
Elucidating the Trodusquemine Action Against Neurodegenerative Disorders by Molecular Modelling Techniques

Doctoral Dissertation submitted to the
Faculty of Informatics of the Università della Svizzera Italiana
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

presented by
Stefano Muscat

Under the supervision of
Prof. Andrea Danani (Advisor)
Prof. Jürgen Schmidhuber (Co-advisor)

June 2023

Dissertation Committee

Internal Committee Members

Prof. Rolf Krause

Prof. Andrea Emilio Rizzoli

External Committee Members

Prof. Marco Agostino Deriu

Prof. Giulia Rossi

Dissertation accepted on 22th June 2023

Research Advisor

The Student's Advisor

PhD Program Director

The PhD program Director *pro tempore*

I certify that except where due acknowledgement has been given, the work presented in this thesis is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; and the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program.

Stefano Muscat

Lugano, June 2023

Abstract

The significant increase in life expectancy worldwide has drawn scientific attention to age-related pathologies, such as Alzheimer's disease (AD). The amyloid hypothesis asserts that AD results from abnormal aggregation of specific amyloid proteins, leading to intracellular inclusions or extracellular aggregates in specific brain areas. Trodusquimine, an aminosterol isolated from the dogfish shark *Squalus acanthias*, has been found to modulate the kinetics of aggregation of the amyloid peptides responsible of AD. However, the mechanism by which aminosterols protect the neuron interacting with the cell membranes and modulate their properties remains unclear.

This PhD thesis employs computational molecular modelling to investigate trodusquimine molecular mechanism of action. Molecular dynamics (MD) simulations are used to elucidate trodusquimine interaction with the cell membrane and its self-aggregation mechanism at the molecular level. Utilizing computational methodologies, the present study provides insights into the modulation of the cell membrane physicochemical properties by trodusquimine. Consequently, this modulation confers a reduced vulnerability of the cell membranes to the deleterious effects of misfolded protein aggregates. However, computational limitations often hinder the ability to investigate biomacromolecule conformational properties. The limitation mainly results from biological system's rough energy landscapes, with many local minima separated by high-energy barriers. Therefore, MD simulations face the challenge of adequately sampling experimentally pertinent timescales ranging from milliseconds to seconds. Presently, classical MD simulations employing atomic resolution are confined to few microseconds. This work benefits from the development of a coarse-grain model to accelerate trodusquimine's dynamics and investigate supramolecular structures. Coarse-grain approach highlights the trodusquimine ability to modulate the coexistence of liquid-gel and liquid-disorder phases, influencing the lipid membrane thermodynamic stability. Furthermore, enhanced sampling techniques and dimensionality reduction methodologies are applied to estimate fundamental thermodynamic properties of aminosterol molecules. The outcomes of this research represent an essential step in designing small molecules capable of protecting neuron cell membranes from the toxicity of amyloid oligomers.

Contents

Abstract	IV
Chapter 1	1
Computational Modelling of Biomolecular Systems	1
1.1. Statistical Ensemble	2
1.2. The Molecular Mechanics Force Field.....	3
The potential energy function.....	4
1.3. Molecular Dynamics	5
Velocity Verlet integration methods	6
1.4. Coarse Grained Modelling	8
Martini Force Field.....	8
Time scale interpretation	9
Martini limits and the polarisable model.....	10
Martini Tools.....	11
1.5. Enhanced sampling methods for molecular dynamics simulations.....	11
Replica Exchange Molecular Dynamics	13
Metadynamics	14
Well-tempered Metadynamics	16
On-the-fly probability-enhanced sampling.....	18
Collective variables and dimensionality reduction algorithms	19
Chapter 2	23
Biological Background.....	23
2.1. Protein folding.....	24
2.2. Protein misfolding	24
2.3. Alzheimer's disease.....	27
Drug development against Alzheimer's disease	29
Targeting amyloid beta aggregation molecular mechanism.....	29
2.4. Aminosterols against neurodegenerative disorders.....	30
Chapter 3	35
Trodesquimine Molecular Mechanism of Action.....	35
3.1. Aim of the Work.....	36

3.2. Interaction with the Cell Membrane.....	38
Introduction	38
Material and Methods.....	39
Results and Discussion.....	42
Conclusion.....	46
3.3. Coarse-Grained models for unraveling the complexities of trodusquamine supramolecular phenomena.....	47
Introduction	47
Material and Methods.....	48
Results and Discussion.....	52
Conclusions	62
3.4. Phase Separation Triggers Trodusquamine Action	63
Introduction	63
Material and Methods.....	64
Results and Discussion.....	66
Conclusions	71
3.5. Enhanced sampling for free energy landscape of membrane insertion of aminosterols	73
Introduction	73
Materials and Methods	75
Results and Discussion.....	77
Conclusions	86
Conclusions and Future Perspectives	88
Appendix	91
A1 - Trodusquamine aggregation.....	91
A2 - Trodusquamine and squalamine bond terms optimization.	92
A3 – TICA training and validation.....	96
A4 – Aminosterols experimental binding affinity.....	97
Bibliography	99

Chapter 1

Computational Modelling of Biomolecular Systems

One of the key objectives of computer scientists is to use mathematical equations to model and understand the behavior of the real world. Computational molecular modeling is a branch of computational modelling that describes and predicts the behavior of complex biological systems, such as drugs, proteins, cell membranes, and nucleic acids. Computational molecular modelling aims to provide a realistic atomistic description of these systems, which can be used to understand and predict their macroscopic properties and to elucidate complex molecular mechanism. The field has gained significant traction in recent years due to the increasing availability of computational power and hardware resources. In this context, computational molecular modeling has emerged as an essential tool, which helps in adding valuable quantitative information to experimental data. It has demonstrated its usefulness in various fields, such as materials science and biomolecular modeling. However, due to the large number of molecules typically present in molecular systems, analytical quantification of their properties is often impossible. Numerical methods, such as Molecular Dynamics (MD), are therefore used to overcome this challenge. MD simulations leverage principles from physics, chemistry, and statistical mechanics to model the dynamics of molecular processes. Molecular modeling has a wide range of applications beyond protein structure and dynamics, folding/unfolding, multiscale modeling, and polymer chain analysis. Despite being based on simple underlying concepts like Newton's second law, it proves to be a powerful method for investigating complex molecular systems.

This chapter provides the theoretical foundation for the present PhD thesis work, aiming to clarify the physical principles underpinning the computational method. It begins with a concise introduction to statistical mechanics in Section 1.1, followed by detailed presentations of the force field description and the MD algorithm in Sections 1.2 and 1.3, respectively. To address the limitations of classical MD, the chapter also explores multiscale descriptions and advanced sampling techniques. Section 1.4 delves into coarse-grained modeling, while Section 1.5 discusses enhanced sampling methods and dimensionality reduction methodologies.

1.1. Statistical Ensemble

To investigate and predict the macroscopic properties of a complex biological system it is necessary to average the microscopic states of a statistical ensemble. A *statistical ensemble* is characterized by a collection of distinct microscopic states that share an identical thermodynamic state, thereby providing a comprehensive representation of the system behavior under a given set of conditions. This ensemble defines all accessible physical states by the molecular system, also called *phase space*. Within this phase space, a micro-state for a system comprising N particles in three dimensions is defined by $6N$ degrees of freedom, accounting for the spatial and dynamic properties of each particle. Specifically, three of these degrees of freedom correspond to the coordinates of the particle position in space, while the remaining three degrees of freedom represent the components of the particle momentum. Each possible state corresponds to one unique point in the phase space.

Table 1: The Micro-Canonical ensemble (NVE) with the number of particles (N), the volume (V) and the energy (E) are fixed and therefore comparable as an isolated system. The Canonical Ensemble (NVT) with number of particles, volume and temperature (T) fixed and therefore comparable a closed system. The Grand Canonical Ensemble (μVT) with chemical potential (μ), volume and temperature fixed and therefore comparable an open system. The Isobaric-Isothermal Ensemble (NPT) characterized by constant number of particles, pressure (P) and temperature.

Ensembles	Fixed variables
Micro-Canonical	N, V, E
Canonical	N, V, T
Isobaric-isothermal	N, p, T
Grand canonical or Gibbs	μ, V, T

Various methodologies have been developed to generate a representative statistical ensemble that effectively captures the behavior and properties of complex molecular systems. Among these techniques, Monte Carlo (MC) simulation and Molecular Dynamics (MD) simulation are the most prevalent and widely employed approaches in the field of computational biophysics and chemistry.

From the phase space it is possible to calculate a macroscopic property as average on ensemble as follow:

$$\langle A \rangle_{Ensemble} = \iint A(p^N, r^N) \rho(p^N, r^N) dp^N dr^N \quad (1.1)$$

where $A(p^N, r^N)$ is the parameter of interest, p the momenta and r the atomic position. The variable ρ is the ensemble probability density:

$$\rho(p^N, r^N) = \frac{1}{Q} e^{[-\beta H(p^N, r^N)]} \quad (1.2)$$

where $\beta = 1/k_B T$ and k_B is the Boltzmann factor, T the system temperature, H the Hamiltonian and Q is the *partition function*. The partition function indicates the normalize sum of Boltzmann factor of all microstates and is defined as:

$$Q = \iint e^{[-\beta H(p^N, r^N)]} dp^N dr^N \quad (1.3)$$

This equation is extremely important because it allows to derive macroscopic thermodynamic properties. However, considering all possible system states, the partition function is very difficult to calculate. To overcome this problem, the *ergodic hypothesis* is used: over long periods of time, the time-average of a certain physical property, represents the ensemble-average of the same property. The basic idea is to allow the system to evolve in time indefinitely, so that it will eventually pass through all its possible states. In conclusion, overcoming the computational challenges associated with calculating the partition function necessitates the implementation of robust sampling methods, such as MC and MD simulations, and the application of the ergodic hypothesis. However, since simulation time is limited, it is crucial to adopt the most effective method to sample a sufficient number of allowable states.

1.2. The Molecular Mechanics Force Field

In molecular dynamics simulations, defining an appropriate potential energy function is a prerequisite for solving Newton's equations of motion and accurately modeling the dynamic behavior of molecular systems. The Molecular Mechanics (MM) method uses Newtonian mechanics to model molecular systems, neglecting the electronic motions and analyzing the system as a set of atoms interacting through a potential energy function. The core of the MM

approach is the set of the equation and parameters used to describe the potential energy function V of a molecular system, also known as *force field*. In this context, numerous force fields have been developed to model a wide range of molecular systems, from small organic molecules to large biomacromolecules like proteins and cell membrane components.

The potential energy function

The force field allows to compute the potential energy of the molecular system in a given conformation as a sum of bonded and non-bonded individual energy contribution. Each term can be modelled in a different way, depending on the particular simulation settings being used¹. The potential energy function can be described as:

$$\begin{aligned}
 V(r_1, r_2, \dots, r_n) = & \sum_{bonds} \frac{1}{2} k_l [l - l_0]^2 + \sum_{angles} \frac{1}{2} k_\theta [\theta - \theta_0]^2 \\
 & + \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{improper\ dihedrals} \frac{1}{2} k_\zeta [\zeta]^2 \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left(\frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{ij}} + 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}^{12}}{r_{i,j}} \right) - \left(\frac{\sigma_{i,j}^6}{r_{i,j}} \right) \right] \right)
 \end{aligned} \tag{1.4}$$

The *force field* equation consists of two distinct components: the set of equations used to generate the potential energies (also known as the potential function), and the parameters used in this set of equations. The *force field* parameters are usually derived empirically or by means of quantistic approach.

The first term of the eq. 1.4 models the interaction between bonded atoms, described with a harmonic potential inducing potential energy increase when the bond length departs from the reference value l_0 . The second term describes the angle among three atoms again modelled by using a harmonic potential. The third contribution is a torsional potential that describes bond rotates, and the fourth is an additional torsional that includes an out-of-plane potential.

The last term models the non-bond interactions that represent a very important component of the MD force field, being the dominant type of interaction between molecules, critical in maintaining the three-dimensional structure of proteins and nucleic acids. The non-bond terms

are usually modelled as a function of an inverse power of the distance. The non-bond interactions generally consist of electrostatic and Van Der Waals components. The electrostatic interactions, which arise due to the presence of charged particles within the molecular system, are usually described by using the Coulomb's law. The *Van Der Waals* forces represent the sum of the attractive or repulsive forces between molecules (or between parts of the same molecule), caused by correlations in the fluctuating polarizations of nearby particles. The *Van Der Waals* interactions are relatively weak compared to covalent bonds but play a fundamental role in defining many properties of organic compounds, including their solubility in polar and non-polar media.

1.3. Molecular Dynamics

Among all computational methods, Molecular Dynamics (MD) is a powerful tool to investigate biomolecular systems. MD simulations are based on the solution of Newton's laws of motion for each atom of a molecular system, providing detailed insights into the dynamic behavior of molecules and generating molecular trajectories. The obtained molecular trajectories can be used to estimate the kinetic and thermodynamic properties of interest. This computational method relies on the principles of classical mechanics where the future state of the system is completely determined by its present state. For a system consisting of N atoms, MD simulations solve the Newton's law of motion considering the r_i position and the mass m_i to obtain the force F_i acting on atom i -th as follow:

$$m_i \frac{d^2 r_i}{dt^2} = F_i \quad (1.5)$$

The force acting on i -th atom can be written according to the potential energy U , a function of the atoms position:

$$F_i = -\nabla_{r_i} U(r_1, \dots, r_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i} \right) \quad (1.6)$$

From this equation, the acceleration can be obtained as the derivative of potential energy with respect to the position r :

$$m_i a_i = m_i \frac{dv_i}{dt} = m_i \frac{d^2 r}{dt^2} = -\nabla_{r_i} U(r_1, \dots, r_N) = F_i \quad (1.7)$$

where v_i is the velocity of the i -th particles.

To solve the aforementioned equation through integration, initial positions and velocities are required. Information regarding initial positions can be obtained from experimental data, commonly derived from techniques such as nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography. For instance, the Protein Data Bank (PDB) database provides a wealth of structural information for a wide range of biomolecules. In contrast, initial velocities must be generated for the simulation. This can be achieved by randomly extracting them from a Maxwell-Boltzmann or Gaussian distribution, selecting a temperature to examine the system of interest, which in turn determines the probability that an atom i possesses a velocity v_x in the x direction at a temperature T . By assigning initial velocities according to this distribution, the simulation starts from a thermodynamically consistent state, allowing it to explore the system's behavior under the desired conditions. The numerical resolution scheme of these equations goes on for several steps until the equilibrium and the convergence of the computed equilibrium property are reached. This iterative process ensures the accurately sampling of the dynamic behavior of the biomolecular system, allowing to compute kinetic and thermodynamic properties, as well as to explore conformational changes, binding events, and other molecular interactions.

Velocity Verlet integration methods

There are several algorithms for integrating the equations of motion using finite difference methods in MD calculations. The *velocity Verlet* method² is probably the most widely used method for integrating the equations of motion in a MD simulation and provides positions, velocities and accelerations at the same time and does not affect precision:

$$r(t + \partial t) = r(t) + \partial t v(t) + \frac{1}{2} \partial t^2 a(t) \quad (1.8)$$

$$v(t + \partial t) = v(t) + \partial t [a(t) + a(t + \partial t)] \quad (1.9)$$

This method is implemented as a three-steps process because the calculation of new velocities requires the accelerations at both t and $(t + \partial t)$. Thus, in the first stage the positions at $(t + \partial t)$ are computed using the velocities and the accelerations at time t . The velocities at time $(t + \frac{1}{2} \partial t)$ are then estimated applying:

$$v\left(t + \frac{1}{2}\partial t\right) = v(t) + \frac{1}{2}\partial t[a(t)] \quad (1.10)$$

New forces are then calculated from the current positions, thus providing $a(t + \partial t)$. In the last stage, the velocities at time $(t + \partial t)$ are determined using:

$$v(t + \partial t) = v\left(t + \frac{1}{2}\partial t\right) + \frac{1}{2}\partial t[a(t + \partial t)] \quad (1.11)$$

The resolution of these equations is done by numerical integration. Starting from the atomic positions r_i , the potential energy U is calculated; the scheme allows to calculate the forces F_i acting on each atom, by deriving the potential energy function; then the integration of the equation of motion leads to the calculation of new position r_i and velocities v_i . The cycle goes on for several steps until the equilibrium and the convergence of the computed equilibrium property are reached.

Selecting an appropriate time step for molecular dynamics simulations is crucial for ensuring both computational efficiency and simulation stability. A large time step can enable the simulation to progress more rapidly, covering a broader range of molecular configurations in a shorter period. However, employing an excessively large time step may lead to instability in the simulation, causing erroneous behavior or even failure of the simulation to converge. Therefore, it is essential to choose a time step that balances the need for efficient sampling of the phase space with the requirement for maintaining the stability of the simulated system. This often involves a careful assessment of the system dynamics and the relevant time scales of the molecular processes being investigated. The chosen time step should be small enough to accurately capture the fastest atomic motions, such as bond vibrations, while not being so small as to unnecessarily increase the computational cost of the simulation. From a practical point of view, an appropriate time step should be less than $1/10^{\text{th}}$ the period of the fastest harmonic oscillator. Table 2 summarize the main time steps employed in MD simulations depending on the level of detail and the specific properties of the molecular system under investigation.

Table 2: Schematic summary of the time steps to be used depending on the type of system and motions.

Graining	Motions Type	Time Step
Coarse Grain	Translation	> 10 fs

Rigid Molecules	Translation, rotation	5 fs
Flexible Molecules/Rigid Bond	Translation, rotation, torsion	2 fs
Flexible Molecules/Bond	Translation, rotation, torsion, vibration	0.5 – 1fs

1.4. Coarse Grained Modelling

Even though the all-atom (AA) MD approach faithfully recreates physical, chemical, and thermodynamic properties of biological systems by representing all atoms, AA representation is limited by its characteristic length and time scales, and by the computing power³. In this context, Coarse Grained (CG) modelling can use simplified mathematics allowing mechanics, classical thermodynamics and kinetic theories to be considered with a large number of degrees of freedom removed⁴. The resulting energy landscape of coarse grained forcefields is smoother than their all-atom counterparts (Figure 1), and thereby helps to avoid local energy minima traps⁵.

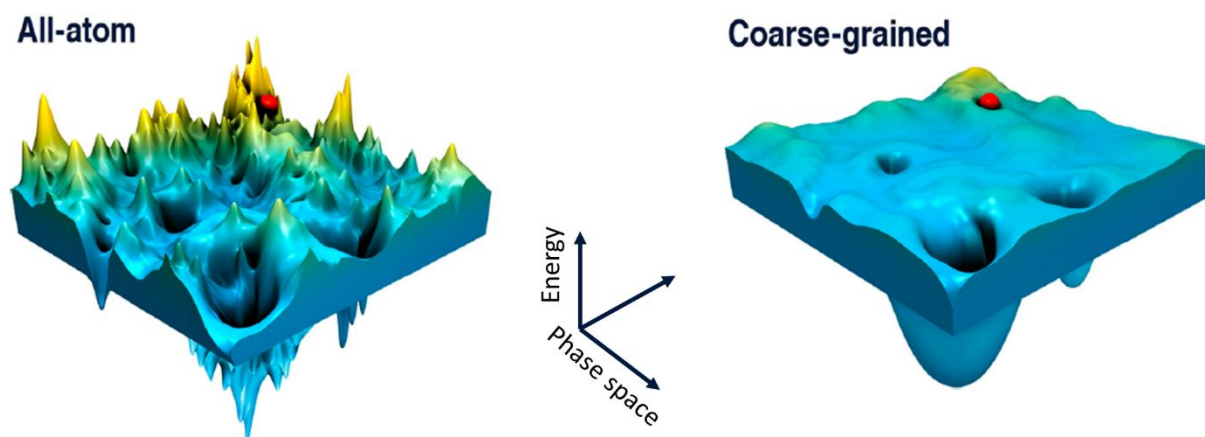


Figure 1: Example of free energy profiles using all-atom (left) and coarse-grained (right) representations. The coarse-grained profile exhibits a smoother and more simplified landscape compared to the highly detailed all-atom profile.

Martini Force Field

The Martini force field is one of the most widely used CG force fields in molecular simulations. Initially developed by the groups of Marrink and Tieleman, the Martini force field was first designed for simulating lipid systems. It has since been extended to cover a wide range of

biomolecules, including proteins⁶, carbohydrates⁷, polymers⁸ and DNA⁹. The idea behind the Martini methodology was to create an extendable CG model based on simple modular building blocks without capture every single detail of a given atomistic simulation, using few parameters and standard interaction potentials to maximize applicability and transferability¹⁰.

Instead of focusing on the accurate emulation of conformational details at a selected state for a specific system, they aimed for a wider range of applications without the necessity to re-parameterize the model each time. Martini methodology follows a top-down approach by extensive calibration of the non-bonded interactions of the chemical building blocks based on experimental data. A specific thermodynamic parameter, the free energy (ΔG), was used as the reference data set to calibrate CG blocks by computing oil/water partitioning coefficients¹⁰.

