlin Heidelberg, 2013). doi:10.1007/978-3-642-16712-6_41.
68. Xiao, W., Wang, R.-S., Handy, D. E. & Loscalzo, J. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxid. Redox Signal.* **28**, 251–272 (2018).
69. Ding, T., Song, G., Liu, X., Xu, M. & Li, Y. Nucleotides as optimal candidates for essential nutrients in living organisms: A review. *J. Funct. Foods* **82**, 104498 (2021).

70. Sousa, F. M., Lima, L. M. P., Arnarez, C., Pereira, M. M. & Melo, M. N. Coarse-Grained Parameterization of Nucleotide Cofactors and Metabolites: Protonation Constants, Partition Coefficients, and Model Topologies. *J. Chem. Inf. Model.* **61**, 335–346 (2021).
71. Schoepp-Cothenet, B. *et al.* On the universal core of bioenergetics. *Biochim. Biophys. Acta* **1827**, 79–93 (2013).
72. Milshteyn, D., Cooper, G. & Deamer, D. Chemiosmotic energy for primitive cellular life: Proton gradients are generated across lipid membranes by redox reactions coupled to meteoritic quinones. *Sci. Rep.* **9**, 12447 (2019).
73. Kaurola, P., Sharma, V., Vonk, A., Vattulainen, I. & Róg, T. Distribution and dynamics of quinones in the lipid bilayer mimicking the inner membrane of mitochondria. *Biochim. Biophys. Acta* **1858**, 2116–2122 (2016).
74. Prince, R. C., Dutton, P. L. & Gunner, M. R. The aprotic electrochemistry of quinones. *Biochim. Biophys. Acta Bioenerg.* **1863**, 148558 (2022).
75. Franza, T. & Gaudu, P. Quinones: more than electron shuttles. *Res. Microbiol.* **173**, 103953 (2022).
76. Kurosu, M. & Begari, E. Vitamin K2 in electron transport system: are enzymes involved in vitamin K2 biosynthesis promising drug targets? *Molecules* **15**, 1531–1553 (2010).
77. Collins, M. D. & Jones, D. Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication. *Microbiol. Rev.* **45**, 316–354 (1981).
78. Nahaie, M. R., Goodfellow, M., Minnikin, D. E. & Hájek, V. Polar lipid and isoprenoid quinone composition in the classification of Staphylococcus. *J. Gen. Microbiol.* **130**, 2427–2437 (1984).
79. The International Union of Pure & (iupac), A. C. quinones. in *The IUPAC Compendium of Chemical Terminology* (International Union of Pure and Applied Chemistry (IUPAC), 2014). doi:10.1351/goldbook.q05015.
80. Liu, S., Cheng, W., Fowle Grider, R., Shen, G. & Li, W. Structures of an intramembrane vitamin K epoxide reductase homolog reveal control mechanisms for electron transfer. *Nat. Commun.* **5**, 3110 (2014).
81. Wakeman, C. A. *et al.* Menaquinone biosynthesis potentiates haem toxicity in Staphylococcus aureus. *Mol. Microbiol.* **86**, 1376–1392 (2012).
82. Schlievert, P. M. *et al.* Menaquinone analogs inhibit growth of bacterial pathogens. *Antimicrob. Agents Chemother.* **57**, 5432–5437 (2013).
83. Kirby, D. T., Savage, J. M. & Plotkin, B. J. Menaquinone (Vitamin K2) Enhancement of *Staphylococcus aureus* Biofilm Formation. *J. Biosci. Med.* **02**, 26–32 (2014).
84. Liu, M. & Lu, S. Plastoquinone and Ubiquinone in Plants: Biosynthesis, Physiological Function and Metabolic Engineering. *Front. Plant Sci.* **7**, 1898 (2016).
85. Schurig-Briccio, L. A. *et al.* Role of respiratory NADH oxidation in the regulation of Staphylococcus aureus virulence. *EMBO Rep.* **21**, e45832 (2020).
86. ENZYME: 1.6.5.9. *KEGG* https://www.genome.jp/dbget-bin/www_bget?ec:1.6.5.9.
87. IUBMB. EC 1.6.5.9. *IUBMB Enzyme Nomenclature* https://iubmb.qmul.ac.uk/enzyme/EC1/6/5/9.html.
88. Brito, J. A. *et al.* NADH:quinone oxidoreductase (NDH-II) from Staphylococcus aureus - holoprotein structure - 2.32 A resolution. (2018) doi:10.2210/pdb5na1/pdb.
89. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
90. Bragg, P. D. & Hou, C. Reduced nicotinamide adenine dinucleotide oxidation in Escherichia coli particles: II. NADH dehydrogenases. *Arch. Biochem. Biophys.* **119**, 202–208 (1967).
91. Rosário, A. L. *et al.* Expression, purification, crystallization and preliminary X-ray diffraction analysis of a type II NADH:quinone oxidoreductase from the human pathogen Staphylococcus aureus. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **71**, 477–482 (2015).
92. Sena, F. V. *et al.* Type-II NADH:quinone oxidoreductase from Staphylococcus aureus has two distinct binding sites and is rate limited by quinone reduction. *Mol. Microbiol.* **98**, 272–288 (2015).