The Martini model is based on a four-to-one mapping, i.e., on average four heavy atoms plus associated hydrogens are represented by a single interaction centre. By reducing the level of detail in the molecular representation, the Martini model significantly decreases the number of interaction sites and hence the interactions to be calculated, leading to a substantial reduction in computational costs. The four-to-one mapping was chosen as an optimum between computational efficiency on the one hand and chemical representability on the other hand¹¹. In order to keep the model as simple as possible, four main types of interaction sites are developed: polar (P), nonpolar (N), apolar (C), and charged (Q)¹⁰. Each particle type has a number of subtypes, which allows more accurate representation of the chemical nature of the underlying atomic structure. In total, 18 different types of interacting particles are generated. Particles subtypes are furthermore distinguished by division in hydrogen bonding capabilities (“d” indicates donor, “a” indicates acceptor, “da” for both, 0 for no-one), or by a number which indicates the polarity degree (from 1, low polarity, to 5, high polarity).

Moreover, for preserving purposes of the small ring compounds geometry, a four-to-one mapping procedure is not the best choice. The strategy adopted to model rings is typically composed by 2 or 3 atoms into one CG beads mapping¹⁰. With this more detailed mapping, enough geometrical detail is kept to mimic the geometry of small compounds such as cyclohexane or benzene and sterol bodies such as cholesterol.

Time scale interpretation

In general, the interpretation of the simulation time scale in CG modelling is not straightforward. Typically, in comparison to its AA counterpart, the dynamics observed with

CG models is faster. The main reason is that the underlying energy landscape is much smoother (see also Figure 1) as a result of the larger particle sizes. Consequently, the time scale in CG simulations does not directly correspond to real time, requiring a conversion factor to bridge the gap. Based on comparison of diffusion constants in CG and atomistic modelled systems, the effective time sampled using Martini CG model was found to be 2- to 10-fold larger. The standard conversion factor used is a factor of 4, which is the speed up factor in the diffusional dynamics of CG water compared to real water.

Martini limits and the polarisable model

As the forcefield Martini grew in popularity and usage, its limits began to become apparent.

These limitations stem primarily from the inherent simplifications associated with the CG process. Some of the most significant limitations include:

1. Entropy/enthalpy compensation influenced by temperature: CG techniques can struggle to accurately capture the delicate balance between entropy and enthalpy at varying temperatures. This limitation can result in inaccuracies when predicting thermodynamic properties, phase transitions, and temperature-dependent behaviors of biomolecules ¹¹⁻¹³.
2. Absence of hydrogen bond directionality: the Martini model simplifies atomic interactions and does not explicitly represent hydrogen bonds or their directionality. The hydrogen bonds omission can significantly impact protein folding studies, as hydrogen bonding plays an essential role in stabilizing secondary structures like alpha-helices and beta-sheets ¹¹⁻¹³.
3. Precision in electrostatic interactions: the Martini force field simplifies the description of electrostatic interactions, limiting its ability to capture the exact nature of these interactions in biomolecular systems. This limitation can be particularly problematic when studying systems where electrostatic interactions are central, such as in protein-protein interactions, or protein-nucleic acid interactions ¹¹⁻¹³.

The primary cause of the last limitation was that water molecules were represented by non-polarizable particles, which were incapable of representing the electrostatic shielding of the actual solvent. This shielding is implicitly described as a function of distance in Martini's standard water model utilizing a dielectric constant of 15 and a shifted electrostatic potential.

To further improve the performance of the Martini force field, several enhancements have been developed. In this context, the multipolar and polarizable variants of the Martini water model^{14,15}, which offer more accurate descriptions of water-water and water-solute interactions. Polarizable ions have also been introduced to better capture ion solvation effects¹⁶. A polarized version of the protein force field has also been parameterized to capture the interactions between charged amino acids more accurately¹⁷. Moreover, polarized Martini beads have been used in various models; in particular, they have shown great promise for stabilizing supramolecular polymers. Finally, in Martini 2.3, Cation-Pi interactions were added, and processing speed was improved through the optimization of electrostatic cutoffs¹³.

Martini Tools

To help with the process of creating, executing, and evaluating MD simulations using Martini force field, a broad variety of tools were built. Automatic topology builder tools are among them that in principle are able to convert a simple smiles string into a complete Martini topology file¹³. Some of these, like the pyCGtool¹⁸ and Swarm-CG¹⁹, are designed to generate bonded potentials, while others, like autoMartini²⁰ and the graph-based cg_param method²¹, are designed to address the difficulty of mapping an underlying chemical structure to its CG Martini representation.

1.5. Enhanced sampling methods for molecular dynamics simulations

Nowadays, MD simulations are increasingly employed in conjunction with experimental techniques to gain a more comprehensive understanding of molecular systems at the atomistic level. Using small integration steps on the order of femtoseconds, MD simulations solve equations of motion numerically and provide atomistic equivalents to macroscopic descriptions. As well as monitoring the time-evolution of a system as a result of perturbations, they are also frequently used for the collection of statistical ensembles. However, in atomistic simulations, sampling is a hurdle. The complexity and size of the investigated systems, coupled with the extensive configurational space that demands exploration, make it arduous to sufficiently sample pertinent regions of phase space within feasible computational timeframes.

In this context, there are generally three main limitations associated with MD simulations: force field accuracy, presence of high energy barriers, and inability to capture phenomena occurring over biologically relevant time scales.

Even the most precise models are not able to deliver high-quality results without sufficient phase space sampling. The sampling difficulty is an ergodicity problem, caused by the gap between macroscopic and atomistic time scales. Furthermore, biological systems exist at different metastable states each of which is separated by kinetic bottlenecks that make transitioning between them extremely difficult. This complex landscape further exacerbates the difficulty in achieving comprehensive sampling, thereby posing a substantial challenge for MD simulations.

Given the Boltzmann configuration distribution in Eq. 1.12 where k_B is the Boltzmann constant, there are three strategies to enhance the sampling of high-energy conformations²², crossing energies barriers:

1. By increasing the temperature T .
2. By modifying the force field potential $U(R)$.
3. By adding an external bias potential $V(CV(R))$.

$$P(R) \propto \exp\left[-\frac{U(R) + V(CV(R))}{k_B T}\right] \quad (1.12)$$

Enhanced sampling can solve this problem establishing an ad hoc modified ensemble to increase the probability of sampling unusual events. Adaptive Biasing Potential (ABP) methods incorporate an external bias potential $V(CV(R))$ into a space of collective variables (CVs) to model bias system dynamics. Adding an external bias potential allows high barriers on the free energy surface to be overcome for flatter CV sampling, leading to an easier-to-sample biased CV-distribution. Several ABP approaches exist, each differing in the construction of bias potential and the convergence of sampling. Among the most commonly used ABP methods are self-healing umbrella sampling²³, adaptive biasing MD (ABMD)²⁴, the metadynamics²⁵, and on-the-fly probability-enhanced sampling (OPES)²⁶. These methods provide diverse strategies to improve sampling efficiency, thereby facilitating a more comprehensive understanding of complex molecular systems. However, the implementation of enhanced sampling techniques in complex molecular systems using a limited set of physically derived CVs is challenging²⁷. Developing CVs involves finding a low-dimensional representation from high-dimensional configurations, similar to dimensionality reduction algorithms. Data-driven strategies and signal analysis methods have been suggested for CV generation, such as principal component analysis (PCA) and time-lagged independent component analysis (TICA). Furthermore, these

techniques can be employed when metastable states, like folded and unfolded states of a peptide or reactants and products of a reaction, are known beforehand.

Replica Exchange Molecular Dynamics

In 1999 Replica Exchange Molecular Dynamics (REMD) was introduced as enhanced sampling method²⁸. Basically, the concept of REMD is to create N replicas of the original system at different temperatures. This temperature-dependent approach is essential because high-temperature systems are generally able to sample a vast volume of phase space, whereas low-temperature systems frequently become trapped in local energy minima due to their restricted sampling within a localized region of phase space.

Through the exchange of complete configurations between the systems at different temperatures, REMD achieves good sampling. Therefore, the inclusion of higher temperature systems ensures that lower temperature systems can access a representative set of low temperature regions of phase space. Essentially, one constructs a set of replicas equilibrated at their own temperatures, initiates a simulation for each replica, and allows their conformations to exchange with an appropriate rate (Figure 2).

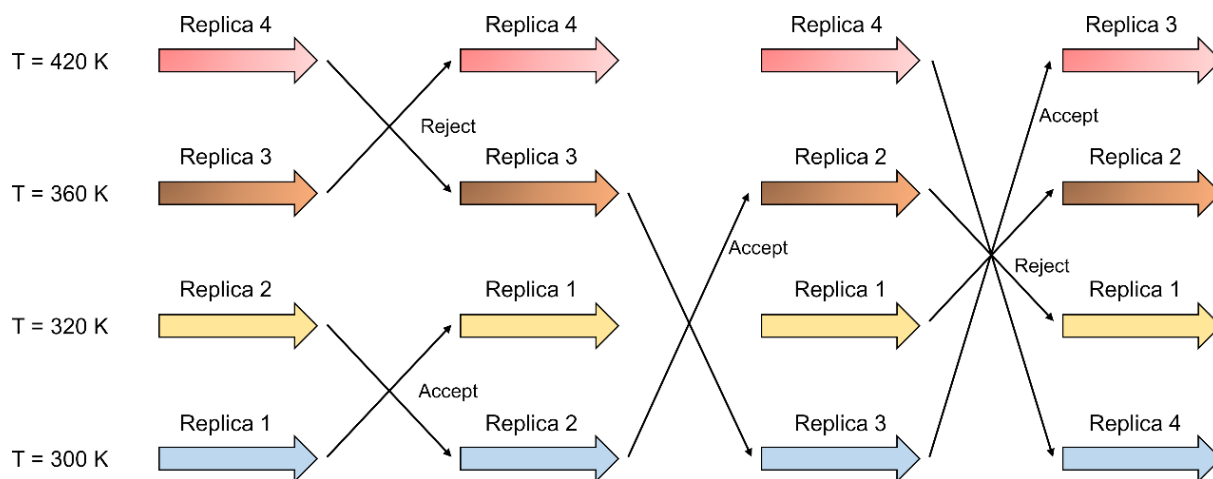


Figure 2. Schematic representation of REMD simulation and temperature exchange.

The transition probability from state X to X' in the REMD process is given by the Metropolis criterion:

$$w(X \rightarrow X') = \begin{cases} 1 & \text{for } \Delta \leq 0 \\ e^{-\Delta} & \text{for } \Delta > 0 \end{cases} \quad (1.13)$$

With

$$\Delta = (\beta_m - \beta_n)\{E(q^j) - E(q^i)\} \quad (1.14)$$

where E is the potential energy, q are the positions of atoms, m and n are the temperature indexes, i and j are the replica indexes, β is the inverse temperature defined by $\beta = 1/k_b T$. After the replica exchange, atomic momenta are rescaled as follows:

$$p^{i'} = \sqrt{\frac{T_n}{T_m}} p^i \quad p^{j'} = \sqrt{\frac{T_m}{T_n}} p^j \quad (1.15)$$

where p are the momenta of atoms. The aforementioned equations demonstrate that the probability of swaps between adjacent replicas is dependent on the difference in potential energy, which necessitates an adequate overlap in their potential distribution. Failure to achieve this overlap may result in kinetic blockage of the system, hindering the exploration of phase space. Consequently, the selection of temperatures becomes a critical factor; the highest temperature must be sufficiently elevated to prevent any replicas from becoming trapped in local energy minima. Additionally, the number of replicas must be large enough to ensure that swapping occurs between all adjacent replicas. Despite its apparent simplicity, REMD has emerged as a widely used technique for investigating complex biomolecular systems. It has proven particularly valuable in the study of protein folding, where it enables the exploration of the complete conformational space of small peptides, as well as the sampling of other significant conformational changes. The effectiveness of REMD in enhancing the sampling of phase space has made it an indispensable tool in the field of computational chemistry and biophysics, contributing to our understanding of the structural and dynamical properties of various molecular systems ²².

Metadynamics

Metadynamics is an enhanced sampling method that employs an additional bias potential or force, acting on a carefully chosen set of degrees of freedom, commonly referred to as collective variables (CVs). This enhanced sampling method introduces an external history-dependent bias potential, which is a function of the CVs, into the Hamiltonian of the molecular system being

studied. The bias potential is expressed as a sum of Gaussians deposited along the system trajectory in the CVs space, with the aim of preventing the system from revisiting previously sampled configurations.

Let S be a set of d functions of the microscopic coordinates R of the system:

$$S(R) = (S_1(R), \dots, S_d(R)) \quad (1.16)$$

At the time t , the metadynamics potential can be written as:

$$V_G(S, t) = \int_0^t dt' \omega \exp \left(- \sum_{i=1}^d \frac{(S_i(R) - S_i(R(t')))^2}{2\sigma_i^2} \right) \quad (1.17)$$

where ω is an energy rate and σ_i is the width of the Gaussian for the i th CV.

The energy rate is constant and usually expressed in terms of a Gaussian height W and a deposition stride τ_G :

$$\omega = \frac{W}{\tau_G} \quad (1.18)$$

Metadynamics has several advantages:

- (i) It accelerates the sampling of rare events by pushing the system away from local free- energy minima.
- (ii) It allows exploring new reaction pathways as the system tends to escape the minima passing through the lowest free-energy saddle point.
- (iii) No *a priori* knowledge of the landscape is required.

After a transient, the bias potential V_G provides an unbiased estimate of the underlying free energy:

$$V_G(S, t \rightarrow \infty) = -F(S) + C \quad (1.19)$$

where C is an irrelevant additive constant, and the free energy $F(S)$ is defined as

$$F(S) = -\frac{1}{\beta} \ln \left(\int dR \delta(S - S(R)) e^{-\beta U(R)} \right) \quad (1.20)$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T the temperature of the system, and $U(R)$ the potential energy function.

However, there are some disadvantages of this method:

- (i) The free energy landscape does not converge to a definite value but fluctuates around the correct result, leading to an average error which is proportional to the square root of the bias potential deposition rate. Consequently, determining when to stop the simulation can be challenging. Ideally, the simulation should be terminated when the motion of the CVs becomes diffusive in the region of interest.
- (ii) The difficulty in identifying a suitable set of CVs for describing complex processes. Choosing appropriate CVs is crucial for accurately capturing the system behavior and driving the sampling process, but this task can be far from trivial and represents a significant challenge for computer simulations. Identifying an optimal set of CVs often requires a deep understanding of the system being studied and its underlying biophysical or chemical properties.

To address the first issues, the well-tempered metadynamics was developed, as described in the following.

Well-tempered Metadynamics

In the well-tempered metadynamics approach²⁹, the bias deposition rate is modified to decrease over the course of the simulation. This modification aims to control the exploration of regions in the free energy surface (FES) that are physically meaningful and relevant to the system under investigation.

This is achieved by applying an history-dependent potential:

$$V(S, t) = \Delta T \ln \left(1 + \frac{\omega N(S, t)}{k_b \Delta T} \right) \quad (1.21)$$

Where the bias potential V is defined as a monotonic function of the histogram $N(S, t)$, which represents the frequency of the sampled states in the collective variable space S during the

simulation. The parameter ω , with dimensions of an energy rate, and ΔT , an input parameter with the same dimensions as temperature, both contribute to the formulation of the bias potential.

It is worth mentioning that $V(s, t)$ changes with the rate of:

$$\dot{V}(S, t) = \frac{\omega \Delta T \delta_{s,s(t)}}{\Delta T + \omega N(S, t)} = \omega e^{-[V(S, t)/\Delta T]} \delta_{s,s(t)} \quad (1.22)$$

The connection with metadynamics is evident if we examine the previous equation and replace $\delta_{s,s(t)}$ with a finite width Gaussian. The height of each Gaussian is determined by $w = \omega e^{-[V(S, t)/\Delta T]} \tau_G$, where τ_G is the time interval at which Gaussians are deposited. Thus, ω represents the initial bias deposition rate. It is worth highlighting that while the bias deposition rate decreases as $1/t$, the dynamics of all the microscopic variables becomes progressively closer to thermodynamic equilibrium as the simulation proceeds. The bias potential converges to:

$$V(S, t \rightarrow \infty) = \frac{-\Delta T}{T + \Delta T} F + C \quad (1.23)$$

Where C is an immaterial constant. It is worth highlighting that in the long-time limit, the probability distribution of the CVs becomes:

$$P(S) \propto e^{\frac{-F(S)}{k_b(T+\Delta T)}} \quad (1.24)$$

While ordinary MD corresponds to the limit $\Delta T \rightarrow 0$, the standard metadynamics is obtained for $\Delta T \rightarrow \infty$. Fine tuning of ΔT allows to regulate the extent of FES exploration. This proves to be a useful procedure for preventing overfilling of the free energy landscape and saving computational time when numerous CVs are employed. The introduction of a history-dependent potential modifies the probability distribution sampled throughout the simulation. In the past, various approaches have been developed to reweight a metadynamics simulation and reconstruct the unbiased distribution for variables other than the CVs, assuming an adiabatic evolution for the bias potential^{30,31}.

Nonetheless, if the assumption of exploring the phase space along selected degrees of freedom does not hold, convergence of the bias potential may be rendered impossible due to hysteresis effects. This highlights the importance of carefully choosing appropriate collective variables

and ensuring that the selected degrees of freedom adequately represent the relevant processes under investigation.

In conclusion, well-tempered metadynamics solves the convergence problems of metadynamics and allows the computational effort to be focused on the physically relevant regions of the conformational space³². The latter property makes it possible to use adaptive bias methods in higher dimensionality cases, thus paving the way for the study of complex systems where it is difficult to select a priori a very small number of relevant degrees of freedom.

On-the-fly probability-enhanced sampling

The On-the-fly Probability Enhanced Sampling (OPES) method represents an advancement of metadynamics, integrating some concepts from the Variationally Enhanced Sampling approach. OPES is specifically designed for ease of use and robustness, even when dealing with suboptimal collective variables. In comparison to metadynamics, OPES exhibits faster convergence towards a quasi-static bias and manages multi-dimensional collective variables with greater efficiency²².

As for metadynamics the OPES method is based on CVs. In general, CVs are function of the microscopic configuration $s = s(x)$. In OPES, the sampled distribution deviates from the Boltzmann distribution and instead conforms to a preselected target distribution that is easier to sample. This approach effectively eliminates kinetic bottlenecks between relevant metastable states, enabling more efficient exploration of the phase space. In details, the physical probability distribution of s ($P(s)$), is adjusted to reach a given target distribution $p^{tg}(s)$, adding a bias potential $V(s)$ defined as:

$$V(s) = -\frac{1}{\beta} \log \left(\frac{p^{tg}(s)}{P(s)} \right) \quad (1.25)$$

where β is the inverse temperature.

At convergence the relationship between the bias and the free energy is:

$$V(s) = -\left(1 - \frac{1}{\gamma}\right) F(s) \quad (1.26)$$

Adding the bias potential, the free-energy barriers along the CVs will be reduced by a factor γ , increasing the system's transition rate between metastable states. Iteratively in each step n , the

probability distribution $P_n(s)$ is estimated, reweighting on the fly the MD simulation which is biased with $V_n(s)$, previously constructed on $P_{n-1}(s)$ following the equation 1.25. During the MD simulations, through a weighted kernel density estimation (KDE) a new kernel is added:

$$P_n(s) = \frac{\sum_k^n w_k G(s, s_k)}{\sum_k^n w_k} \quad (1.27)$$

Where w_k are the weights, defined as $w_k = e^{\beta V_{k-1}(s_k)}$.

During the MD simulations, the FES estimation change, allowing to quickly obtain a coarse-grained estimation, that slowly converge.

Collective variables and dimensionality reduction algorithms

Choosing appropriate collective variables (CVs) is a critical aspect of enhanced sampling methods. CVs are functions of the atomic coordinates R (see eq. 1.16) and are intended to represent the slow modes of the system, reducing the complexity of the molecular system. In the context of enhanced sampling methods, CVs directly influence the reconstructed FES, as they capture the essential features of the system and should ideally provide a meaningful representation of the system's conformational landscape.

In general, the FES is defined from its partition function $Q(s)$ (see eq. 1.3) as:

$$FES(s) = -kT \log Q(s) + \tilde{C} \quad (1.28)$$

where \tilde{C} is also a constant independent of s . Since we are interested in a free energy difference between two conformations, \tilde{C} is ignored in most cases.

Within this framework, applying enhanced sampling methods to investigate intricate molecular systems with only a limited number of physically derived CVs presents significant challenges. Effectively sampling a complex system may require up to hundreds of CVs²⁷. The design of CVs, which seeks to uncover a low-dimensional representation from high-dimensional configurations, aligns with the goals of dimensionality reduction algorithms. These techniques rely on the assumption that data points exist near a low-dimensional manifold, even when situated in a high-dimensional space.