REFERENCES

93. Sena, F. V. *et al.* Regulation of the mechanism of Type-II NADH: Quinone oxidoreductase from S. aureus. *Redox Biol* **16**, 209–214 (2018).
94. Hanukoglu, I. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* **43**, 206–209 (2015).
95. Lencina, A. M. *et al.* Type 2 NADH Dehydrogenase Is the Only Point of Entry for Electrons into the Streptococcus agalactiae Respiratory Chain and Is a Potential Drug Target. *MBio* **9**, (2018).
96. Sousa, F. M. *et al.* The key role of glutamate 172 in the mechanism of type II NADH:quinone oxidoreductase of Staphylococcus aureus. *Biochim. Biophys. Acta Bioenerg.* **1858**, 823–832 (2017).
97. Marreiros, B. C. *et al.* Structural and Functional insights into the catalytic mechanism of the Type II NADH:quinone oxidoreductase family. *Sci. Rep.* **7**, 42303 (2017).
98. Kaila, V. R. I. & Wikström, M. Architecture of bacterial respiratory chains. *Nat. Rev. Microbiol.* **19**, 319–330 (2021).
99. Kitajima-Ihara, T. & Yagi, T. Rotenone-insensitive internal NADH-quinone oxidoreductase of Saccharomyces cerevisiae mitochondria: the enzyme expressed in Escherichia coli acts as a member of the respiratory chain in the host cells. *FEBS Lett.* **421**, 37–40 (1998).
100. Dawson, T. M. & Dawson, V. L. Molecular pathways of neurodegeneration in Parkinson's disease. *Science* **302**, 819–822 (2003).
101. Seo, B. B. *et al.* Molecular remedy of complex I defects: rotenone-insensitive internal NADH-quinone oxidoreductase of Saccharomyces cerevisiae mitochondria restores the NADH oxidase activity of complex I-deficient mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9167–9171 (1998).
102. Seo, B. B., Matsuno-Yagi, A. & Yagi, T. Modulation of oxidative phosphorylation of human kidney 293 cells by transfection with the internal rotenone-insensitive NADH-quinone oxidoreductase (NDI1) gene of Saccharomyces cerevisiae. *Biochim. Biophys. Acta* **1412**, 56–65 (1999).
103. Seo, B. B., Wang, J., Flotte, T. R., Yagi, T. & Matsuno-Yagi, A. Use of the NADH-quinone oxidoreductase (NDI1) gene of Saccharomyces cerevisiae as a possible cure for complex I defects in human cells. *J. Biol. Chem.* **275**, 37774–37778 (2000).
104. Yagi, T. *et al.* Can a single subunit yeast NADH dehydrogenase (Ndi1) remedy diseases caused by respiratory complex I defects? *Rejuvenation Res.* **9**, 191–197 (2006).
105. Catania, A. *et al.* Arabidopsis thaliana alternative dehydrogenases: a potential therapy for mitochondrial complex I deficiency? Perspectives and pitfalls. *Orphanet J. Rare Dis.* **14**, 236 (2019).
106. Hazell, A. S., Itzhak, Y., Liu, H. & Norenberg, M. D. 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) decreases glutamate uptake in cultured astrocytes. *J. Neurochem.* **68**, 2216–2219 (1997).
107. Cook, G. M., Greening, C., Hards, K. & Berney, M. Energetics of pathogenic bacteria and opportunities for drug development. *Adv. Microb. Physiol.* **65**, 1–62 (2014).
108. Hummer, G. & Wikström, M. Molecular simulation and modeling of complex I. *Biochim. Biophys. Acta* **1857**, 915–921 (2016).
109. Leach, A. R. & Ar., L. *Molecular Modelling: Principles and Applications*. (Pearson Education, 2001).
110. van Gunsteren, W. F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed Engl.* **45**, 4064–4092 (2006).
111. Schlick, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*. (Springer New York, 2010). doi:10.1007/978-1-4419-6351-2.
112. Profeta, S., Jr. Molecular Modeling. *Kirk-Othmer Encyclopedia of Chemical Technology* (2005) doi:10.1002/0471238961.1315120516181506.a01.pub2.
113. Meller. Molecular dynamics. *Encyclopedia of life sciences* (2003).
114. Marshall, G. R. Limiting assumptions in molecular modeling: electrostatics. *J. Comput. Aided Mol. Des.* **27**, 107–114 (2013).
115. Frank, I. Classical motion of the nuclei in a molecule: A concept without alternatives. *ChemistrySelect* **5**, 1872–1877 (2020).

116. Car, R., de Angelis, F., Giannozzi, P. & Marzari, N. First-Principles Molecular Dynamics. in *Handbook of Materials Modeling: Methods* (ed. Yip, S.) 59–76 (Springer Netherlands, 2005). doi:10.1007/978-1-4020-3286-8_5.
117. Li, J. Basic Molecular Dynamics. in *Handbook of Materials Modeling: Methods* (ed. Yip, S.) 565–588 (Springer Netherlands, 2005). doi:10.1007/978-1-4020-3286-8_29.
118. Cai, Li & Yip. 1.09-Molecular Dynamics. *Comprehensive nuclear materials* **1**, 249–265 (2012).
119. Zhou, K. & Liu, B. *Molecular Dynamics Simulation: Fundamentals and Applications*. (Elsevier Science, 2022).
120. Becker, O. M., MacKerell, A. D., Jr, Roux, B. & Watanabe, M. *Computational Biochemistry and Biophysics*. (Taylor & Francis, 2001). doi:10.1023/A:1012420708544.
121. Doltsinis, N. Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches. (2006).
122. Monticelli, L. & Tieleman, D. P. Force Fields for Classical Molecular Dynamics. in *Biomolecular Simulations: Methods and Protocols* (eds. Monticelli, L. & Salonen, E.) 197–213 (Humana Press, 2013). doi:10.1007/978-1-62703-017-5_8.
123. Verma, A., Rangappa, S. M., Ogata, S. & Siengchin, S. *Forcefields for Atomistic-Scale Simulations: Materials and Applications*. (Springer Nature Singapore, 2022).
124. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).
125. Molecular Modelling & Simulation Team. Force fields and interactions. https://computecanada.github.io/molmodsim-md-theory-lesson-novice/01-Force_Fields_and_Interactions/index.html (2018).
126. Bunker, A. & Róg, T. Mechanistic Understanding From Molecular Dynamics Simulation in Pharmaceutical Research 1: Drug Delivery. *Front Mol Biosci* **7**, 604770 (2020).
127. Kmiecik, S. *et al.* Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **116**, 7898–7936 (2016).
128. Andrew McCammon, J. & Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*. (Cambridge University Press, 1987). doi:10.1017/CBO9781139167864.
129. Liguori, N., Croce, R., Marrink, S. J. & Thallmair, S. Molecular dynamics simulations in photosynthesis. *Photosynth. Res.* **144**, 273–295 (2020).
130. Chen, C. *et al.* A comparison of united atom, explicit atom, and coarse-grained simulation models for poly(ethylene oxide). *J. Chem. Phys.* **124**, 234901 (2006).
131. Ewen, J. P. *et al.* A Comparison of Classical Force-Fields for Molecular Dynamics Simulations of Lubricants. *Materials* **9**, (2016).
132. Goiko, M., de Bruyn, J. R. & Heit, B. Membrane Diffusion Occurs by Continuous-Time Random Walk Sustained by Vesicular Trafficking. *Biophys. J.* **114**, 2887–2899 (2018).
133. Borges-Araújo, L., Souza, P. C. T., Fernandes, F. & Melo, M. N. Improved Parameterization of Phosphatidylinositide Lipid Headgroups for the Martini 3 Coarse-Grain Force Field. *J. Chem. Theory Comput.* **18**, 357–373 (2022).
134. Karl Spinti, J., Nunes, F. N. & Melo, M. N. Room for Improvement in the Initial Martini 3 Parameterization of Peptide Interactions. *ChemRxiv* (2022) doi:10.26434/chemrxiv-2022-wgz1j.
135. Marrink, S. J. *et al.* Two decades of Martini: Better beads, broader scope. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* (2022) doi:10.1002/wcms.1620.
136. Ingólfsson, H. I. *et al.* The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 225–248 (2014).
137. Souza, P. C. T. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **18**, 382–388 (2021).
138. Chang, C.-E. A., Huang, Y.-M. M., Mueller, L. J. & You, W. Investigation of Structural Dynamics of Enzymes and Protonation States of Substrates Using Computational Tools. *Catalysts* **6**, (2016).