For this reason, a variety of data-driven methodologies and signal analysis methods have been presented as potential avenues for CV creation³³. When the metastable states that are involved

in the uncommon event are known in advance, such as the folded and unfolded states of a peptide or the reactants and products of a reaction, certain of these approaches can be utilized.

Principal Components Analysis

Principal Components Analysis (PCA) is a powerful approach to determine a set of equations of motion for a small number of modes corresponding to the largest eigenvalues which provide a reduced representation of the long-time dynamics of a specific molecular event. After removing the translational and rotational motions from the system trajectory composed by N atoms, the covariance matrix (C) of $3N$ cartesian coordinates can be computed as follows:

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \quad (1.29)$$

Prior to calculating C , the simulation trajectory is superimposed onto a reference structure to eliminate translation and rotation of the entire biomolecular system. Diagonalization of C results in a set of $3N$ orthonormal eigenvectors (e_j) with corresponding eigenvalues (σ_j^2). These eigenvectors are arranged in descending order of eigenvalues and are referred to as PCA vectors.

The projection $p_j(t)$, known as the j -th principal component (PC), represents the position of the molecular system along the j -th PCA vector, as illustrated below:

$$p_j(t) = [x(t) - \langle x \rangle] \cdot e_j \quad (1.30)$$

The protein motion along a simulation trajectory, quantified as mean square fluctuation (MSF), can be decomposed into contributions from different PCs:

$$\langle (x - \langle x \rangle)^2 \rangle = \sum_{j=1}^{3N} \text{var}(p_j) = \sum_{i=1}^{3N} \sigma_i^2 \quad (1.31)$$

Typically, the first PCs account for a substantial portion of the atomic MSF. Considering the protein folding event, if the molecular event under examination involves considerable protein movements, the initial few principal components serve as a reasonable basis set for analyzing the biomolecular simulation and hence could be used as CV during enhanced sampling simulations. However, PCA drawbacks are evident. The PCs are a linear combination of system coordinates. Nonetheless, relevant molecular events are nonlinear. As linear PCA is unable to

accurately model a nonlinear phenomenon, a nonlinear dimensionality reduction technique is essential for constructing the CVs.

Time-lagged Independent Component Analysis

Time-Lagged Independent Component Analysis (TICA) is a linear transformation method. In contrast to PCA, which finds coordinates of maximal variance, TICA finds coordinates of maximal autocorrelation at the given lag time, specifically tailored for obtaining CVs in molecular systems. The driving principle behind TICA is rooted in the idea that optimal CVs ought to represent the slow modes of a molecular system, as these modes exhibit correlation functions that decay slowly over time.

More in detail, TICA determines those “slowest” independent collective degrees of freedom v_k onto which the projections $y_k(t) = y_k \cdot x(t)$ have the largest time-autocorrelation

$$\frac{\langle y_k(t)y_k(t + \tau) \rangle_t}{\langle y_k(t)^2 \rangle_t} \quad (1.32)$$

where τ is a chosen lag time.

Equivalently, using the matrix notation:

$$C(\tau) = (\langle x_i(t)x_j(t + \tau) \rangle_t)_{ij} \quad (1.33)$$

Where each degree of freedom v_k maximizes

$$\frac{v_k^T C(\tau) v_k}{v_k^T C(0) v_k} \quad (1.34)$$

Subject to the constraint that it is orthogonal to all preceding degrees of freedom, the v_k solutions are derived from the generalized eigenvalue problem.

$$C(\tau)v_k = \lambda_k C(0)v_k \quad (1.35)$$

Recently, the effectiveness of linear methods in identifying CVs has been significantly enhanced by incorporating Neural Networks (NNs), which capitalize on their capacity to approximate non-linear functions of multiple variables. This integration has resulted in the development of highly efficient CVs, as demonstrated by innovative approaches like the Reweighted Autoencoded Variational Bias (RAVE)³⁴ and Deep-TICA³⁵. These methods

harness the power of deep learning to better capture the complex relationships and dynamics within molecular systems, leading to improved sampling performance and a more accurate representation of the underlying free energy landscapes ²².

A line of attack in these cases has been to collect a number of configurations from short unbiased MD runs in the different metastable states and use these data to train a supervised classification algorithm. Furthermore, the interface between the NN and MD code is currently not optimized, leading to significant degradation in computational performance.

Chapter 2

Biological Background

Proteins are essential for a multitude of physiological functions and undergo numerous conformational changes. Between proteins creation by the ribosome and its eventual breakdown via proteolysis, a protein molecule can assume a vast variety of conformational states within a biological system, with these states interconverting over a broad range of timescales. Factors such as pH, temperature, and interaction with biological surfaces can affect protein folding leading to incorrect arrangements. Misfolded proteins are associated with several increasingly common human disorders, including Alzheimer's, Parkinson's, type II diabetes, and various systemic amyloidosis.

The misfolding process triggers a cascade of events that culminates in the formation of fibrillar protein aggregates responsible for several human diseases. The presence of stable fibrillar aggregates in the organs of patients suffering from protein deposition diseases led initially to the reasonable postulate that mature fibrils are the main cause of the diseases. However, recent findings have raised the possibility that fibril precursors, such as soluble oligomers, represent the pathogenic species responsible for the disease onset and severity.

In the context of neurodegenerative disorders, a promising approach to prevent toxic aggregate formation involves anti-amyloid molecules. Current treatments mainly address symptoms without reversing neurodegenerative processes. Redirecting the abnormal aggregation cascade using small molecules is a more feasible strategy than complete inhibition. Aminosterol compounds, discovered in dogfish sharks, offer optimism for effective treatments. These compounds have potential to influence protein aggregation in neurodegenerative diseases like Alzheimer's and Parkinson's.

This chapter outlines the biological context of the scientific problem being investigated. Specifically, Section 2.1 presents the protein folding process, while Section 2.2 discusses protein misfolding. Section 2.3 explores the connection between protein misfolding and Alzheimer's Disease, highlighting the current state of drug discovery. Finally, Section 2.4 is dedicated to the role of aminosterols in addressing neurodegenerative disorders.

2.1. Protein folding

Proteins, which are produced by ribosomes through the instructions contained in the DNA, are the second most common molecule in biology after water and are involved in every cellular biological function. The protein folding results in a high-compact three-dimensional structure and is the most sensational and fundamental self-assembly biological event. Self-assembly with high accuracy and fidelity of even the most complex molecular structures is one of the hallmarks of a living system. One of the great difficulties of contemporary research is the investigation of the procedures through which this occurs.

It is well known that folding catalysts and molecular chaperones are just two of the many auxiliary factors present in the cells of living organisms that aid in the folding process. These components facilitate the folding of polypeptide chains without define the polypeptides native shapes. The thermodynamically most stable configurations of proteins under physiological conditions usually always correspond to their native states. However, the protein folding process is a multidimensional problem that have influence on how a specific protein works. In fact, a systematic search for this structure, would take an unimaginably long amount of time because a huge number of potential conformations for any given polypeptide chain existed.

During the folding process, the protein reaches its final three-dimensional conformational structure, which is decided by an overall free energy balance. Protein native structure creation is thought to be guided, in part, by the energy landscape, which is the core principle behind the energy landscape theory. The native-state-forming interaction allows the polypeptide chain to systematically explore different structures until it finds one with the lowest energy. There are certain kinetically desirable stages that the unfolded protein must take to return to its natural state. There are kinetic hurdles to attaining the native state for large proteins, as well as for some smaller proteins. Therefore, when given the freedom to fold on their own, these proteins become "kinetically stuck" at local energy minima.

2.2. Protein misfolding

The protein three-dimensional structure is linked to others biological function as guidance of cell growth and differentiation and trafficking of molecules³⁶. The proper biological function is admitted only for correct folded proteins which are stable throughout time in complex biological contexts and can interact selectively with their endogenous partners. These complex roles further underscore the critical importance of protein three-dimensional structure in

maintaining the complex and dynamic processes that underpin the overall functionality of living organisms. Furthermore, peptides and proteins have coevolved with their biological contexts to stay in their soluble forms. However, there are situations in which they can clump together forming ineffective and even harmful protein aggregates. In this context several pathological conditions related to neurodegenerative disorders are triggered by a molecular phenomenon known as misfolding, where the specific protein folding is lost.

Although the protein aggregates implicated in various neurodegenerative disorders differ in their specific composition, it is striking that the underlying process of protein misfolding, its intermediate states, end-products, and principal characteristics exhibit remarkable similarities³⁷. From a mechanical and biological standpoint, the most crucial of these states are summarized in Figure 3. Several proteins of interest such as α -synuclein, tau, and the islet amyloid polypeptide (IAPP) are naturally unfolded or intrinsically disordered in solution, although it is believed that many of these proteins fold into more well-defined structures in response to contact with specific biological players. Both intrinsically disordered systems and globular proteins have been implicated in protein aggregation associated with illness, highlighting the complexity of protein folding processes and their critical role in maintaining proper cellular function and overall organism health.

The protein folding is heavily influenced by the specific cellular environment in which it takes place, such as the cytoplasm, mitochondria, or endoplasmic reticulum, among others. Additionally, the temporal aspect of the folding process, whether it occurs concurrently with or following protein synthesis, also plays a crucial role in determining the folding outcome³⁸. However, improper interactions with other molecules in the crowded cellular environment are possible for incompletely folded proteins due to their exposure of native-state-buried portions of structure to the solvent³⁹.

Protein folding and unfolding are the ultimate mechanisms for inducing and suppressing a wide variety of biological behaviours such as membrane translocation, trafficking, secretion, immune response modulation, and cell cycle control. Therefore, improper folding or the failure to maintain proper folding will lead to the dysfunction of biological systems and, ultimately, illness⁴⁰. Occasionally, proteins with a strong propensity to misfold bypass all the safeguards and instead form insoluble clumps either inside the cell or, more frequently, in the extracellular space. Deposition of such aggregates in tissues including the brain, heart, and spleen is directly linked to a growing number of diseases.

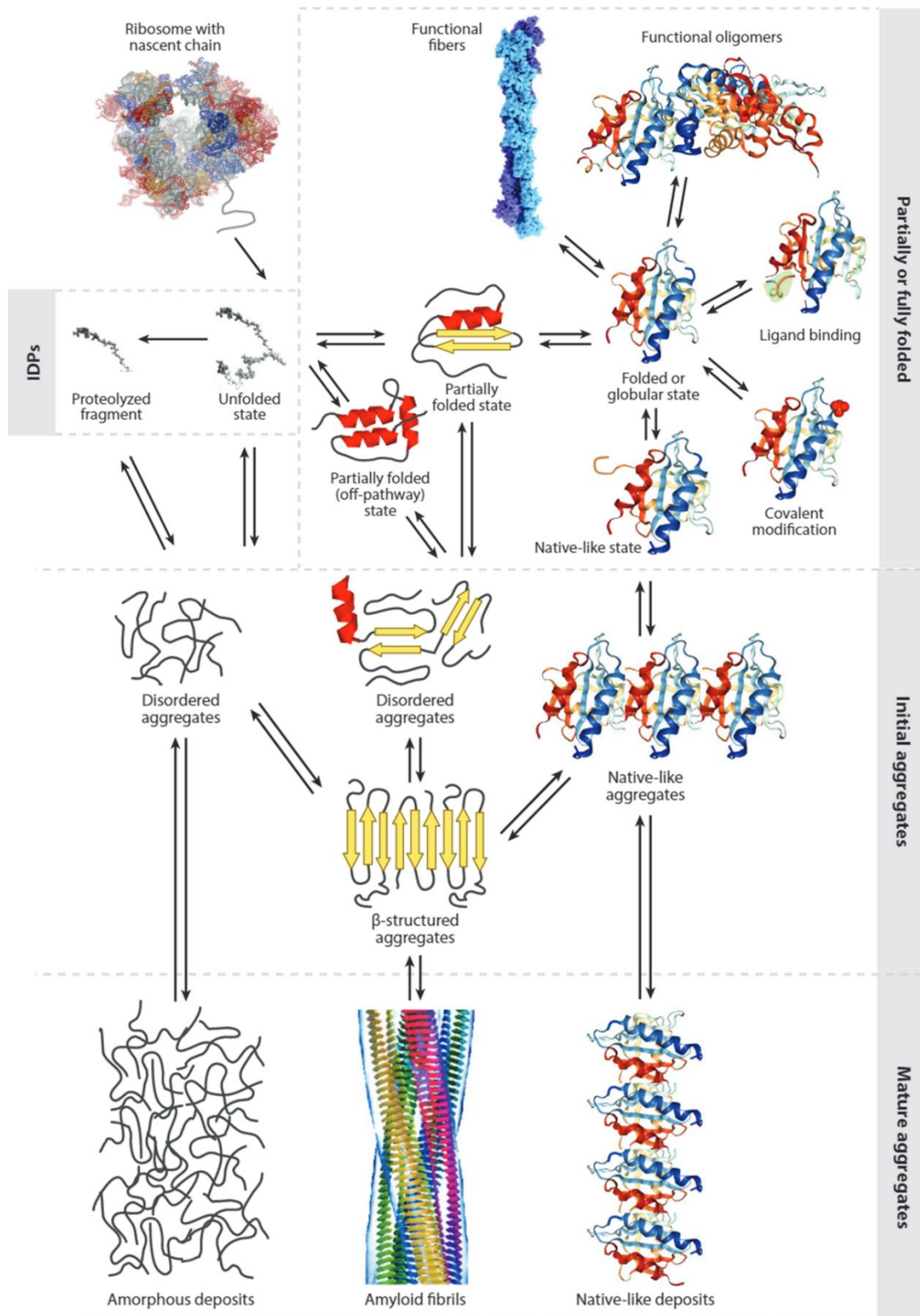


Figure 3: Diagram depicting the various conformations that a polypeptide chain can take after biosynthesis and the possible transitions between them. Intrinsically disordered proteins (IDPs), partially folded proteins, newly formed oligomers, and fully formed aggregates are all represented by different boxes. The proteostasis network meticulously controls all of these conformational states and their transitions in the living organism. Amyloid fibrils (bottom, center), native-like deposits (bottom, right), and amorphous deposits (bottom, left) are all products of protein aggregation, and all three are linked to pathogenic conditions when they form in an unregulated fashion⁴¹.

The proteins most frequently associated with the accumulation of misfolded aggregates in the brain in the context of neurodegenerative disorders encompass amyloid-beta ($A\beta$) and tau in Alzheimer's disease (AD), α -synuclein in Parkinson's disease (PD), multiple system atrophy, and dementia with Lewy bodies^{36,41}. Aggregates can form when molecules are in such intermediate, partially folded conformations. In response to these hypotheses, a plethora of experimental and theoretical studies of the folding process have been conducted^{42,43}.

2.3. Alzheimer's disease

Alzheimer's disease (AD) is a highly debilitating pathological condition that is getting more prevalent as the world's population ages. It is estimated that there are already 50 million AD sufferers in the globe, and that number will rise to more than 150 million by 2050⁴⁴. AD is directly related to the abnormal aggregation of misfolded amyloidogenic proteins and one of the most widely accepted theories to explain the irreversible progress of the disease is the amyloid hypothesis. In detail, the amyloid hypothesis proposes amyloid β ($A\beta$) peptide as the main cause of the AD⁴⁵. $A\beta$ peptides are formed from 36 to 43 amino acids originated from the cleave of the amyloid precursor protein (APP). In the non-amyloidogenic pathway, APP is cleaved preferentially by α -secretase. In the amyloidogenic pathway, neurotoxic $A\beta$ peptides are released after sequential cleavage of APP by β and γ -secretases in the transmembrane region, and further accumulate into oligomeric aggregates (Figure 4). In this context, AD is characterized by $A\beta$ brain levels 4–6 times higher due to altered APP processing and poor $A\beta$ elimination⁴⁶.

The misfolding of the extracellular $A\beta$ protein accumulated in senile plaques and the intracellular deposition of the misfolded tau protein in neurofibrillary tangles cause memory loss and confusion, leading in personality and cognitive decline over time⁴⁷. In detail, the aggregation process of $A\beta$ transpires through several distinct stages. Initially, monomers undergo self-assembly to form toxic oligomers. Subsequently, in the second stage, fibril growth occurs via primary and secondary nucleation mechanisms. This progression ultimately culminates in the formation of senile plaques, which are characterized by the coexistence of fibrils and a small concentration of monomers⁴⁸.

Post-mortem examination of brain tissue from AD patients has revealed the presence of plaques and tangles in close proximity to deteriorating neurons and vascular structures. Brain

accumulation of both A β and tau causes several pathophysiological changes that affect cognitive function in AD patients. One of these pathophysiological changes involves intracellular calcium dysregulation in neurons and astrocytes that can result in excitotoxicity, a pathological process in which excessive glutamate receptor activation leads to neuronal damage and eventual cell death. Another pathophysiological change involves astrogliosis and microgliosis, where reactive astrocytes and microglia affect synaptic transmission, which contributes to cognitive impairment. A third pathophysiological change is a reduced lysosomal degradation capacity of neurons that leads to autophagy and neuronal loss. The fourth change is development of cerebral amyloid angiopathy (CAA), where A β builds up in cerebral blood vessels resulting in ruptured vessel walls over time. Collectively, these pathophysiological changes make AD a multifactorial disease and add to the challenges in finding an effective treatment for patients.

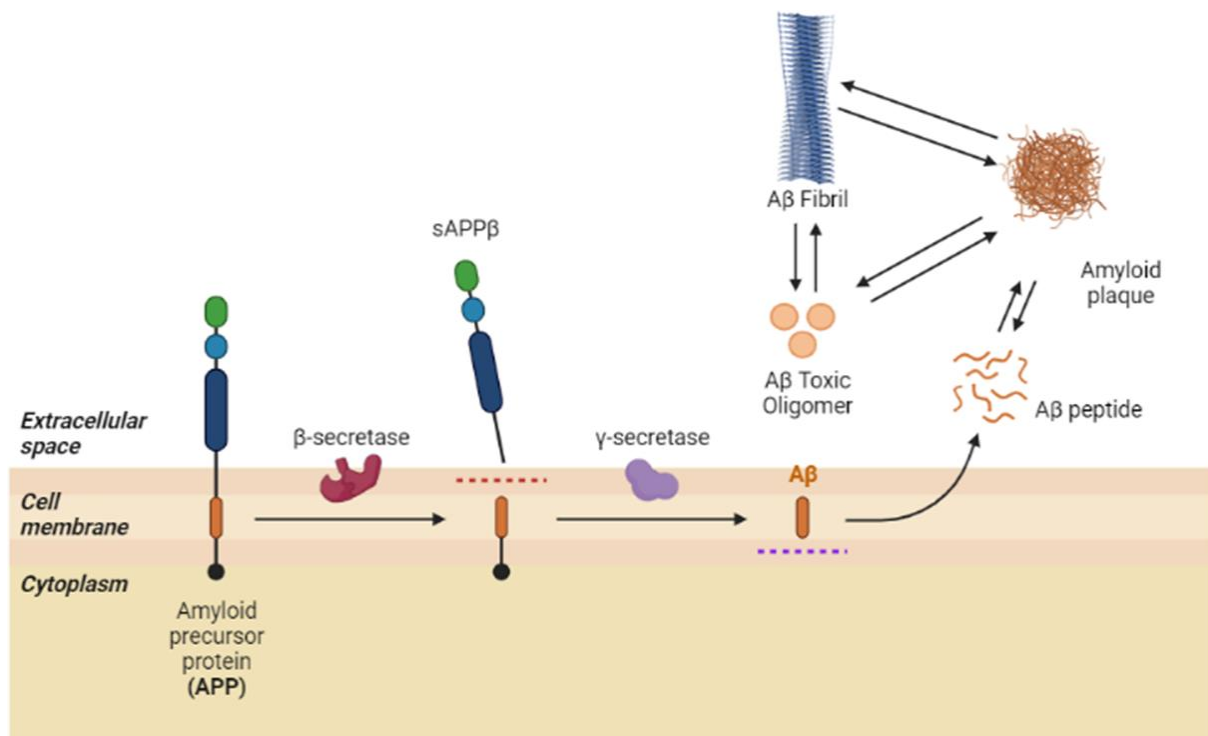


Figure 4: An overarching model of how amyloidogenic peptides and proteins aggregate. β and γ -secretases secretase enzymes cleave the amyloid precursor protein (APP) in the transmembrane region producing neurotoxic A β peptides. Misfolding and aggregation of monomeric proteins can result in the development of forward oligomers and fibrils, which can aggregate to create amyloid plaques. Fibrils break apart as they reach maturity, releasing extremely diffusible reverse toxic oligomers.

Drug development against Alzheimer's disease

Drug development for AD has proved incredibly challenging as evidenced by the fact that no new medicine has been approved by the FDA since 2003⁴⁹. The expanding public health concern presented by AD necessitates the development of therapies to prevent the illness, postpone the onset, decrease the course, and ameliorate the overall symptoms⁴⁹. Currently, only four cholinesterase inhibitors and one N-methyl-D-aspartate (NMDA) receptor antagonist are approved for the treatment of AD⁴⁹⁻⁵¹. Four of the six 3 compounds investigated during clinical studies Phase 3 has the A β protein as the primary pharmacologic target. About the 65% of the 221 studies of disease-modifying drugs that were registered between 2002 and 2012 were aimed in this direction. The target has not been verified, and no group of medicines has shown effectiveness against this target in human clinical trials. Anti-A β drugs have proven pharmacological and behavioral benefit in several animal models of amyloidosis; nevertheless, numerous limitations have been highlighted in the application to human⁴⁹. Overall, AD is more costly to the U.S. economy than either cardiovascular disease or cancer. In order to alleviate this socioeconomic impact and improve patient outcomes, enhancements to the drug development pipeline are essential. Such improvements may be achieved through the creation of more accurate animal models that better mimic the pathophysiology of AD in humans, exploring a broader array of therapeutic targets, and adopting multi-targeted approaches that address several disease pathways at once.

Targeting amyloid beta aggregation molecular mechanism

A β_{42} represents an intrinsically disordered peptide, and the inhibition of its aggregation has been identified as a crucial therapeutic approach against AD. Despite the promise of this strategy, no small-molecule compounds designed to impede its self-assembly have demonstrated significant clinical efficacy to date^{52,53}. One potential explanation can be attributed to the insufficient understanding of the complex molecular mechanisms responsible for generating harmful species associated with A β_{42} . Additionally, the precise means by which small molecules might disrupt the aggregation pathway of A β_{42} remain poorly elucidated. Furthermore, A β_{42} peptide has been found to encode multiple aggregated isoforms, each exhibiting distinct secondary and tertiary structures, increasing the complexity of design new drugs that can effectively target different A β_{42} polymorphisms⁵⁴.

Moreover, several studies suggest the primary role of prefibrillar oligomeric species as pathogenic agent in AD rather than the mature amyloid fibrils and plaques. As a result, effective

therapeutic interventions are expected to extend beyond the suppression of A β ₄₂ fibril formation. In this context, the aim should be the selective targeting of specific oligomeric species by precisely intervening during the intricate aggregation process of A β ₄₂. A therapeutic strategy of this nature has become increasingly plausible due to recent significant breakthroughs in our comprehension of the molecular processes that underlie amyloid formation. These advancements can be attributed to the pioneering application of chemical kinetics in the investigation of protein aggregation phenomena⁵⁵. Upon reaching a small yet critical concentration of A β ₄₂ aggregates through the primary nucleation of monomers, surface-catalyzed secondary nucleation emerges as the predominant process. In this context, the surfaces of the pre-formed A β ₄₂ fibrils act as catalytic sites, enhancing the generation of A β ₄₂ toxic oligomeric species⁵⁵. Subsequently, these oligomers undergo growth and conversion into additional fibrils, which in turn foster the production of more toxic species. This highly efficient catalytic cycle perpetuates the formation and propagation of these detrimental oligomeric entities⁵⁶.

In this context, it has been observed that some of the most toxic oligomers of A β peptides are embedded within the plasma membrane of neuronal cells, where they form calcium-permeable amyloid pores⁵⁷. Among the cell membrane lipid composition, cholesterol assumes a critical role, as its concentration within the plasma membrane of neural cells serves as a key determinant of A β peptide neurotoxicity^{58,59}. Cholesterol has been found to specifically interact with monomeric A β peptides⁶⁰, facilitate their insertion into the plasma membrane, and promote their functional oligomerization into calcium-permeable ion channels. In this context, it was observed that to prevent the formation of these oligomers and disassemble preformed channels, it is necessary for therapeutic molecules to access the lipid environment, a characteristic primarily exhibited by lipophilic and/or amphipathic compounds⁶¹.

2.4. Aminosterols against neurodegenerative disorders

A feasible strategy to prevent the neurodegeneration is based on the development of anti-amyloid molecules, i.e., those capable of preventing the generation of toxic aggregates. Currently, there is a critical need for effective drugs that can considerably delay the development of neurodegenerative illnesses. Upon diagnosis, patients suffering from neurodegenerative diseases such as Alzheimer's disease and Parkinson's disease are frequently prescribed medications that primarily address symptoms but lack the capacity to halt or reverse the underlying neurodegenerative processes^{62,63}.

Substantial research efforts have been directed toward unraveling the potential of various molecules to influence the aggregation reactions of proteins implicated in neurodegenerative diseases leading to the identification of molecular species that either promote or inhibit fibril formation⁶⁴.

Numerous studies highlight the ability of small molecules to redirect the abnormal aggregation cascade rather than its entire inhibition. Upon reflection, this observation appears reasonable, given the vast amount of buried surface area within protein aggregates compared to the relatively diminutive size of inhibitor molecules. Consequently, employing small molecules to modify the nucleation pathway, either by disrupting specific intermolecular contacts⁶⁵ or promoting non-canonical ones, seems to be a more viable approach for preventing the formation of toxic aggregates than attempting to antagonize all potential intermolecular interactions.

In recent years, a novel class of aminosterol compounds with antimicrobial activity, initially discovered and identified within the digestive tracts of dogfish sharks (*Squalus acanthias*), has emerged as a source of optimism in the search for effective treatments for neurodegenerative diseases such as AD and PD⁶⁶.

Squalamine

The first aminosterol discovered was squalamine (Figure 5). At physiological pH, squalamine adopts a zwitterionic form, characterized by the presence of both positive and negative charges within the molecule, giving it the property of being soluble in both aqueous and organic environments. Squalamine has demonstrated a diverse range of biological activities, both in vitro and in vivo. Studies have shown its potential to inhibit pathological angiogenesis associated with retinopathy and cancer, which is the abnormal growth of blood vessels that can lead to disease progression. Additionally, squalamine has exhibited antiviral activity against a variety of human viruses, including both RNA- and DNA-enveloped viruses, indicating its potential as an antiviral therapeutic agent. Furthermore, in mouse models, squalamine has been observed to counteract age-related declines in colonic motility and jejunal vagal firing rates, suggesting its potential to ameliorate certain age-related physiological impairments. In 2017 a study highlighted the squalamine potential therapeutic action in combating PD in vivo and in vitro. The proposed mechanism was based on either targeting the kinetics of the protein aggregation process or directly acting upon the aggregates of misfolded proteins⁶⁷. Specifically, squalamine was reported to inhibit α -synuclein aggregation, protecting neuronal cells from the

detrimental effects of toxic oligomers, and ameliorate motility deficits and paralysis in a *Caenorhabditis elegans* model of PD expressing human α -synuclein⁶⁷. Furthermore, evidence coming from dopaminergic neurons of patients with PD described the primary nucleation of α -synuclein induced by the cell membrane⁶⁸. In vitro experiments have showed the squalamine ability to suppress the α -synuclein primary nucleation preventing its activity towards living cells.

Trodsuquemine

In 2001, during squalamine extraction procedure, seven other aminosterols in considerably lower concentrations than squalamine were discovered. These newly identified molecules exhibited structural similarities to squalamine, with variations primarily occurring in the cholestane-like side chain, ranging from carbon-20 (C-20) to carbon-27 (C-27). One of these new aminosterols was trodsuquemine (Figure 5), originally named MSI-1436, which is able to cross the blood-brain barrier and can stimulate the regeneration of tissues and injured organs^{69,70}.

This new natural product was found to have a wide spectrum of antimicrobial activity, similar to squalamine. The minimum inhibitory concentration of trodsuquemine was found to be approximately four times lower against *E. coli*, *P. aeruginosa*, and *C. albicans* related to squalamine⁶⁶.

Nowadays, squalamine and trodsuquemine can be chemically synthesized in the laboratory, facilitating systematic chemical and clinical investigations. From a structural viewpoint, squalamine and trodsuquemine are very similar (Figure 5). Both are cationic amphipathic aminosterols composed of a sulphate moiety, a central sterol group, and an alkyl polyamine group on the other side of the sterol: squalamine has a spermidine moiety (7 methylene and 3 amino groups) while trodsuquemine consist of a spermine moiety (10 methylene and 4 amino groups).

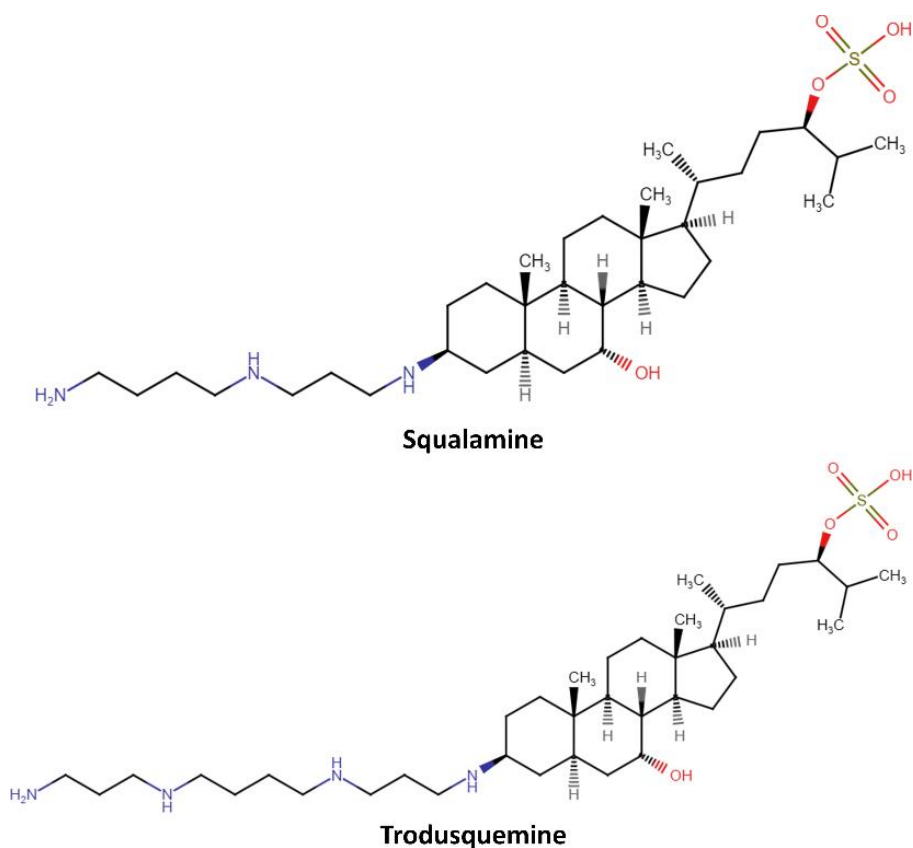


Figure 5: Squalamine and trodusquemine chemical structures. Both aminosterols are formed by a sulfate group (right) and a sterol (centre). The structural distinction between squalamine and trodusquemine lies in the polyamine moiety connected to carbon-3 (C-3) of the molecule. In squalamine, the polyamine attached to C-3 is spermidine, while in trodusquemine, it is spermine.

Trodusquemine on protein aggregation

Trodusquemine has demonstrated the ability to inhibit α -synuclein aggregation by displacing both monomers and oligomers of the protein from cell membranes. However, unlike the previous report on squalamine, which focused solely on lipid-induced nucleation, trodusquemine was shown to inhibit not only the lipid-induced primary nucleation of α -synuclein but also a key secondary process involved in α -synuclein aggregation, specifically secondary nucleation⁷¹.

In contrast to α -synuclein, in vitro and in vivo studies have shown that trodusquemine was able to accelerate the A β aggregation process responsible of the AD, maintaining unaltered the primary nucleation and enhancing the secondary nucleation⁷². As a result, trodusquemine appears to shift the A β oligomer toxic population by accelerating their conversion into relatively non-toxic A β fibrils, effectively depleting the pool of oligomers responsible for causing cellular dysfunction^{72,73}. Recent studies have highlighted the recovery of spatial memory deficits, the

reduction in size of amyloid plaques and the prevention of neuronal loss after administration of trodusquemine⁷⁴. Furthermore, the interaction of toxic A β oligomers with cell membranes and the resulting toxicity cascade was prevented by trodusquemine in neuronal cell cultures.

There are at least two possible mechanisms for reducing the toxicity of A β oligomers by the action of trodusquemine: i) aminosterol binds to the oligomers responsible for AD preventing interaction with the cell membrane, ii) trodusquemine interacts directly with the plasma membrane protecting it from aggression by abnormal A β aggregates. It was recently found how trodusquemine does not affect the structural and morphological toxic characteristic of misfolded aggregates, thus discouraging the first hypothesis⁷⁵.

Understanding the overall mechanism of the trodusquemine behaviour in a physiological environment is fundamental to clarify the protective mechanism exerted against A β oligomers. In this context, the mechanism by which aminosterols interact with itself and with the cell membranes, how they change their properties and how this process leads to an A β oligomer resistant membrane is not yet clear.

Chapter 3

Trodesquamine Molecular Mechanism of Action

Neurodegenerative diseases are a series of pathologies characterized by a progressive and irreversible loss of neurons in the central nervous system. The mechanisms involved in these diseases are very complex and consequently still mostly unknown. However, according to the amyloid hypothesis, the cause can be attributed to the conversion of specific peptides or proteins from their native, soluble, and intrinsically disordered form into well-defined, insoluble fibrillar aggregates. Aminosterols have recently been proposed as active protagonists against protein misfolding in neurodegenerative diseases and several studies have highlighted the ability of these molecules to block the neurotoxic effect of α -synuclein and A β aggregates, known to be among the main causes of the development of Parkinson's and Alzheimer's diseases.

The Section 3.2 focuses on examining the interactions of trodesquamine with the outer leaflet of the neuronal cell membrane. The computational results complement experimental evidence by offering valuable, quantitative insights, serving as a molecular microscope. Specifically, trodesquamine can alter the physico-chemical properties of the lipid bilayer, rendering it more resistant to the attack of toxic A β oligomers.

The Section 3.3 emphasizes the importance of developing a multiscale computational approach to comprehend supramolecular phenomena. Specifically, a coarse-grained model of trodesquamine is created and validated. Subsequently, the self-assembly and aggregation behavior of the aminosterol in various physiological environments are investigated.

The section 3.4 explores the use of MD simulations to understand lipid-raft interactions and the behavior of trodesquamine within a biphasic membrane. Due to size and time scale limitations, atomistic MD simulations are challenging, thus Martini coarse-grained force field simulations are employed. These simulations offer valuable structural insights into lipid-raft interactions and their significance in cellular biology.

The section 3.5 focuses on compute essential thermodynamics properties for drug design as binding affinity of aminosterols for the cell membrane. A complex computational protocol is employed, including coarse-grained representation, enhanced MD simulations and dimensionality reduction techniques.

3.1. Aim of the Work

In the context of Alzheimer's disease, trodusquemine has demonstrated its efficacy in rescuing spatial memory deficits, reducing plaque size, and preventing neuronal loss in vivo mouse models⁷⁴. Moreover, it can effectively cross the blood-brain barrier following intraperitoneal administration. Initially isolated from the liver of the dogfish shark (*Squalus acanthias*)⁷⁶, trodusquemine, is currently undergoing in Phase 1 and 1b clinical trials for its potential in treating obesity and diabetes. It has shown notable success in stimulating tissue and organ regeneration following injury in animal models, without affecting uninjured tissues^{69,70}. Despite the promising findings, unraveling the precise molecular mechanism of action of trodusquemine remains a crucial challenge.

To achieve this objective, the present work employs molecular dynamics simulations to provide a comprehensive understanding of the phenomenon, from the moment the molecule is in solution to its interaction with the biological target. Figure 6 presents a schematic representation of the cascade of molecular phenomena involving trodusquemine. Additionally, investigating this complex pattern necessitates the use of a dual-resolution trodusquemine model, depending on the temporal and spatial scale of interest.

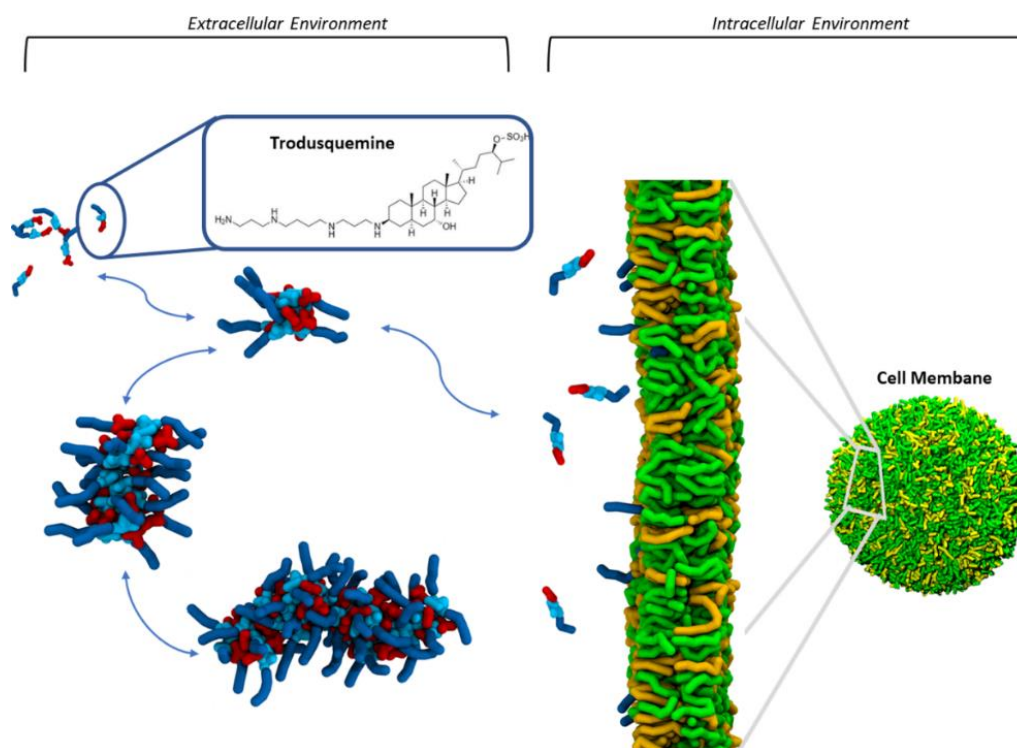


Figure 6: Schematic trodusquemine aggregation (left) and release mechanism (right). The trodusquemine is shown in blue (tail), cyan (core) and red (head). The release mechanism shows the interaction of trodusquemine with a hypothetical membrane patch (green and yellow).

In detail, the following sections will describe the interaction of trodusquemine with neuron-like cellular membranes at the atomistic scale and how it alters their physicochemical properties. Subsequently, a coarse-grained model of the aminosterol is developed and validated to elucidate the molecular mechanism of trodusquemine self-assembly up to supramolecular scales. The multiscale model is employed to understand the trodusquemine influence on the dynamics of lipid raft formation and demixing, which are molecular phenomena that develop on timescales inaccessible to atomistic modeling. Finally, enhanced sampling techniques are utilized to quantify the interaction of aminosterols with the cellular membrane, leveraging sophisticated methods of dimensionality reduction.

3.2. Interaction with the Cell Membrane

Introduction

There are many different neurodegenerative disorders, including Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and spongiform encephalopathies (SE), that are characterized by progressive and irreversible neuronal loss⁴¹. It is believed that a great deal of neurodegenerative diseases is related to fibrillar aggregates formed by proteins or peptides converting from their native soluble forms. It is possible for fibrillar aggregates to form various types of extracellular deposits or intracellular inclusions, including amyloid plaques, neurofibrillary tangles, and Lewy bodies⁴¹.

Scientists have found that aminosterols isolated from the dogfish shark *Squalus acanthias* prevent amyloid formation and misfolded oligomeric species from interacting with biological membranes, making them promising drug candidates⁷⁷. In detail, squalamine and trodusquemine are the aminosterols that have been explored the most. Squalamine is undertaking phase-2 and phase-1 clinical trials for the treatment of Parkinson's disease-related constipation and dementia, respectively (NCT03781791 and NCT03938922). Trodusquemine is an equally intriguing molecule because it is effective at lower doses than squalamine,^{72,77} has been reported to cross the blood-brain barrier after intraperitoneal administration⁷⁸, and stimulates the regeneration of injured tissues and organs without affecting uninjured tissues, as observed in zebrafish and mice⁷⁹. Recent Alzheimer's disease mouse models have demonstrated that trodusquemine can prevent hippocampal neuron loss and spatial memory deficits, decrease plaque size, and prevent neuronal death⁸⁰.

Squalamine and trodusquemine are cationic amphipathic aminosterols that share a sterol group, a sulphate moiety at position 24, and an alkyl polyamine group at position 3, which is a spermidine moiety in squalamine (7 methylene and 3 amino groups) and a spermine moiety in trodusquemine (10 methylene and 4 amino groups)⁷⁶. Furthermore, trodusquemine has a larger net positive charge than squalamine at physiological pH due to its extra amino group.