REFERENCES

139. Vanommeslaeghe, K., Guvench, O. & MacKerell, A. D., Jr. Molecular mechanics. *Curr. Pharm. Des.* **20**, 3281–3292 (2014).
140. Bauer, P., Hess, B. & Lindahl, E. *GROMACS 2022.3 Manual*. (2022). doi:10.5281/zenodo.7037337.
141. Bulacu, M. *et al.* Improved Angle Potentials for Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **9**, 3282–3292 (2013).
142. Tan, C., Jung, J., Kobayashi, C. & Sugita, Y. A singularity-free torsion angle potential for coarse-grained molecular dynamics simulations. *J. Chem. Phys.* **153**, 044110 (2020).
143. C. Silva, C., Cunha, C. & Vieira, M. *Química em reação - Física e Química A - 10.º ano*. (Porto Editora, 2021).
144. Naeem, R. Lennard-Jones Potential. *Chemistry LibreTexts* https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Physical_Properties_of_Matter/Atomic_and_Molecular_Properties/Intermolecular_Forces/Specific_Interactions/Lennard-Jones_Potential (2013).
145. de Jong, D. H., Baoukina, S., Ingólfsson, H. I. & Marrink, S. J. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Comput. Phys. Commun.* **199**, 1–7 (2016).
146. Joshi, S. Y. & Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Simul.* **47**, 786–803 (2021).
147. Atkins, P. & de Paula, J. *Physical Chemistry for the Life Sciences*. (W. H. Freeman, 2011).
148. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D. & Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **6**, e1000810 (2010).
149. Alessandri, R. *et al.* Martini 3 coarse-grained force field: Small molecules. *Adv. Theory Simul.* **5**, 2100391 (2022).
150. Ingólfsson, H. I. *et al.* Lipid organization of the plasma membrane. *J. Am. Chem. Soc.* **136**, 14554–14559 (2014).
151. Bruininks, B. M., Souza, P. C., Ingolfsson, H. & Marrink, S. J. A molecular view on the escape of lipoplexed DNA from the endosome. *Elife* **9**, (2020).
152. de Jong, D. H. *et al.* Atomistic and Coarse Grain Topologies for the Cofactors Associated with the Photosystem II Core Complex. *J. Phys. Chem. B* **119**, 7791–7803 (2015).
153. Bruininks, B. M. H., Souza, P. C. T. & Marrink, S. J. A Practical View of the Martini Force Field. *Methods Mol. Biol.* **2022**, 105–127 (2019).
154. Srinivasan, S., Zoni, V. & Vanni, S. Estimating the accuracy of the MARTINI model towards the investigation of peripheral protein-membrane interactions. *Faraday Discuss.* **232**, 131–148 (2021).
155. Lamprakis, C. *et al.* Evaluating the Efficiency of the Martini Force Field to Study Protein Dimerization in Aqueous and Membrane Environments. *J. Chem. Theory Comput.* **17**, 3088–3102 (2021).
156. Marrink, S. J. & Mark, A. E. Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. *J. Am. Chem. Soc.* **125**, 15233–15242 (2003).
157. Alessandri, R. *et al.* Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **15**, 5448–5460 (2019).
158. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
159. Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2021).
160. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17 (2012).
161. Barriga, R. MasterThesis: Scripts used during my Master Thesis 2021/2022. *Github* https://github.com/RodrigoBarrigaAlves/MasterThesis (2022).
162. Barnoud, J. CG Builder. *CGBuilder — Visual tool to build CG molecule models* https://jbarnoud.github.io/cgbuilder/ (2021).
163. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–8, 27–8 (1996).