In vitro study shown that squalamine displaced monomeric α -synuclein, a protein linked to PD, from lipid vesicles and inhibited its lipid-induced aggregation into fibrils. Furthermore, squalamine introduced into the extracellular media reduced the toxicity of oligomers by preventing them from associating with neuronal membranes⁸¹. Perni and co-workers by treating a transgenic *C. elegans* model of PD that overexpresses human α -synuclein with squalamine,

observed that the worms' locomotor skills were identical to those of control worms and preventing any α -synuclein inclusions⁸¹. Another study showed that trodusquemine inhibited α -synuclein aggregation and lipid-induced secondary nucleation without any impact on fibril elongation. In this context, trodusquemine compared with squalamine was equally effective at lower dosages in preventing α -synuclein deposition and restoring motility in transgenic worms, as well as inhibiting the interaction of oligomers with the membranes of cultured neurons.

By increasing secondary nucleation and, to a lesser extent, fibril elongation, trodusquemine has been shown to accelerate amyloid aggregation, while having no effect on the rate of primary nucleation. Overexpression of human A β in *C. elegans* worms resulted in increased A β deposition in trodusquemine-containing medium, confirming the in vitro results. Finally, trodusquemine was shown to repair the motility deficit of the worms and prolong their longevity in vivo and in transgenic mouse AD models has reversed memory loss in spatial contexts protecting neurons⁸⁰. Interestingly, evidence on neural cell cultures showed the reduced toxicity of the A β oligomers against the cell membranes and the conversion to fibrils accelerated in the presence of trodusquemine⁸⁰.

Both aminosterols have been shown to displace misfolded oligomers of α -synuclein and A β from cell membranes, although the precise molecular mechanism it is unknown. Furthermore, evidence of the aminosterols bind the cell membrane are available only for the squalamine only indirectly.

Herein, we examine trodusquemine and provide data showing that trodusquemine binds to the lipid bilayer of reconstituted liposomes generated by four natural lipids that mimics the out-layer composition of neurons. Based on our findings, we know that this interaction alters the physicochemical characteristics of lipid membrane phase domains in a way that effectively blocks the binding of misfolded protein oligomers. Using a multidisciplinary approach, this work not only sheds light on the compound's mechanism of action but also on how altering the physicochemical features of lipid membranes might make them more resistant to the interaction and harmful consequences of misfolded protein aggregates.

Material and Methods

Two molecular systems were investigated: a phospholipid bilayer constituted by 400 lipids in water environment and the same bilayer in presence of trodusquemine with trodusquemine:lipid ratio of 1:200. Each system composition is reported in Table 3.

Table 3: Molecular systems composition.

Sys.	DOPC	SSM	GM1	CHL	TROD	Time [ns]	Ions [mM]
Mem	262 (65.5%)	130 (32.5%)	4 (1%)	4 (1%)	-	1000	150
TROD	262 (65.5%)	130 (32.5%)	4 (1%)	4 (1%)	2	1000	150

This composition was chosen to mimic the external leaflet of neural cell membranes⁸². The 1% CHOL concentration, which is below the physiological level, was chosen to favour well-separated ordered domains and disordered regions⁸³. However, all-atom MD simulations are unable to observe a distinct phase separation within computationally reasonable timeframes. Nonetheless, the same experimental LUVs composition was employed to ensure a proper comparison.

Each membrane of 400 lipids and 3 nm of water thickness were obtained using the CHARMM-Builder^{84,85}. Then 150 mM NaCl was added to reproduce the physiological environment with an excess of 4 Cl⁻ ions to neutralize the net system charge. For each system, 1000 steps energy minimization was conducted using the steep descent algorithm. Two 25 ps position restrained simulations with decreasing force constants (2000 and 1000 kJ mol⁻¹·nm⁻²) were performed under the canonical ensemble at 300 K. Then, three position-restrained simulations of 50 ps (force constants of 1000, 400, 200 kJ mol⁻¹·nm⁻²) were performed using Berendsen semi-isotropic pressure coupling scheme (time step = 2 fs; pressure time constant $\tau_P = 5.0$ ps)⁸⁶ at a reference pressure of 1 atm. Finally, 1 μ s long production MD was carried out using the v-rescale thermostat (T = 300 K; $\tau_T = 1.0$ ps)⁸⁷, together with the semi-isotropic Parrinello-Rahman barostat (P = 1 atm, $\tau_P = 2.0$ ps)⁸⁸.

Each molecular system was investigated by two independent replicas to increase the statistical data. The simulation time step was set to 2 fs in conjunction with the LINCS algorithm⁸⁹. The particle mesh Ewald (PME) method was employed to calculate long-range electrostatic interactions⁹⁰. The van der Waals interactions were calculated by applying a cut-off distance of 12 Å and switching the potential from 10 Å. The trodusquimine protonation state was evaluated at pH 7.0 using Avogadro software⁹¹ and the most probable protonation state i.e. trodusquimine

protonate was used. Forcefield topology of trodusquimine molecule was taken from the CHARMM General Force Field⁹². Phospholipid and water topology were parametrized by the CHARMM36m force-field⁹³ and TIP3p model, respectively. All MD simulations were performed using the GROMACS 2018 package⁹⁴. The Visual Molecular Dynamics (VMD) software was used to monitor all simulation trajectories⁹⁵.

Membrane properties – Lipid diffusion

The impact of the trodusquimine on the membrane was evaluated in terms of diffusion coefficient and membrane mechanical proprieties. The lateral diffusion coefficient (D) was obtained from the mean square displacement (MSD) of the DOPC and SM molecules, removing the effect of the center of mass motion of each phospholipid molecule on each leaflet, using the Einstein relation, and performing a linear fitting of the MSD data on the time interval between 50 and 75 ns on time window of the last 300 ns.

Membrane properties – Bending modulus

Mechanical properties were estimated in terms of bending modulus for monolayer as previously done in literature^{96–98}. In a greater detail, the bending rigidity is associated with the ability of the lipid membrane components to change their orientation with respect to each other⁹⁶. The above-mentioned lipid ability is quantified as splay angle (α , ranging in 0–90 degree) and directly related to the splay modulus χ_{12} as follows. By monitoring the variation of splay angles on both leaflets, it is possible to obtain a normalized probability distribution used to calculate the potential of mean force (PMF) as:

$$PMF(\alpha) = -k_B T \ln \frac{P(\alpha)}{P_0(\alpha)} \quad (3.1)$$

where $P_0(\alpha) = \sin(\alpha)$ is the probability distribution of a hypothetical non-interacting particle system⁹⁹, k_B Boltzmann factor and T the temperature of the system.

In the investigated systems, the splay modulus χ_{12} , can be extracted by a quadratic fit of the PMF data^{100–102} and is representative of the single membrane leaflet bending modulus, K_m . For a multi-components membrane the total bending module is a weight average between each lipid type⁹⁶. In detail the bending module was calculated in the last 200 ns of each MD replica, considering the DOPC and SM lipids.

Results and Discussion

Squalamine and trodusquemine have demonstrated inhibitory effects on the interaction between misfolded protein oligomers and cell membranes. The underlying mechanisms of their ability to prevent oligomer binding and toxicity may be attributed to at least two non-mutually exclusive possibilities. Firstly, aminosterols could bind to oligomers and prevent their binding to cell membranes. Secondly, aminosterols could bind to cell membranes and prevent the binding of oligomers. Recently trodusquemine administered at a concentration of 5 μM , which is consistent with the concentration used in previous studies, did not have any impact on the structural and morphological characteristics of misfolded oligomers.

In this context, MD simulations were run to provide atomistic insights into the interaction of trodusquemine with the membrane bilayer having the same lipid composition of experimental LUVs and SLBs (2 simulations, 1 μs aggregated simulation time) and illustrated the ability of the molecule to penetrate the membrane outer layer after a few hundred ns⁷³.

At physiological pH 7, the protonation state of trodusquemine is assessed to determine its net charge. Trodusquemine has a net charge of +3, as illustrated in Figure 7. This charge results from the negatively charged sulphate (SO_4^-) group and the protonated amino groups present on the spermine tail. Knowing the protonation state is essential for building a consistent and accurate model of trodusquemine, which is critical for studying its interactions and behavior in a biological environment.

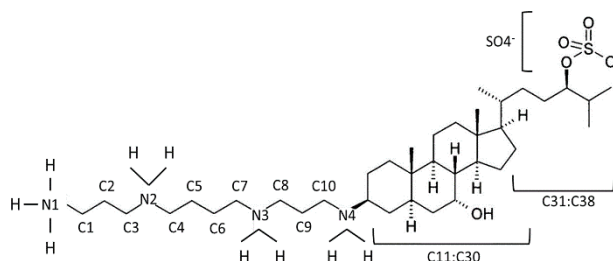


Figure 7: Trodusquemine protonation state evaluated at pH 7.0 through Avogadro software.

MD simulations can provide quantitative and qualitative information, complementary to experimental data. Detailed Figure 9A shows cell culture medium of human neuroblastoma SH-SY5Y cells incubated with trodusquemine and analysed by confocal microscopy. The images show the nuclei of cells stained with the Hoechst dye (Figure 9A, blue fluorescence) and trodusquemine (Figure 9A, red fluorescence) largely localising on the plasma membrane. However, the resolution of confocal microscopy is limited to micrometre-scale. MD

simulations make it possible to describe in detail on an atomistic level the behaviour of trodusquemine in the presence of a lipid membrane. In details, trodusquemine is positioned at the interface between the hydrophilic and hydrophobic layers. While the sterol ring (atoms C11:C30) is buried into the hydrophobic domain of the membrane, the SO₄⁻ group and the spermine (atoms N1:N4, C1:C10) are in proximity of the solvent (Figure 8B and Figure 9A). However, the sulphate head is only partially in contact with the hydrophilic region of the membrane, as shown by analysing the position of trodusquemine with respect to the density profile of each system component (Figure 9B), and also by the visual inspection of the molecular system (Figure 8B).

Furthermore, MD investigations shows that trodusquemine exposes a solvent accessible surface area (SASA) of 40.6 ± 6.4 % of the total trodusquemine surface, in agreement with the experimental value of $38 \pm 1.5\%$ obtained with quartz crystal microbalance on supported lipid bilayers⁷³. The positively charged tail of trodusquemine is exposed to the solvent and modulates the total surface charge by partially reducing the natural negative charge of the membrane. Experimental evidence by zeta potential measurements quantifies the negative charge reduction by 5.6 ± 1.2 mV relative to LUVs in presence of 5 μ M trodusquemine⁷³.

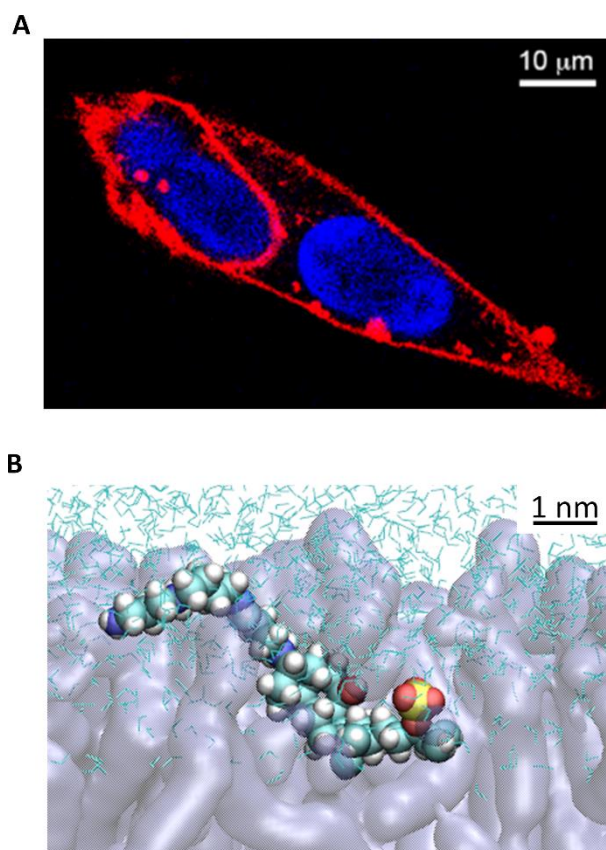


Figure 8: A) Magnified confocal image of an SH-SY5Y cell treated with 4 μM Alexa Fluor® 594-labelled trodusquimine, readapted from ⁷³. Hoechst-labeled nuclei and trodusquimine are indicated by blue and red fluorescence, respectively. B) Representative MD snapshot of trodusquimine within the membrane (N atoms in blue, C in cyan, H in white, O in red and S in yellow, lipids in transparent violet, water in cyan sticks).

The impact of trodusquimine on the membrane bilayer properties was evaluated by computing the lateral diffusion coefficient (D) of membrane lipids and the membrane bending modulus. The D of pure membrane is $4.5 \pm 0.8 \times 10^{-8} \text{ cm}^2/\text{s}$ and $4.5 \pm 0.3 \times 10^{-8} \text{ cm}^2/\text{s}$ for the DOPC and SM lipids, respectively, in line with the similar membrane composition previously investigated^{103,104}. Trodusquimine increases the D value by 33% for DOPC ($6.0 \pm 0.6 \times 10^{-8} \text{ cm}^2/\text{s}$) and 21% for SM ($5.5 \pm 0.6 \times 10^{-8} \text{ cm}^2/\text{s}$), as reported in Figure 9C. The monolayer bending modulus was found to be $10.0 \pm 0.3 \text{ k}_\text{B}\text{T}$, in accordance with the previous similar composition membrane¹⁰⁵ and does not change upon trodusquimine addition ($10.0 \pm 0.3 \text{ k}_\text{B}\text{T}$), as reported in Figure 9D.

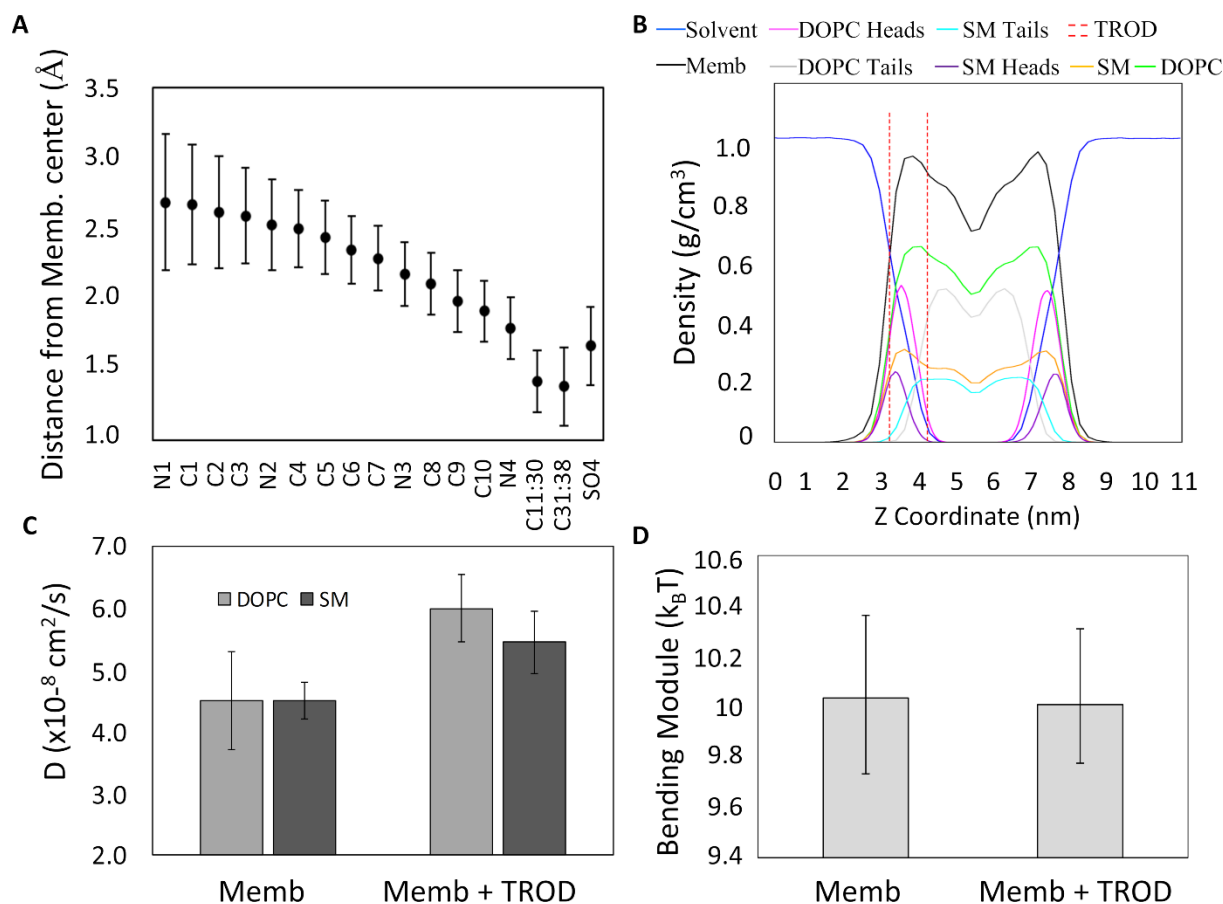


Figure 9: Trodusquimine is partially inserted into the lipid bilayers, modulating the membrane properties. A) Distance between each heavy atom of the trodusquimine and the membrane centre of mass, evaluated during the MD simulation at the equilibrium. B) Lipid density profile for the indicated system components along the Z coordinate (membrane centre positioned at 0.0 nm). The partial insertion of trodusquimine is indicated by red dashed lines. C) Lateral diffusion coefficient for DOPC and SM molecules for membrane without and with trodusquimine. (D) Monolayer bending modulus for membrane without and with trodusquimine. All experimental errors are standard deviation ⁷³.

The physicochemical alterations are widely recognized for their ability to hinder the binding of misfolded oligomers to cellular membranes, thus shielding the cells from their toxicity effects. Our findings collectively depict how the integration of sterol molecules harboring polyamine tails within cell membranes can induce changes in their physicochemical characteristics, rendering them more resilient against protein aggregates linked to neurodegeneration. This implies that potential therapeutic approaches may be formulated to make resistant cell membranes against misfolded protein assemblies. This finding may provide new insights into the mechanisms underlying the interaction between aminosterols and misfolded protein oligomers and may have implications for the development of potential therapeutics for protein misfolding diseases.

Conclusion

In conclusion, the use of MD simulations in conjunction with experiments proves to be an efficient and effective approach in elucidating molecular-level mechanisms of trodusquemine molecule. Our findings indicate that trodusquemine incorporates itself into lipid membranes, predominantly within the superficial hydrophilic regions. This interaction enables trodusquemine to effectively modulate the physicochemical properties of the lipid membranes. Remarkably, trodusquemine induces substantial changes in the lipid membranes. These alterations include a reduction in net charge, and a remodeling of the spatial distribution of lipid components increasing their diffusion. Our findings underscore the broader implications of how natural products that modulate cell membrane properties can serve as potential protective agents against the cytotoxicity induced by these deleterious protein aggregates. Furthermore, the ability to computationally model systems at the molecular level and visualize their behavior *in silico* provides valuable insights that have led to a better understanding of the behavior of trodusquemine and its interactions with different lipid components. This insight may pave the way for the development of novel therapeutic strategies targeting membrane-associated processes in neurodegenerative diseases.

3.3. Coarse-Grained models for unraveling the complexities of trodusquemine supramolecular phenomena

Introduction

All-atom (AA) simulations became the standard in biomolecular simulations in the 1980s and 1990s because of rising processing capacity and software optimization. AA representation accurately reproduces physical, chemical, and thermodynamic properties of molecular systems. However, the notion that, as capacity continued to expand, all processes might be represented with such sophisticated models proved to be unfounded. Even with access to supercomputer capacity, all-atom simulations are still restricted to spatio-temporal sizes of less than hundreds of nanometers and a few microseconds of simulation time^{106,107}, with the possibility of local minima trapping and oversampling in metastable states¹³. Furthermore, these precise molecular systems might objectively capture a limited number of molecular phenomena.

Inside most biological systems of interest, major macromolecules interact across extensive spatio-temporal scales. As a result, AA models cannot be used effectively for the observation of several critical phenomena and molecular processes that occur on lengthy characteristic timeframes. Several improved sampling techniques, such as metadynamics²⁹ and replica exchange molecular dynamics¹⁰⁸, have mitigated some of these limitations by enabling the rapid exploration of conformational space adding a bias potential. However, the aforementioned methods need substantial understanding of the biomolecular system under investigation and are impractical for greater systems.