164. (Marrink-lab). Martinize2. *GitHub* https://github.com/marrink-lab/vermouth-martinize.
165. Periole, X., Cavalli, M., Marrink, S.-J. & Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **5**, 2531–2543 (2009).
166. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
167. Zhang, J., Zhang, H., Wu, T., Wang, Q. & van der Spoel, D. Comparison of Implicit and Explicit Solvent Models for the Calculation of Solvation Free Energy in Organic Solvents. *J. Chem. Theory Comput.* **13**, 1034–1043 (2017).
168. Fehér, P. P. & Stirling, A. Assessment of reactivities with explicit and implicit solvent models: QM/MM and gas-phase evaluation of three different Ag-catalysed furan ring formation routes. *New J. Chem.* **43**, 15706–15713 (2019).
169. Zhao, Y.-P., Wang, F.-C. & Chi, M. Molecular Dynamics Simulation and Molecular Orbital Method. in *Handbook of Adhesion Technology* (eds. da Silva, L. F. M., Öchsner, A. & Adams, R. D.) 1559–1595 (Springer International Publishing, 2018). doi:10.1007/978-3-319-55411-2_52.
170. Molecular Modelling & Simulation Team. Periodic boundary conditions. *Practical considerations for Molecular Dynamics* https://computecanada.github.io/molmodsim-md-theory-lesson-novice/04-Periodic_Boundary/index.html (2018).
171. Wassenaar, T. A., Ingólfsson, H. I., Böckmann, R. A., Tieleman, D. P. & Marrink, S. J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J. Chem. Theory Comput.* **11**, 2144–2155 (2015).
172. Keith, J. A. *et al.* Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.* **121**, 9816–9872 (2021).
173. Mishra, S. K. & Ram, B. Steepest Descent Method. in *Introduction to Unconstrained Optimization with R* (eds. Mishra, S. K. & Ram, B.) 131–173 (Springer Singapore, 2019). doi:10.1007/978-981-15-0894-3_6.
174. Caceres-Delpiano, J., Wang, L.-P. & Essex, J. W. The automated optimisation of a coarse-grained force field using free energy data. *Phys. Chem. Chem. Phys.* **23**, 24842–24851 (2021).
175. Walton, E. B. & Vanvliet, K. J. Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **74**, 061901 (2006).
176. Gallo, M. T. *et al.* Novel procedure for thermal equilibration in molecular dynamics simulation. *Mol. Simul.* **35**, 349–357 (2009).
177. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
178. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
179. Molecular Modelling & Simulation Team. Controlling temperature. *Practical considerations for Molecular Dynamics* https://computecanada.github.io/molmodsim-md-theory-lesson-novice/07-thermostats/index.html (2022).
180. Molecular Modelling & Simulation Team. Controlling pressure. *Practical considerations for Molecular Dynamics* https://computecanada.github.io/molmodsim-md-theory-lesson-novice/08-barostats/index.html (2022).
181. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
182. Del Tatto, V., Raiteri, P., Bernetti, M. & Bussi, G. Molecular Dynamics of Solids at Constant Pressure and Stress Using Anisotropic Stochastic Cell Rescaling. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **12**, 1139 (2022).
183. Pedersen, B. Barostats in molecular dynamics. *CompChems* https://www.compchems.com/barostats-in-molecular-dynamics/ (2022).

REFERENCES

184. Schmid, N. *et al.* Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **40**, 843–856 (2011).
185. Páll, S. & Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **184**, 2641–2650 (2013).
186. Malde, A. K. *et al.* An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **7**, 4026–4037 (2011).
187. Petri, J. *et al.* Structure of the NDH-2 - HQNO inhibited complex provides molecular insight into quinone-binding site inhibitors. *Biochim. Biophys. Acta Bioenerg.* **1859**, 482–490 (2018).
188. Lindahl, Abraham, Hess & Spoel, V. der. *GROMACS 2019.6 Manual*. (Zenodo, 2020). doi:10.5281/zenodo.3685925.
189. Goga, N., Rzepiela, A. J., de Vries, A. H., Marrink, S. J. & Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **8**, 3637–3649 (2012).
190. Rehal, R. P. *et al.* The influence of mild acidity on lysyl-phosphatidylglycerol biosynthesis and lipid membrane physico-chemical properties in methicillin-resistant Staphylococcus aureus. *Chem. Phys. Lipids* **206**, 60–70 (2017).
191. Klimovich, P. V., Shirts, M. R. & Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.* **29**, 397–411 (2015).
192. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
193. Bannan, C. C., Calabró, G., Kyu, D. Y. & Mobley, D. L. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.* **12**, 4015–4024 (2016).
194. Daberdaku, S. Identification of protein pockets and cavities by Euclidean Distance Transform. in (11 2018).
195. Gromiha, M. M., Nagarajan, R. & Selvaraj, S. Protein Structural Bioinformatics: An Overview. in *Encyclopedia of Bioinformatics and Computational Biology* (eds. Ranganathan, S., Gribskov, M., Nakai, K. & Schönbach, C.) 445–459 (Academic Press, 2019). doi:10.1016/B978-0-12-809633-8.20278-1.
196. Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **16**, 548–558 (1983).
197. Hermosilla, P. *et al.* Interactive GPU-based generation of solvent-excluded surfaces. *Vis. Comput.* **33**, 869–881 (2017).
198. Daberdaku, S. Command: surface. *UCSF Resource for Biocomputing, Visualization, and Informatics* https://www.cgl.ucsf.edu/chimerax/docs/user/commands/surface.html (2022).
199. Rowland, R. S. & Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J. Phys. Chem.* **100**, 7384–7391 (1996).
200. Bondi, A. van der Waals Volumes and Radii. *J. Phys. Chem.* **68**, 441–451 (1964).
201. Melo, M. N., Ingólfsson, H. I. & Marrink, S. J. Parameters for Martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **143**, 243152 (2015).
202. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
203. Roel-Touris, J. & Bonvin, A. M. J. J. Coarse-grained (hybrid) integrative modeling of biomolecular interactions. *Comput. Struct. Biotechnol. J.* **18**, 1182–1190 (2020).
204. Henderson, A. R. The bootstrap: a technique for data-driven statistics. Using computer-intensive analyses to explore experimental data. *Clin. Chim. Acta* **359**, 1–26 (2005).
205. Lebigot, E. O. *Uncertainties: a Python package for calculations with uncertainties*. (2010).
206. Hansch, C. *Exploring Quantitative Structure-Activity Relationships (QSAR); Hydrophobic, Electronic, and Steric Constants*. (1995).
207. Sangster, J. Octanol-Water Partition Coefficients of Simple Organic Compounds. *J. Phys. Chem. Ref. Data* **18**, 1111–1229 (1989).