Coarse graining (CG) is a common method for getting over the gap real-life molecular description. CG model involves streamlining the description of fine grain molecular models by decreasing their resolution and aggregating multiple atoms into CG beads. Minimal models of biomolecules are developed for CG molecular dynamics (MD) to capture their basic biophysics. Due to its easy-to-use building block feature, the Martini model is one of the most used CG models in biomolecular modeling¹⁰⁹. The Martini model primary premise is that specific CG beads may represent small chemical fragments. The main non-bonded bead interactions are described by Lennard–Jones potentials calibrated using experimental thermodynamic data, such as partitioning free energies. Bonded interactions are instead produced by optimizing bond distributions obtained from a reference AA simulation. Therefore, the Martini model incorporates both top-down and bottom-up parameterization methodologies^{110,111}. The Martini

CG philosophy is similar to other CG force fields, but different from bottom-up models that aim to precisely mimic higher resolution models' structural features. Furthermore, the transferability of Martini scheme makes this force field useful for researching structural differences or different molecular systems. However, molecular modelers must typically modify their original Martini models to better represent a specific molecule's chemistry and architecture¹¹². Martini force field is used in a vast array of applications in varied domains: structural biology^{113–115}, biophysics^{116,117}, biomedicine¹¹⁸, nanotechnology^{119,120}, and materials design^{121,122}. In this context, the increasing challenges in understanding biological processes have led to a growth in efforts towards CG modelling, able to observe computationally demanding phenomena such as agglomeration, phase transitions, and self-assembly that require large time scales and system sizes¹²³. The above-mentioned molecular phenomena are particularly relevant for many biological processes that depend on macromolecules as building blocks for large-scale complexes, such as viruses, ribosomes, and cytoskeletal filaments^{124,125}. Such macromolecular complexes often consist of numerous copies of the same macromolecule, which aggregate through non-covalent interactions into ordered and functional suprastructures^{125,126}.

In this chapter, MD simulations were employed to study the interaction of trodusquemine with the cell membrane and with itself to obtain a detailed understanding of the molecular mechanisms underlying its biological activity. By employing both AA and CG resolutions, we aim to capture a comprehensive picture of the system, spanning from the intricate interplay of individual atoms to the larger structural organization of the molecules. AA simulations provide a high-resolution view of the molecular interactions, while CG simulations offer insights into the overall dynamics and emergent properties of the system. This multi-scale approach enables us to investigate the molecular conformations, and lipid interactions of trodusquemine at the cell membrane, while also providing a deeper insight into the self-assembly and aggregation behavior of the molecule in different physiological environments.

Material and Methods

Trodusquemine Coarse-Grained Model Design

The AA model was used as an input for the generation of the CG trodusquemine model. Building the trodusquemine CG model is a trial-and-error process that follows precise steps (see Figure 10). Starting from AA simulations, the CG coordinates and the topology were

generated by grouping atoms in CG beads¹²⁷. The mapping choice followed three criteria: i) preserve the chemical features, ii) maintain geometric symmetry, iii) reproduce as truthfully as possible the AA conformation of trodusquemine in water and lipid environment.

Trodusquemine consists of a negatively charged sulphate group (head) and a polar sterol ring (core) covalently bound to the positively charged polyamine spermine (tail). Figure 11 shows the trodusquemine mapping: the trodusquemine head was mapped by a Q_a negatively charged bead, the apolar part of the core was mapped by SC_2 beads that mimic the hydrophobic behaviour of the sterol ring, instead the polar hydroxyl group (OH) was mapped by SP_1 bead and finally, the charged tail was mapped by Q_d positively charged beads. CG nonbonded parameters have been taken from the Martini forcefield¹²⁸. In order to generate the CG trodusquemine bond terms (bonds, angles and dihedrals), the AA trajectories and AA-CG maps were used as input for PyCGTOOL¹⁸.

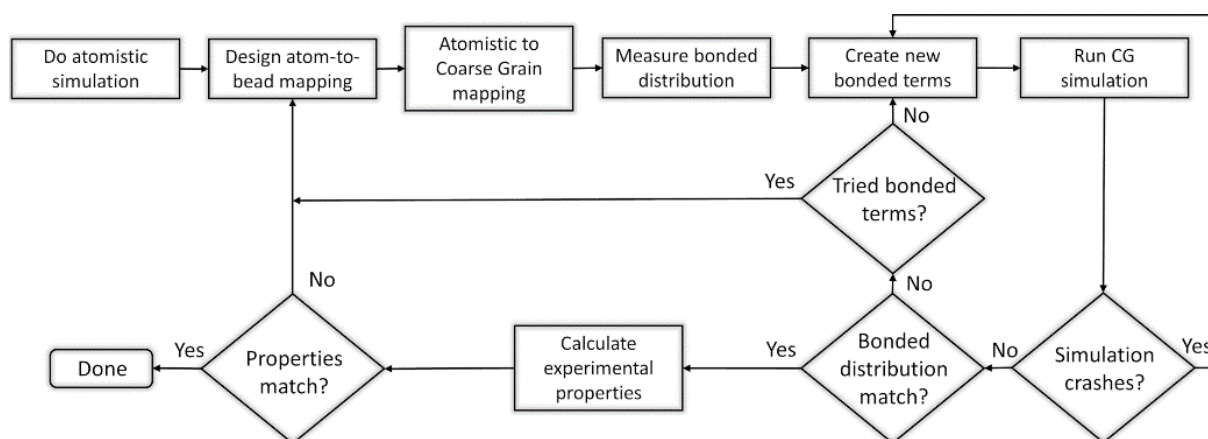


Figure 10: Flow chart of Coarse-Grained model generation.

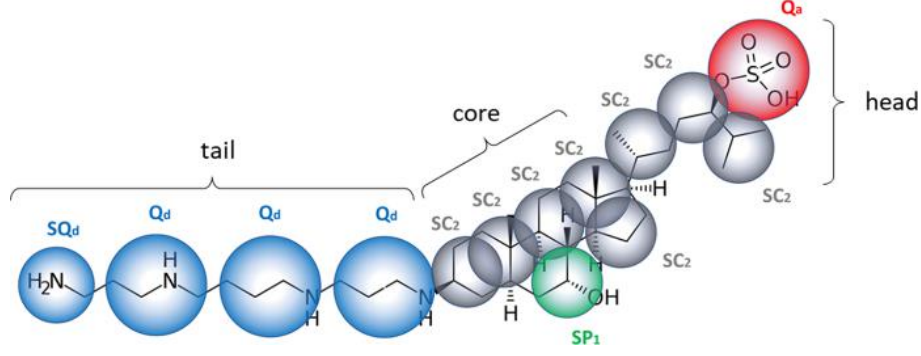


Figure 11: Coarse grain mapping of trodusquemine. The tail was mapped by Q_d beads which are positively charged beads (+1). The core mainly formed by carbon atoms was mapped by SC_2 beads which are apolar beads and the hydrophobic behaviour of hydroxyl group (OH) was mapped by SP_1 bead which is a polar bead. Finally, the trodusquemine negatively charged phosphate group was mapped by a Q_a bead which is a negatively charged bead.

Computational Optimization of Coarse-Grained Force Field Parameters

The correct generation of CG bond parameters is an iterative process that starts from atomistic simulations. In this context, topological CG parameters are obtained by using the iterative Boltzmann inversion^{129,130} (IBI). IBI methodology is based on the Boltzmann inversion^{131,132} equation:

$$V(r) = -k_B T \ln(g(r)) \quad (3.1)$$

Where $V(r)$ is the potential energy, k_B is the constant of Boltzmann, T is absolute temperature and $g(r)$ is the distribution function between a pair of sites. Afterwards that, the iteration process begins, leading to a correction of the starting potential:

$$V_{i+1}(r) = V_i - \alpha k_B T \frac{\ln(g(r)_i)}{\ln(g(r)_{i+1})} \quad (3.2)$$

Where V_{i+1} is the actual potential energy, V_i is the previous potential energy, k_B is the constant of Boltzmann, T is absolute temperature, $g(r)_{i+1}$ is the distribution function and α is called dumping factor which prevents abnormal variation of the potential energy, during the iteration procedure. Extracting the distribution function $g(r)$ from the previous equation, we can obtain the following equivalence:

$$e^{-V(r)/k_B T} = e^{-K(x-\mu)^2/2k_B T} \quad (3.3)$$

$$e^{-(x-\mu)^2/2\sigma^2} \leftrightarrow e^{-K(x-\mu)^2/2k_B T} \quad (3.4)$$

Where $\frac{K(x-\mu)}{2}$ is a generic harmonic potential $V(r)$. The last equivalence is valid only if it is true that the standardized normal gaussian distribution can be considered equivalent to the target functional form, since taking the assumption of independent degrees of freedom, and hence no correlation.

Finally, simply replacing new assumed $g(r)$ in the equation 3.2, it results:

$$V(r) = -k_B T \ln(g(r)) = -k_B T \ln(e^{-(x-\mu)^2/2\sigma^2}) = -k_B T \frac{-(x-\mu)^2}{2\sigma^2} \quad (3.5)$$

$$V(r) = -\frac{K(x-\mu)^2}{2} = -k_B T \frac{-(x-\mu)^2}{2\sigma^2} = \frac{k_B T}{\sigma^2} \quad (3.6)$$

Such particular computing method of Boltzmann inversion ^{131,132} technique is applicable for the estimation of the bonds, angles and constraint topological parameters, only using standard deviation σ to get harmonic constant K . As for IBI computation, even equation 3.6 can be used to implement an iterative modified Boltzmann inversion (ImBI):

$$V_{i+1}(r) = V_i(r) \pm \beta V_i(r) \frac{\sigma_{i+1}^2}{\sigma_i^2} \quad (3.7)$$

Where β is a value with the aim of modulating abrupt variation of potential energy values, like IBI-dumping constant, which range is from 0 to 1. In conclusion, the correction factor $+\beta V_i(r) \frac{\sigma_{i+1}^2}{\sigma_i^2}$ is adopted if σ_{i+1}^2 is greater than σ_i^2 , contrariwise $-\beta V_i(r) \frac{\sigma_{i+1}^2}{\sigma_i^2}$ is selected if σ_{i+1}^2 is lower than σ_i^2 .

Trodesquimine self-aggregation – Simulations set-up

The aggregation propensity of trodesquimine was investigated in two physiological environments following the experimental indications: a water environment and a phosphate buffer. Five molecular systems were built for each molecular environment, consisting of 2, 3, 4, 6, and 8 trodesquimine molecules, respectively. Water molecules were added to a periodic cubic box. The total system charge was neutralized by adding three Cl^- counter ions in case of water environment and adding one PO_4^{2-} plus one PO_4^{1-} molecule in case of phosphate buffer for each trodesquimine molecule, respectively.

The resulting systems were minimized by 1000 steps with the steep descent algorithm. The systems temperatures were equilibrated at 300 K by 200 ps under the canonical ensemble using v-rescale⁸⁷ thermostat (temperature time constant $\tau_T = 0.1$ ps) and at 1 atm by 200 ps under NPT ensemble using Berendsen⁸⁶ isotropic pressure coupling scheme (pressure time constant $\tau_P = 2$ ps). Finally, production MD (2 fs as time step) in NPT ensemble was carried out using the V-rescale thermostat ($T = 300$ K; $\tau_T = 1.0$ ps), together with the isotropic Parrinello-Rahman barostat ($P = 1$ atm, $\tau_P = 2.0$ ps)⁸⁸. The summary of each molecular system investigated was reported in the Appendix (Table A1).

The system analysis was carried out on the last 20 ns of each molecular system. The particle mesh Ewald (PME) method⁹⁰ was used to calculate electrostatic interactions, with a real-space cut-off of 1.2 nm. The van der Waals interactions were calculated by applying a cut-off distance of 1.2 nm and switching the potential from 1.0 nm. The trodusquemine force-field topology was taken from the CHARMM General Force Field⁹² as previously done in literature⁷³. Water and ions topology was parametrized by the CHARMM36m force-field⁹³. All MD simulations were carried out using GROMACS 2020 software package⁹⁴. The Visual Molecular Dynamics (VMD) software was used to monitor all simulation trajectories⁹⁵.

The electrostatic surface potential investigation was performed on representative frames extracted from the ensemble trajectory of trodusquemine complexes. Electrostatic potentials in the presence of phosphates were compared with electrostatic potential calculated for trodusquemine aggregates in water. The electrostatic potentials were computed by the APBS package¹³³. In detail, the non-linear Poisson-Boltzmann equation was applied using single Debye-Huckel sphere boundary conditions on a $129 \times 129 \times 129$ grid with a spacing of 0.7 Å centered at the COM of the molecular system. The relative dielectric constants of the solute and the solvent were set to 4 and 78.4, respectively.

Results and Discussion

Trodusquemine CG Model in water environment

The Martini model employs Lennard-Jones (LJ) potentials to describe the main non-bonded bead interactions, which are calibrated using experimental thermodynamic data such as partitioning free energies. Bonded interactions, on the other hand, are optimized based on bond distributions from reference AA simulations. Consequently, the Martini model incorporates both top-down and bottom-up parameterization methodologies, distinguishing it from other CG

force fields and bottom-up models that aim to reproduce structural features of higher resolution models with high accuracy.

A range of software and approaches have been developed to fine-tune bonded and nonbonded interactions in CG models, including MSCGFM, VOTCA, BOCS, Magic, and PyCG-TOOL. These tools employ techniques such as direct Boltzmann inversion (DBI), iterative Boltzmann inversion (IBI), inverse Monte Carlo (IMC), and the g-YBG equation to optimize interactions while keeping nonbonded ones constant.

Each trodusquemine model (AA and CG) was investigated by three independent MD simulation of 100 ns each and the last 20 ns were used for the validation step. The trodusquemine validation process followed two steps: the topology parameters validation and the conformational dynamics verification. Iterative modified Boltzmann inversion (ImBI) technique¹³⁰ had provided bonded terms that match the reference AA model. The average and standard deviation between AA and CG of bond terms and angle terms are shown in Figure 12A-B. The bond and angle parameters of the CG model are in agreement with the values of the AA model and in accordance with other CG models developed in literature¹²⁷.

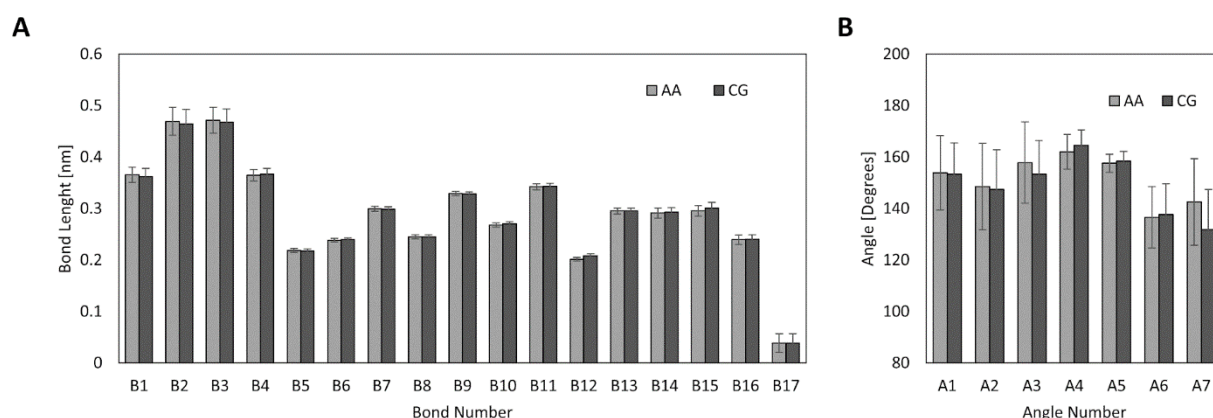


Figure 12: Validation of topological parameters of CG and AA trodusquemine model: A) bond length, B) angle.

Finally, structural conformation of CG trodusquemine models was evaluated in comparison with AA trajectories, by measuring the radius of gyration (R_g) and the solvent accessible surface areas (SASA). Figure 13A-B shows good agreement of radius of gyration and SASA between CG and AA models. In conclusion, the CG model was able to well reproduce the AA behaviour in water environment, assuming similar geometric conformation.

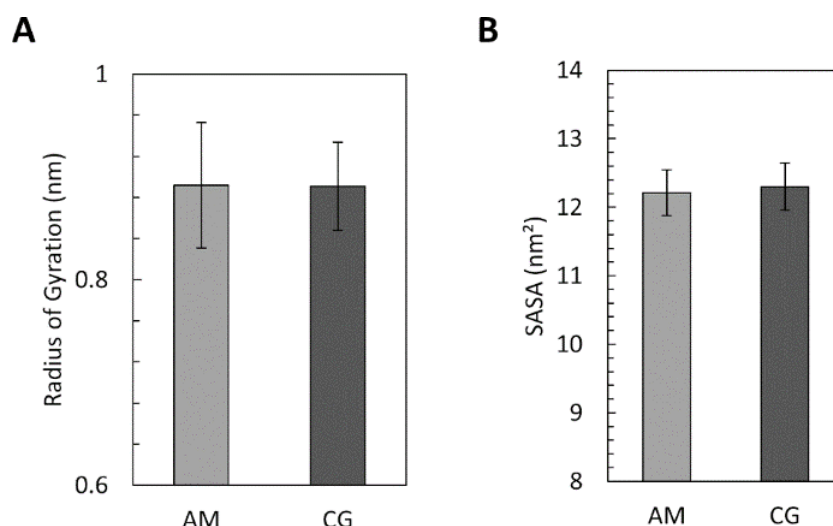


Figure 13: Validation of conformation parameter of trodusquemine molecule: A) radius of gyration, B) SASA.

Trodusquemine CG model in lipid environment

Subsequently, the developed CG model was examined to assess its behavior within a lipid environment, as trodusquemine possesses the ability to partially penetrate the lipid membrane⁷³ owing to its bola-amphiphilic chemical structure. After hundreds of nanoseconds, the trodusquemine CG model built was able to enter in the lipid bilayer. Figure 14 shows a representative snapshot of the CG system. Overall, the trodusquemine CG assumes a ‘boat’ shape, similar of its AA counterpart.

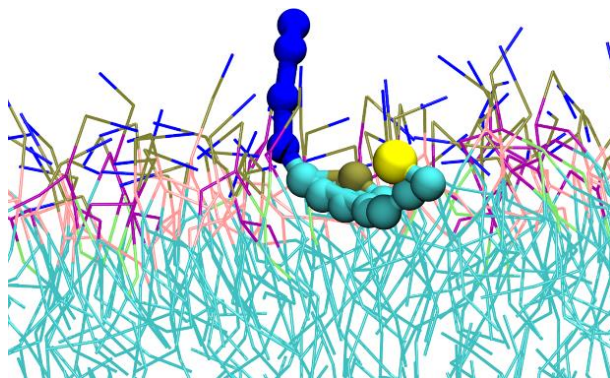


Figure 14: Representative snapshot of trodusquemine CG model (ball and stick representation) within the membrane. The N atoms are in blue, C in cyan, O and S in yellow. The lipids are shown as lines. The water is not shown for clarity.

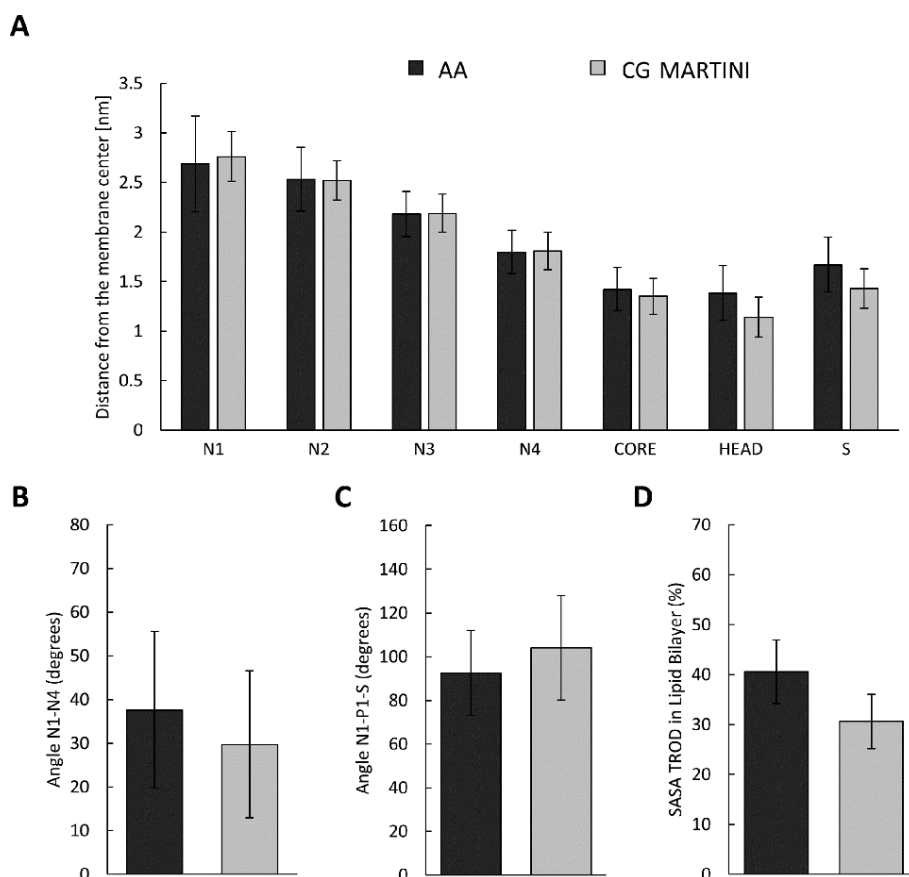


Figure 15: Trodusquimine CG model validation in lipid environment comparing with AA model. A) Distance between representative heavy atoms of the trodusquimine (tail N atoms, core and head) and the membrane centre of mass, evaluated during the MD simulation at the equilibrium. B)-C) Average angle of the tail and head of the trodusquimine. D) Percentage of Solvent Accessible Surface Area of trodusquimine in lipid bilayer.