208. Rich, P. R. & Harper, R. Partition coefficients of quinones and hydroquinones and their relation to biochemical reactivity. *FEBS Lett.* **269**, 139–144 (1990).
209. Schmalhorst, P. S., Deluweit, F., Scherrers, R., Heisenberg, C.-P. & Sikora, M. Overcoming the Limitations of the MARTINI Force Field in Simulations of Polysaccharides. *J. Chem. Theory Comput.* **13**, 5039–5053 (2017).
210. Abraham, M. H., Whiting, G. S., Fuchs, R. & Chambers, E. J. Thermodynamics of solute transfer from water to hexadecane. *J. Chem. Soc. Perkin Trans. 2* 291–300 (1990) doi:10.1039/P29900000291.
211. Borges-Araújo, L. & N. Melo, M. 2: Lipid bilayers II - Complex mixtures using Insane. *Martini* http://cgmartini.nl/index.php/2021-martini-online-workshop/tutorials/562-1-lipid-bilayers-ii (2021).
212. Yano, T., Li, L.-S., Weinstein, E., Teh, J.-S. & Rubin, H. Steady-state kinetics and inhibitory action of antitubercular phenothiazines on mycobacterium tuberculosis type-II NADH-menaquinone oxidoreductase (NDH-2). *J. Biol. Chem.* **281**, 11456–11463 (2006).
213. Schurig-Briccio, L. A., Yano, T., Rubin, H. & Gennis, R. B. Characterization of the type 2 NADH:menaquinone oxidoreductases from Staphylococcus aureus and the bactericidal action of phenothiazines. *Biochim. Biophys. Acta* **1837**, 954–963 (2014).
214. Marcus, R. A. & Sutin, N. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **811**, 265–322 (1985).
215. Kuss-Petermann, M. & Wenger, O. S. Electron Transfer Rate Maxima at Large Donor-Acceptor Distances. *J. Am. Chem. Soc.* **138**, 1349–1358 (2016).

9 SUPPLEMENTARY INFORMATION

You may find all the Supplementary Material in the following link: https://drive.google.com/drive/folders/1dVqtz8L0lzGae23aAR6nWLKy6-sALITj?usp=sharing

Table S1 - All 35 molecules parameterised. In this table, whenever possible, compounds are identified with their PubChem CID. Whenever that is not possible, a SMILES code is given.

Group	Abbreviation	Compounds	PubChem CID
Nucleotides		**Adenine set**	
	ADOS	Adenosine	60961
	ADP	Adenosine Diphosphate	6022
	AMP	Adenosine Monophosphate	6083
	ATP	Adenosine Triphosphate	5957
		Flavin set	
	$FADH_2$	Dihydroflavine Adenine Dinucleotide	446013
	FAD	Flavin Adenine Dinucleotide	643975
	FMN	Flavin Mononucleotide	643976
	RBFL	Riboflavin	493570
		Nicotinamide set	
	NAD^+	Nicotinamide Adenine Dinucleotide (oxidised)	15938971
	NADH	Nicotinamide Adenine Dinucleotide (reduced)	21604869
	$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate (oxidised)	15938972
	NADPH	Nicotinamide Adenine Dinucleotide Phosphate (reduced)	15983949
		Thiamine set	
	THI	Thiamine	1130
	TMP	Thiamine Monophosphate	15942892

	TPP	Thiamine Pyrophosphate	15938963
Quinones		**Isoprenoid (ISO) set**	
	G8YLG	Tetraene-like molecule/Geran-8-yl geran	446073
		Menaquinone set	
	DMNFH	2,3-Dimethyl-1,4-naphthalenediol	5327094
	DMNFQ	2,3-Dimethyl-1,4-naphthoquinone	16615
	GBZ	Geranylbenzoquinone analog/ Methyl-geranylbenzoquinone	*a
	MNDOL	Menadiol	10209
	MND	Menadione	4055
	MNQOL	Menaquinol-8	45479636
	MNQ	Menaquinone-8	5376507
		HQNO set	
	HAQOH	1,4-Dihydroxyquinolin-1-ium	*b
	HQNOH	2-Heptyl-1-Hydroxyquinolin-1-ium-4-ol	4634558
	HQNO	2-Heptyl-4-quinolinol 1-oxide	1561
	HAQO	4-Hydroxyquinoline 1-oxide	76437
		Plastoquinone set	
	PQOL	Plastoquinol-8	*c
	PQ	Plastoquinone-8	101967154
	XQH	*o*-Xylene-3,6-diol	69100
	XQ	*o*-Xyloquinone	10688
		Ubiquinone set	
	UBQ0H	Ubiquinol-0	11183027
	UBQH	Ubiquinol-8	25074411
	UBQ0	Ubiquinone-0	69068