In particular, the hydrophilic charged tail is exposed to solvent, the hydrophobic core interacts mainly with the lipid bilayer and the negative charged head tries to interact with the water environment (Figure 15 A). The trodusquimine angle conformation of CG model shows a good agreement with the AA model (Figure 15 B-C). Furthermore, the small molecule is positioned at the interface between the hydrophilic and hydrophobic layers, with a SASA of $30.6 \pm 5.4\%$ of the total trodusquimine surface (Figure 15 D), in agreement with the AA model value of $40.6 \pm 6.4\%$ and experimental value of $38 \pm 1.5\%$ ⁷³.

Trodusquimine Self-Aggregation

Trodusquimine, characterized by its asymmetrical bola-amphiphilic chemical structure, features a negatively charged group on one side of the sterol rigid core and a positively charged polyamino functionality at the opposite end. The conformation of the polyamino group is expected to be influenced by pH and salt concentration in solution, as the effective screening

of surface charges by counterions may result in a more coiled conformation of the polyamino tail.

Bolaamphiphiles represent a special type of amphiphiles possessing two hydrophilic groups attached to both ends of an elongated hydrophobic segment. These molecules are known to significantly stabilize lipid membranes¹³⁴ and to form other peculiar self-assembled structures, such as liquid crystalline phases in the bulk state and gels. In aqueous systems bola-amphiles are reported to aggregate in a variety of concentration-dependent structures differing in form and size, typical aggregation structure includes disks, micelles or vesicles, and at higher concentration tubular or helical fibers¹³⁵. Formation of nanowells and pores in lipid membranes represents an additional example of their unique self-assembly capability. More recently, bolaamphiphiles have also been reported to form “bolasomes” either alone or together with fusogenic lipid such as DOPE¹³⁶.

It is thus crucial to investigate trodusquemine aggregation behavior in solution and its interfacial properties, as these features will determine the extent and mechanism of its interaction with amyloid-prone proteins and cellular membranes, inspiring new delivery strategies¹³⁷ and drug formulations. During the drug development process, in silico molecular studies are typically conducted before thoroughly evaluating drug behavior in vitro and in vivo. These investigations are often performed in simplified molecular environments, with proper pH control being a critical parameter. An ideal biological buffer should possess an appropriate pKa and buffer capacity, exhibit high solubility in water, and not negatively affect biochemical reactions.

Although drug candidates undergo rigorous testing for biological activity, possible side reactions with buffer systems are often overlooked. Up until now, there has been a scarcity of research published on this subject, with one notable study examining the adverse interactions between tris(hydroxymethyl)aminomethane (TRIS) and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) buffers and the antitumor drug cisplatin¹³⁸. Nevertheless, systematic exploration into the reactivity of phosphate buffers with specific drugs is lacking.

This knowledge deficiency is particularly disconcerting given that phosphate functions as both an intracellular and extracellular buffering system, and its reactivity with structurally complex drugs might not be readily deducible based solely on the constituent functional groups¹³⁸. Buffers may play an important role in several molecular phenomena including drug-protein

interaction and drug-antibody interaction¹³⁹. As such, a thorough evaluation of the impact of phosphate buffer on trodusquemine aggregation is crucial for advancements in drug development.

The tendency of trodusquemine molecules to aggregate was analyzed in different environmental conditions in terms of electrostatic potential, hydrophobic contribution of SASA and RMSF of the trodusquemine aggregates. In detail, Figure 16 shows the final shape and conformation of the trodusquemine aggregates in the different environments. It is worth mentioning that in both cases, the hydrophobic sterol rings of trodusquemine molecules interact closely with each other's, while the hydrophilic tails are exposed outwards to the solvent. In a water environment, the four trodusquemine system forms initially a cylindrical shaped aggregate, which will be used as nucleation block for the complex elongation. In this connection, the hydrophobic sterols represent the centre of the nucleation block, while the hydrophilic tails are spread towards the centrifugal direction. Increasing the number of trodusquemine molecules, the hydrophobic sterols interact with each other with tails facing outward in multiple directions, resulting in a helical rod shape.

The presence of phosphate buffer strongly stabilizes the assembly. In detail, the sterol rings are still stacked at the centre of the aggregate, but the hydrophilic tails are aligned in two opposite directions. The reason of this behaviour is the electrostatic attraction between the trodusquemine positively charged tails and the negatively charged phosphates. In this context, the final shape of the aggregate is different from the previous case, and it is similar to an ellipsoidal disk.

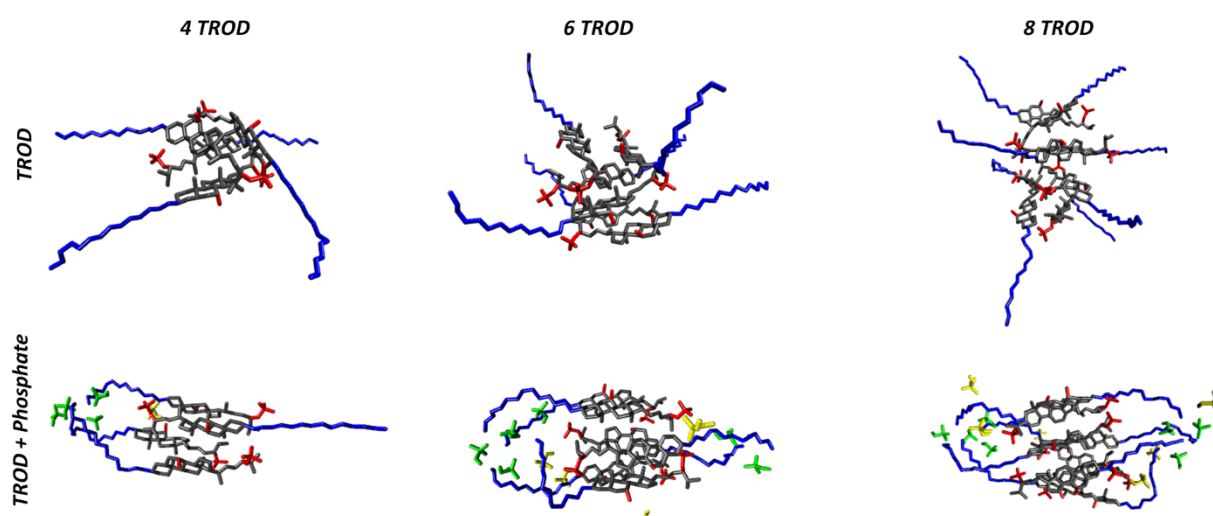


Figure 16: Representative trodusquemine aggregates in water and phosphate buffer (PO_4^{1-} in yellow and PO_4^{2-} in green) at equilibrium composed of 4, 6, and 8 trodusquemine molecules. The trodusquemine spermine tail is shown in blue, the hydrophobic sterol ring in grey, and the head SO_4^- in red. The water molecules are not shown.

In this connection, several structural properties were evaluated to characterize the overall size and shape of trodusquemine aggregates: the radius of gyration (R_g), the ratio of principal moments of inertia (I_i/I_j), and relative shape anisotropy (K^2)¹⁴⁰. The structural analyses of both the 8 trodusquemine aggregates in water and in phosphate buffer environments are shown in Table 4. It is interesting to note that the I_i/I_j values for the 8 trodusquemine aggregate in water are closer to 1 rather than the I_i/I_j values for the 8 trodusquemine aggregate in phosphate buffer. This is a further indication that the aggregate in phosphate buffer have a less spherical shape rather than the aggregate in water. In addition, the K^2 values are also in agreement with the previous analysis, supporting the indication that the aggregate in phosphate buffer have a less spherical shape rather than the aggregate in water. However, the more spherical shape highlighted by the aggregate in water is due to the few trodusquemine used for the simulation. It is reasonable to hypothesize that the elongation of the aggregate will pursuit a final shape similar to a helical rod, as mentioned before.

Table 4: Structural Properties: Radius of Gyration (R_g), Ratio of Principal Moments of Inertia (I_i/I_j), and Relative Shape Anisotropy (K^2) with the respective Standard Deviations of 8 TROD and 8 TROD + phosphate systems. In detail, the more the I_i/I_j values are closer to 1, the more the aggregate is closer to an ideal spherical shape. Moreover, an ideal linear array of atoms is characterized with $K^2=1$, whereas a molecule with ideal spherical symmetry features $K^2=0$

System	$R_g(\text{nm})$	I_1/I_2	I_1/I_3	I_2/I_3	K^2
8 TROD	1.30 ± 0.05	0.78 ± 0.12	0.69 ± 0.11	0.88 ± 0.05	0.014 ± 0.011
8 TROD + Phosphate	1.20 ± 0.04	0.60 ± 0.16	0.53 ± 0.11	0.88 ± 0.07	0.036 ± 0.018

To further inspect the role of phosphate buffer on the trodusquemine assemblies we quantified the surface electrostatic potential (Figure 17). In details, the electrostatic potential of the trodusquemine aggregates is influenced by the two types of environments. Buffer molecules, as well as monatomic ions present in the solution modulate the attractive and repulsive interactions between trodusquemine molecules through adsorption on the aminosterols surface. Overall, the positive charged trodusquemine tails are facing outwards and the inner surface of the trodusquemine aggregates is largely apolar. In the water environment, despite the opposite net charge between negatively charged chlorine ions and positively charged trodusquemine

tails, chlorine ions are unable to coordinate the trodusquimine tails. Instead, the overall surface electrostatic charge of spermine tails is reduced in the presence of buffer solution, allowing them to organize in a more orderly fashion.

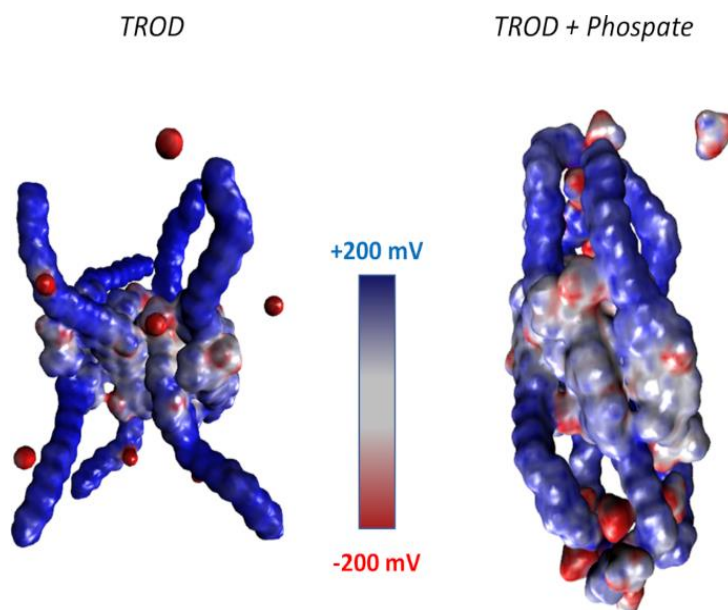


Figure 17: Electrostatic surface map potential of trodusquimine aggregates with Cl^- (left) and phosphate buffer (right).

It is reasonable to postulate that part of the trodusquimine ability to aggregate is attributable to its amphiphilic nature. The trodusquimine amphiphilic nature can be investigated in terms of the percentage of hydrophobic surface exposed to solvent compared to the total surface (Figure 18a). Increasing the size of the trodusquimine assembly from 4 to 8 molecules, the percentage of the hydrophobic surface remains constant suggesting that further increasing the number of trodusquimine molecules, the size of the aggregates would continue to increase. The phosphate molecules modulate the hydrophilic behaviour of trodusquimine macroaggregates, allowing to expose a more significant portion of hydrophobic surface up to 72 % in the case of eight trodusquimine molecules compared to 67.8 % in the absence of phosphates. Furthermore, the RMSF on the nitrogen atoms of trodusquimine tails appears much lower in the presence of phosphates (Figure 18b), suggesting a stabilized role of the latter. In a similar way, previous studies observed the increased stability of the bovine serum albumin and interleukin in the presence of TRIS ^{141,142}. It is reasonable to hypothesize that the aforementioned modulations of physicochemical properties could impact the assembly size, allowing larger aggregates to be formed.

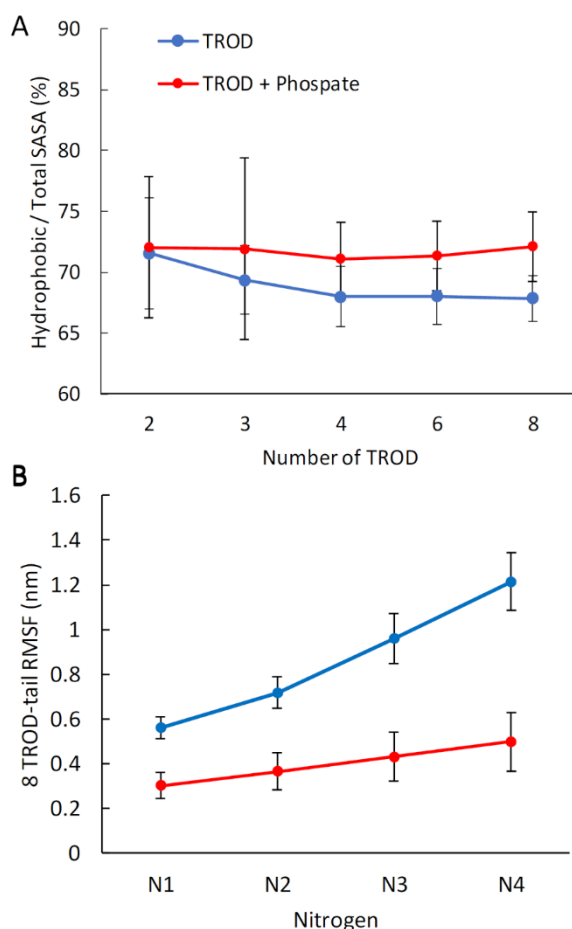


Figure 18: (a) Percentage of hydrophobic SASA of trodusquimine in water and water plus phosphate as a function of the aggregate dimension. (b) Root mean square fluctuation nitrogen trodusquimine tail in the water and water plus phosphate for the aggregate composed of 8 trodusquimine molecules.

However, phosphates do not participate equally in the modulation of the properties of trodusquimine aggregates. Figure 19 shows the number of hydrogen bonds (h-bonds) between each type of phosphate molecule and trodusquimine, normalized to the total number of trodusquimine. Increasing the number of trodusquimine molecules, the h-bonds involving PO_4^{2-} increase up to 2.7 average h-bonds for each trodusquimine. Instead, the h-bonds between trodusquimine aggregates and PO_4^{1-} tend to be constant as the aggregate grows.

This observation is consistent with prior research, which highlights the importance of hydroxyl groups in TRIS for forming hydrogen bonds with several amino acids present in proteins and peptides, including glutamic acid, aspartic acid, alanine, glycine, tryptophan, and cysteine, thus safeguarding them from chemical degradation¹⁴³. Additionally, previous studies have reported that enhancing TRIS concentrations can lead to increased hydrogen bonding capacity¹⁴³, which further corroborates our findings on the distinct involvement of phosphates in the modulation of trodusquimine aggregates.

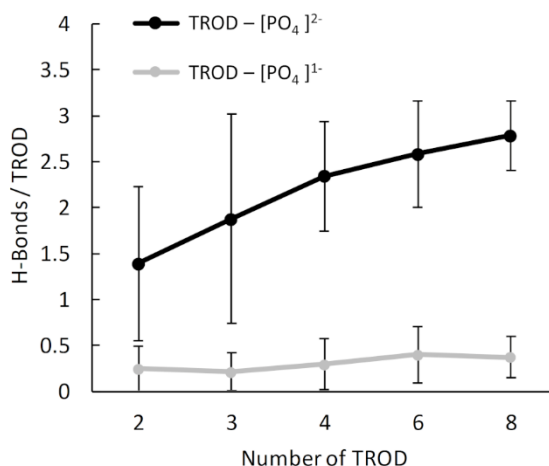


Figure 19: Number of hydrogen bonds between trodusquimine and phosphate (PO_4^{1-} in grey and PO_4^{2-} in black) normalized on the total number of trodusquimine as a function of the trodusquimine aggregates dimension.

Trodusquimine Self-Aggregation: coarse-grained description

The trodusquimine CG model developed is essential to investigate the self-aggregation phenomena, which require long time scale and large molecular systems. Three molecular systems were built consisting of 8, 20 and 50 trodusquimine respectively. Water molecules were added to a periodic cubic box and the total system charge was neutralized adding Na^+ and Cl^- ions at a concentration of 150 mM. Then three independent MD simulations of 2 μs were carried out. During MD production the trodusquimine can form aggregates with a rod shape (Figure 20B). Increasing the number of trodusquimine, the rod length increases. The total Solvent Accessible Surface Area (SASA) of a single molecule in water environment is $10.38 \pm 0.14 \text{ nm}^2$ of which $5.84 \pm 0.07 \text{ nm}^2$ is the hydrophobic contribution. Figure 20A shows the hydrophobic SASA of trodusquimine, normalized on the total number of molecules involved in the aggregate, as a function of total molecules in the MD simulation. Increasing the number of trodusquimine, the relative hydrophobic SASA tends to decrease. The rod shape ensures a relative hydrophobic SASA reduction equal to 31.2%, 39.6% and 42% for an aggregate composed of 8, 20 and 50 trodusquimine respectively compared to a single trodusquimine in water environment.

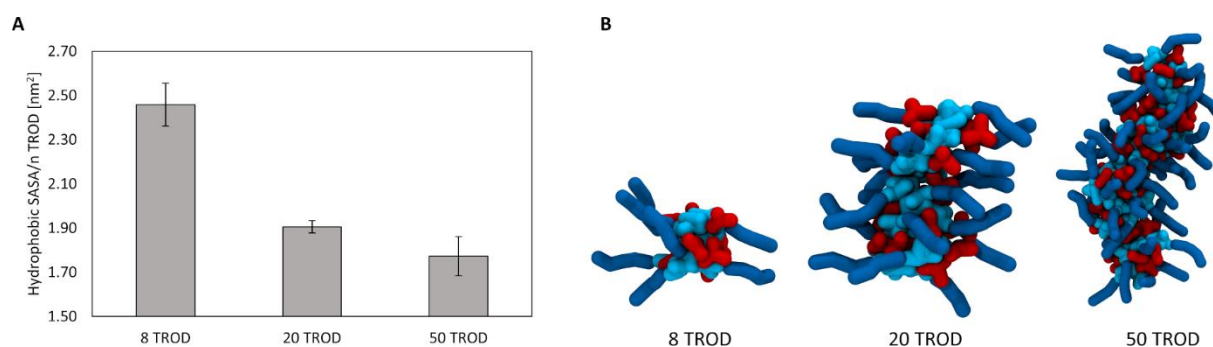


Figure 20: A) Relative SASA of trodusquimine in water environment during the self-aggregation process. B) Trodusquimine rod aggregates, increasing the number of interacting molecules.

Conclusions

The present study offers a comprehensive examination of the molecular mechanisms that underlie the biological behaviour of trodusquimine by employing a multi-scale approach that incorporates both AA and CG resolutions. Utilizing the widely adopted CG Martini model, we are able to delve into the molecular conformations and lipid interactions of trodusquimine at the cell membrane, as well as thoroughly investigate its self-assembly and aggregation behavior across a range of physiological environments.

Furthermore, we highlight the crucial role played by buffers, particularly phosphates, in modulating the properties of trodusquimine aggregates. Our analysis emphasizes the differential involvement of specific phosphate species in hydrogen bond formation, thereby shedding light on the complex interplay between molecular structure and environmental conditions. The phosphate molecules modulate the hydrophilic behaviour of trodusquimine macroaggregates, allowing to expose a more significant portion of hydrophobic surface. This understanding of phosphate's role in stabilizing and modulating the trodusquimine aggregates may have important implications for drug development and optimizing molecular interactions.

Finally, our findings reveal a strong agreement in structural conformation between the CG and AA models, with the CG model capably replicating AA behavior within a water environment. This successful replication of behavior demonstrates the applicability and robustness of the Martini model in biomolecular modeling. The CG model developed will be able to be used to study numerous other phenomena of biological interest, overcoming the accessible time scale limitations of AA resolution.