	UBQN	Ubiquinone-8	5283546

*SMILES is given when the molecule does not exist on PubChem.
[a]O=C1C(C/C=C(C)/CC/C=C(C)\C)=C(C)C(C=C1)=O
[b][H]OC(C1=CC=CC=C21)=CC=[N+]2O[H]
[c][H]OC(C=C1C/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)/CC/C([H])=C(C)\C)=C(C)C(C)=C1O[H]

Table S2 - Relative CG-to-AA difference (% of AA SASA) of all molecules and some of their composing moieties. Uncertainties are between parentheses and correspond to the uncertainty of the last digit. For instance, -0.69(7) = -0.69 ± 0.07.

Group	Abbreviation	Compounds	CG SASA	AA SASA	Relative CG-to-AA difference (% of AA SASA)
Nucleotides		**Adenine set**			
	ADOS[a]	Adenosine	5.274(3)	5.311(2)	-0.69(7)
	ADP[a]	Adenosine Diphosphate	6.92(1)	6.94(1)	-0.2(2)
	AMP[a]	Adenosine Monophosphate	6.207(4)	6.1883(4)	0.29(9)
	ATP	Adenosine Triphosphate	7.61(2)	7.65(2)	-0.6(3)
		Flavin set			
	FLA[b]	Flavin	4.9955(9)	5.0708(7)	-1.48(2)
	FAD	Flavin Adenine Dinucleotide	9.7(5)	9.9(5)	-2(7)
	FMN[b]	Flavin Mononucleotide	7.35(2)	7.60(3)	-3.3(5)
	RIB[b]	Ribitol	3.40(3)	3.680(1)	-7.6(7)
	RBFL[b]	Riboflavin	6.50(3)	6.77(3)	-4.0(6)
		Nicotinamide set			
	NAD^+	Nicotinamide Adenine Dinucleotide (oxidized)	9.3(1)	9.2(1)	2(2)

	NADH	Nicotinamide Adenine Dinucleotide (reduced)	9.9(2)	9.8(2)	1(2)
	$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate (oxidized)	9.9(2)	9.8(2)	1(3)
	NADPH	Nicotinamide Adenine Dinucleotide Phosphate (reduced)	11.0(1)	10.9(1)	1(2)
		Thiamine set			
	THI[d]	Thiamine	5.445(5)	5.602(8)	-2.8(2)
	TMP[d]	Thiamine Monophosphate	6.07(3)	6.16(3)	-1.6(7)
	TPP	Thiamine Pyrophosphate	6.44(5)	6.72(5)	-4(1)
Quinones		**Isoprenoid (ISO) set**			
	G8YLG	Geran-8-yl geran	6.46(4)	7.04(5)	-8.3(9)
		Menaquinone set			
	DMNFH[$]	2,3-Dimethyl-1,4-naphthalenediol	4.276	4.463	-4.190
	DMNFQ[$]	2,3-Dimethyl-1,4-naphthoquinone	4.205	4.315	-2.549
	GBZ	Geranylbenzoquinone analog	5.616(5)	6.00(5)	-6.3(8)
	MNDOL[$]	Menadiol	4.133	4.234	-2.390
	MND[$]	Menadione	4.126	4.23	-2.46
	MNQOL[$]	Menaquinol-8	14.613	14.888	-1.8470
	MNQ[$]	Menaquinone-8	14.682	15.177	-3.2620

	HQNO set			
HAQOH[$]	1,4-Dihydroxyquinolin-1-ium	4.031	4.08	-1.20
HQNOH[$]	2-Heptyl-1-Hydroxyquinolin-1-ium-4-ol	6.151	6.45	-4.64
HQNO[$]	2-Heptyl-4-quinolinol 1-oxide	6.204	6.155	0.7961
HAQO[$]	4-Hydroxyquinoline 1-oxide	3.995	3.865	3.364
	Plastoquinone set			
PQOL[$]	Plastoquinol-8	14.145	14.786	-4.3350
PQ[$]	Plastoquinone-8	14.624	15.535	-5.8640
XQH[$]	*o*-Xylene-3,6-diol	3.689	3.81	-3.18
XQ[$]	*o*-Xyloquinone	3.697	3.77	-1.94
	Ubiquinone set			
UBQ0H[$]	Ubiquinol-0	4.271	4.393	-2.778
UBQH[$]	Ubiquinol-8	14.8	15.521	-4.65
UBQ0	Ubiquinone-0	4.235(2)	4.433(2)	-4.48(7)
UBQN[$]	Ubiquinone-8	14.785	15.406	-4.0310

*$FADH_2$ SASAs are considered to be the same as FAD.

[a]These SASA values are from ATP trajectory but only taking into account the beads that are part of this moiety.

[b]These SASA values are from FAD trajectory but only taking into account the beads that are part of this moiety.

[c]These SASA values are from NADH trajectory but only taking into account the beads that are part of this moiety.

[d]These SASA values are from TPP trajectory but only taking into account the beads that are part of this moiety.

[$]SASAs without uncertainties were only retrieved using one frame since these molecules had very few or any degrees of freedom

Table S3 – Calculated and experimental log P values. Uncertainties are between parentheses and correspond to the uncertainty of the last digit.

Group	Abbreviation	Compounds	Calculated log P	Experimental log P
Nucleotides		**Adenine set**		
	ADN	Adenine	-0.47(1)	-0.09[a]
	ADOS	Adenosine	-2.38(2)	-1.05[a]
	ADP	Adenosine Diphosphate	Protonated form: -6.16(4) Deprotonated form: -7.12(4)	Protonated form: -3.1(2)[b] Deprotonated form: -3.5(3)[b]
	AMP	Adenosine Monophosphate	Protonated form: -4.48(2) Deprotonated from: -5.78(3)	Protonated form: -2.7(5)[b] Deprotonated form: -3.5(5)[b]
	ATP	Adenosine Triphosphate	Protonated form: -7.40(4) Deprotonated form: -8.19(5)	Protonated form: -3.3(1)[b] Deprotonated form: -3.7(2)[b]
	RBOS	Ribose	-2.89(1)	-2.32[a]
		Flavin set		
	$FADH_2$	Dihydroflavine Adenine Dinucleotide	-5.66(5)	----------
	FAD	Flavin Adenine Dinucleotide	-5.22(5)	-2.5(1)[b]
	FMN	Flavin Mononucleotide	Protonated form: -3.40(3) Deprotonated form: -4.45(5)	Protonated form: -3.0(7)[b] Deprotonated form: -3.0(6)[b]
	RBFL	Riboflavin	-1.42(2)	-1.5(1)[b]
		Nicotinamide set		
	NAD^+	Nicotinamide Adenine Dinucleotide (oxidized)	-6.63(5)	-3.2(1)[b]

	NADH	Nicotinamide Adenine Dinucleotide (reduced)	-5.80(5)	-3.1(5)[b]
	$NADP^+$	Nicotinamide Adenine Dinucleotide Phosphate (oxidized)	Protonated form: -8.29(6) Deprotonated form: -9.04(6)	Protonated form: -3.0(5)[b] Deprotonated form: -3.0(2)[b]
	NADPH	Nicotinamide Adenine Dinucleotide Phosphate (reduced)	Protonated form: -7.65(6) Deprotonated form: -8.93(7)	Protonated form: -3.5(5)[b] Deprotonated form: 3.2(2)[b]
		Thiamine set		
	THI	Thiamine	-0.44(2)	----------
	TMP	Thiamine Monophosphate	Protonated form: -2.58(2) Deprotonated form: -3.96(3)	----------
	TPP	Thiamine Pyrophosphate	Protonated form: -4.32(3) Deprotonated form: -5.22(4)	Protonated form: -3.1(4)[b] Deprotonated form: -3.1(4)[b]
Quinones		**Menaquinone set**		
	DMNFH	2,3-Dimethyl-1,4-naphthalenediol	1.99(1)	----------
	DMNFQ	2,3-Dimethyl-1,4-naphthoquinone	2.00(1)	----------
	MNDOL	Menadiol	1.62(1)	----------
	MND	Menadione	1.64(1)	2.2[a]
		HQNO set		
	HAQOH	1,4-Dihydroxyquinolin-1-ium	1.48(1)	----------
	HAQO	4-Hydroxyquinoline 1-oxide	1.49(1)	----------

Plastoquinone set			
XQH	*o*-Xylene-3,6-diol	1.284(9)	1.36[c]
XQ	*o*-Xyloquinone	1.317(9)	1.23[d]
Ubiquinone set			
UBQ0H	Ubiquinol-0	1.07(1)	1.05[d]
UBQ0	Ubiquinone-0	1.12(1)	0.8[d]

[a]Hansch[206].
[b]Sousa *et al*[70].
[c]Sangster[207].
[d]Rich & Harper[208].