3.4. Phase Separation Triggers Trodusquemine Action

Introduction

Membranes are vital elements of all cells, establishing the cell's uniqueness and specifying an extensive array of internal compartments. Conventional cell membranes are densely populated with proteins occupying an estimated membrane surface up to 30% and can comprise hundreds of diverse lipids, unevenly dispersed between the two lipid layers^{144,145}. The fluid-mosaic model describes the cellular membrane as a fluid and its components are characterized by a random distribution, resulting in lateral and rotational freedom. Lipids and proteins engage in a captivating and dynamic interaction within cell membranes. In this intricate protein-lipid interplay, lipids facilitate the movement of proteins, while proteins contribute to the mobility of lipids. This reciprocal relationship ensures the proper functioning of cellular processes and maintains the structural integrity of the membrane.

The heterogeneity of plasma membrane components means that a number of molecular events can occur, including the phase separation of the lipid components. Phase separation in complex liquids and biological fluids sometimes occurs, enabling the formation of compartments with unique dynamic properties and structural patterns¹⁴⁶, that have been named lipid raft¹⁴⁷. The lipid raft hypothesis emphasizes the importance of lipids in cellular functions such as DNA repair, membrane trafficking and signal transduction. Additionally, lipid rafts have been linked to the aggregation of intrinsically disordered proteins, a crucial process for the formation of cellular condensates that are associated with neurodegenerative diseases^{148,149}. Lipid rafts play a critical role in human physiology and the pathogenesis of diverse diseases. Numerous studies have extensively documented the importance of lipid rafts in neurodegenerative diseases such as Alzheimer's and Parkinson's diseases¹⁵⁰. For example, the amyloid precursor protein (APP), a transmembrane protein predominantly found in neurons, is not inherently a raft protein. However, after APP palmitoylation, a considerable proportion of APP tends to localize within lipid rafts¹⁵¹. APP cleavage is generally believed to be modulated within the lipid raft microenvironment¹⁵², and these microdomains contain APP-derived proteolytic fragments as well as enzymes involved in APP amyloidogenic processing. Furthermore, the production of amyloid-beta (A β) peptides is preferentially localized within lipid rafts¹⁵³, emphasizing the importance of these microdomains in the context of APP processing and neurodegenerative diseases¹⁵⁰.

In this context, it has become increasingly evident that the aggregation processes of proteins or peptides implicated in neurodegeneration are accelerated by their binding to lipid membranes, further emphasizing the essential role lipid rafts play in the progression of these diseases. Nowadays, the modulation of lipid raft properties has emerged as a promising avenue for developing additional therapeutic strategies against neurodegenerative diseases¹⁵⁴. In this context, recent studies using atomic force microscopy images have shown the trodusquemine ability to modulate phase separation in LUVs on supported lipid bilayer. However, the investigation of lipid rafts through experimental techniques has presented considerable challenges in provide detailed lateral membrane organization due to their dynamic nature and small size¹⁴⁴.

In this context, molecular dynamics (MD) simulations provide a deeper understanding of the interplay between lipid-raft and molecular components. In silico methodologies allow for a more detailed and controlled analysis of lipid raft behavior and interactions, providing valuable insights that complement experimental findings and enhance our knowledge of their significance in cellular biology.

In this chapter, the interaction of the trodusquemine molecule with a membrane that exhibits phase separation is evaluated. In detail, we studied a biphasic membrane containing a disordered phase and a gel-phase domain, which in cells are generally referred to in the literature as lipid-rafts (Ld-So)¹⁵⁵. However, demixing of lipids into phases separated by means of atomistic MD simulations is hard to achieve due to size and time scale limitations¹⁵⁶. Instead, MD simulations employing Martini CG force field, have shown phase separation events on time scales from micro to milliseconds, providing structural insight to interleaflet domain interaction¹⁵⁷.

Material and Methods

Molecular dynamics simulations – lipid bilayer

The interaction of trodusquemine molecules with lipid raft was evaluated by MD simulations. Five types of molecular systems were investigated: a phospholipid bilayer constituted by 400 lipids and the same bilayer in presence of 1, 2, 3, 4, and 5 trodusquemine molecules, respectively. A fully hydrated lipid bilayer of dibehenoyl-phosphatidylcholine (DBPC) and dilinoleyl-phosphatidylcholine (DIPC) with a mole ratio of 3:1 was created using the bilayer builder insane¹⁵⁸. DBPC has long saturated chains (22:0), whereas DIPC's are shorter and

polyunsaturated (18:2). The gel-to-liquid transition temperature of DBPC and DIPC are 348 K and 216 K respectively (Figure 21). We have chosen DBPC and DIPC as lipids components of a toy model membrane because have gel-to-liquid transition temperature well separated, accelerating the phase separation process. To mimic the physiological environment, 150 mM NaCl was added, along with an excess of 3 Cl⁻ ions for each trodusquemine molecule to neutralize the net system charge. The system was energy-minimized by the steepest descent method for 1500 steps and then equilibrated on NPT ensemble for 0.5 ns using V-rescale thermostat⁸⁷ and Berendsen⁸⁶ semi-isotropic barostat as coupling methods. Each type of lipid is coupled separately in temperature as recently suggested in phase separation studies¹⁵⁶. Each molecular system was subjected on 2 μ s of MD production in the NPT ensemble: the temperature was kept at 293 K using velocity rescaling thermostat ($\tau_T = 1.0$ ps) and the pressure at 1 bar using Parrinello-Raham barostat ($\tau_P = 12$ ps)⁸⁸. The electrostatic interactions were evaluated by particle mesh Ewald (PME) method using 1.2 nm as cut-off⁹⁰. The van der Waals interactions were calculated by applying a cut-off distance of 1.2 nm and switching the potential from 1.0 nm. MARTINI force-field topology was used to parametrize the molecular systems¹⁰, choosing the polarizable water model. All MD simulations were carried out using the GROMACS 2020 software package⁹⁴.

Molecular dynamics simulations – lipid monolayer

In order to verify whether trodusquemine is able to interact with a lipid monolayer, two trodusquemine molecules were inserted in the water slab environment in vacuum with two symmetric monolayers at the two water- vacuum interfaces. The molecular system was built starting from a lipid bilayer, using a bilayer builder. The monolayers consisted of the same bilayer lipid composition: DBPC DIPC with a mole ratio of 3:1. The total system charge was neutralized adding Na⁺ and Cl⁻ ions at a concentration of 150 mM in the water slab. The system was energy-minimized by the steepest descent method for 1500 steps and then equilibrated at 293 K and 1 bar on NPT ensemble for 0.5 ns using V-rescale thermostat⁸⁷ and Berendsen⁸⁶ semi-isotropic barostat as coupling methods. Finally, the molecular system was investigated by 3 μ s of MD production in the canonical ensemble at 293 K.

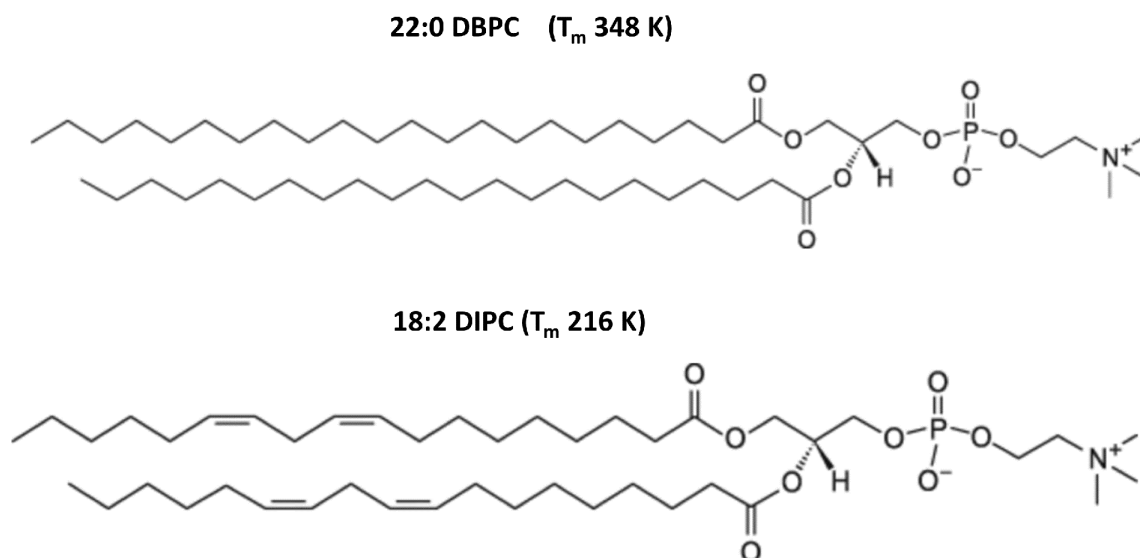


Figure 21: Types of lipids used to investigate the lipid bilayer/monolayer in the presence of phase separation. The dibehenoyl-phosphatidylcholine (DBPC) has a transition temperature of 348K and the dilinoleyl-phosphatidylcholine (DIPC) has a transition temperature of 216K.

Results and Discussion

The following results are derived from a comprehensive study that integrates both experimental and computational approaches to elucidate the intricate behavior of lipid-raft in cell membranes and their interactions with trodusquemine. For the experimental investigation, the study of real cell membranes is difficult to achieve for its complexity and to perform a more in-depth study artificial biomimetic models, such as Langmuir monolayers, are commonly used. Specifically, the experimental investigation was conducted on lipid monolayers, a mimetic model that is simple enough to be accessible by physical measurement¹⁵⁹. To complement the experimental data, MD simulations were carried out to examine the behavior of lipid membranes in greater detail.

In the first stage of the computational analysis, MD simulations were conducted on double layers, offering a more complex model of lipid membrane behavior. Subsequently, the simulations were extended to monolayers, providing an opportunity to assess the representativeness of the double-layer model in capturing the essential characteristics of monolayer systems. This comparative analysis was deemed necessary because the monolayer simulations were found to be prone to certain computational stability issues. Furthermore, the bilayer model was used since lipid monolayers are incapable of accurately describing molecular interactions at gas/liquid interfaces, especially when it comes to computing electrostatic interactions^{128,160}.

Trodusquemine and Lipid Raft Interaction in lipid bilayer

We have built a fully hydrated lipid bilayer of DBPC and DIPC with mole ratio of 3:1. In particular, Figure 22A and B provide detailed visual representations of the initial system configuration from both lateral and top perspectives, highlighting the structural intricacies of the system in question. In these diagrams, it is evident that the two distinct types of phospholipids are dispersed randomly across the membrane plane. Initially, the two well mixed lipid types result in an area per lipid of $61.8 \pm 0.21 \text{ \AA}^2$ as observed in Figure 22C. As the MD production progresses, the phospholipids demonstrate an inherent ability to reorganize themselves, resulting in a dynamic rearrangement of their spatial distribution within the membrane. After 700 ns of MD simulation, the area per lipid converges to $52.25 \pm 0.17 \text{ \AA}^2$ (Figure 22C). Furthermore, the area per lipid provides valuable information on the stability of the lipid membrane and its overall organization.

The final lipid distribution, as seen in Figure 22D-E, exhibits a clear phase separation characterized by the clustering of each type of phospholipid into either the Liquid Order (Lo) or Liquid Disorder (Ld) phase. The observed phase separation is further supported by the hexagonality parameter, as shown in Figure 22F. This metric serves as a quantifiable measure of lipid packing and organization, and its analysis corroborates the presence of distinct Lo and Ld phases.

In accordance with the raft-hypothesis¹⁶¹ and previous computational studies¹⁶² the Lo phase predominantly consists of saturated DBPC lipids, which are known for their role in forming tightly packed and ordered regions within the membrane. In contrast, the Ld phase is primarily composed of polyunsaturated DIPC lipids, which impart a more disordered and fluid character to the membrane.

The observed lipid reorganization is a complex process, involving a delicate interplay between various factors such as lipid-lipid interactions and the influence of environmental conditions on the membrane overall stability.

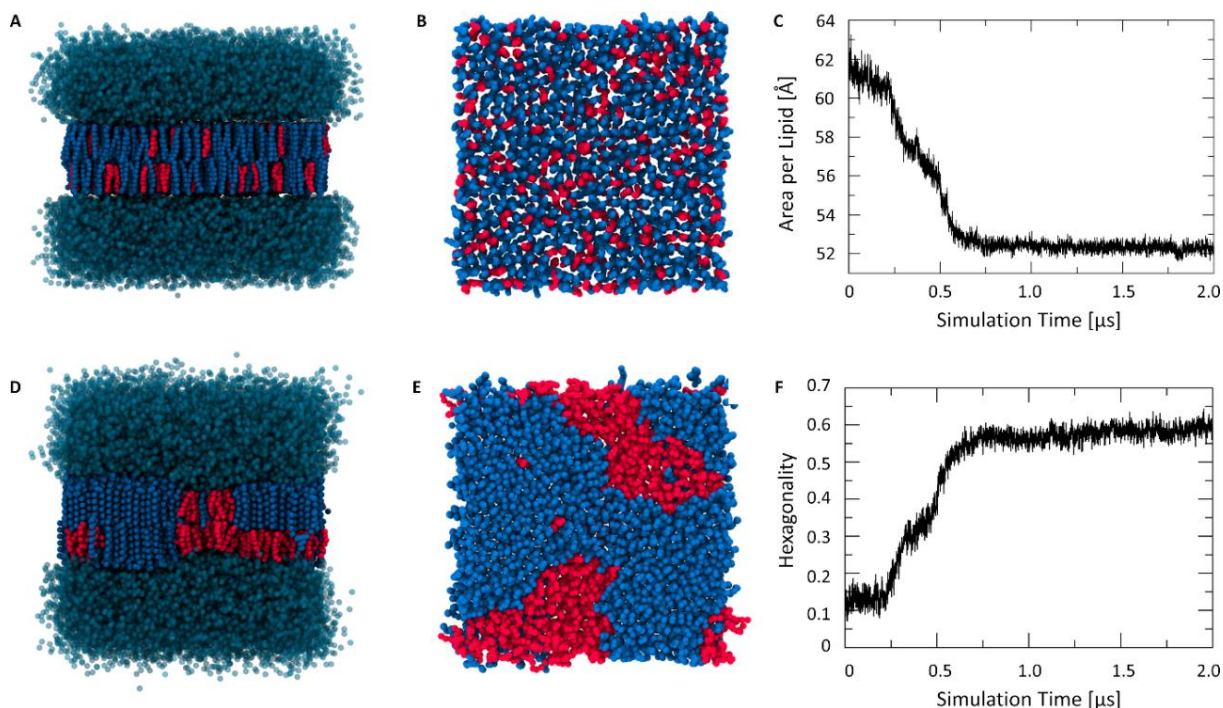


Figure 22: Initial configuration of DBPC (blue) and DIPC (red) random oriented: A) lateral and B) top view. C) Area per lipid during the MD production. Final configuration of MD production with phase separation in D) lateral and E) top view. F) Hexagonality parameter to monitor the phase separation process.

The phase-separated system, characterized by Lo and Ld regions, was employed to investigate whether the trodusquemine maintains its ability to interact with the lipid membrane under such conditions. As illustrated in Figure 23B, the phase separation results in a pronounced difference in thickness between the lipids occupying the Lo state and those in the Ld state. This disparity in membrane thickness can be attributed to the distinct physicochemical properties of the lipids in each phase, with the Lo phase characterized by tightly packed and ordered saturated lipids, while the Ld phase comprises more loosely arranged and disordered polyunsaturated lipids.

The considerable thickness gap between the two lipid types leads to a unique phenomenon where a hydrophobic portion of DBPC becomes exposed to the solvent. This exposure can potentially impact the membrane's structural properties, as well as the behavior of embedded proteins and small molecules, such as trodusquemine, in the lipid environment.

Despite the distinct phase separation observed in the cell membrane, trodusquamine preserves its capacity to enter and interact with the membrane, as observed during MD simulations. Furthermore, these simulations reveal that trodusquamine exhibits a preference for associating with the boundary between the Lo and Ld phases, as illustrated in Figure 23A. In particular, trodusquamine is able to interact with the hydrophobic surface exposed to solvent of DBPC (Lo state) along the phase separation boundary (Figure 23A,C). Experimentally, trodusquamine modulates the coexistence of the liquid-gel and liquid-disorder phase⁷³. The MD simulations reveal that trodusquamine molecules interacting with the edge emphasize the thickness mismatch, attaining values of 1.68 nm when four trodusquamine molecules are present (Figure 23B). The trodusquamine ability to influence the equilibrium between these phases suggests that it may possess the capacity to alter the physicochemical properties and overall stability of the lipid membrane.

With a growing up of trodusquamine in simulation, the effect does not change, increasing the interactions with the hydrophobic surface of DBPC, reaching up to 6% of the total hydrophobic surface covered (Figure 23C).

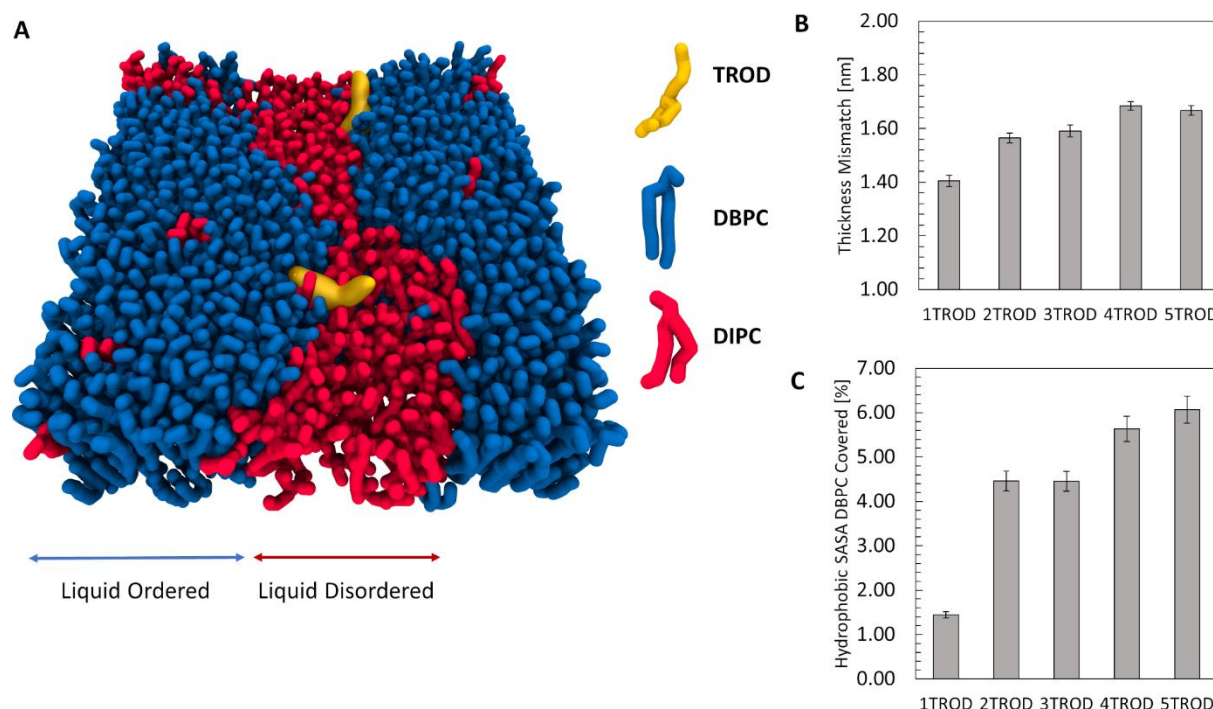


Figure 23: A) Final configuration of trodusquamine localized at the phase separation boundary. B) Thickness of DBPC and DIPC lipids. C) Percentage of hydrophobic SASA of DBPC covered by the trodusquamine interaction as a function of the number of trodusquamine considered.

In the context of AD, neuronal vulnerability to A β -induced toxicity is contingent upon A β binding to cell membranes and its presence in lipid rafts¹⁶³. Elevated levels of soluble A β dimers have been identified in lipid raft fractions of human and transgenic mouse model AD brains¹⁶⁴. Notably, A β accumulation in presynaptic terminals within the AD cortex colocalises with lipid raft markers such as cholesterol and ganglioside GM1¹⁶⁵. These findings collectively suggest that A β aggregation within lipid rafts may contribute to AD neuropathology. Furthermore, the abnormal composition of lipid rafts in AD brains, characterized by increased order and viscosity¹⁶⁶, highlights the potential for modulating lipid raft composition as a therapeutic strategy in addressing AD-related neurodegeneration¹⁶⁷.

Trodesquimine and Lipid Raft Interaction in lipid monolayers

Furthermore, the ability of trodesquimine to interact with a lipid monolayer was tested, inserting two trodesquimine molecules in a water slab environment between the two monolayers. Figure 24a shows the starting configuration of the MD simulation.

During 3 μ s of MD simulation, the ability of trodesquimine to interact with the phase separation edge agrees with the molecular events observed in presence of a phospholipid bilayer (Figure 24B). In this context, the lipid bilayer is also well representative of trodesquimine behavior in the presence of phase separation of a lipid monolayer.

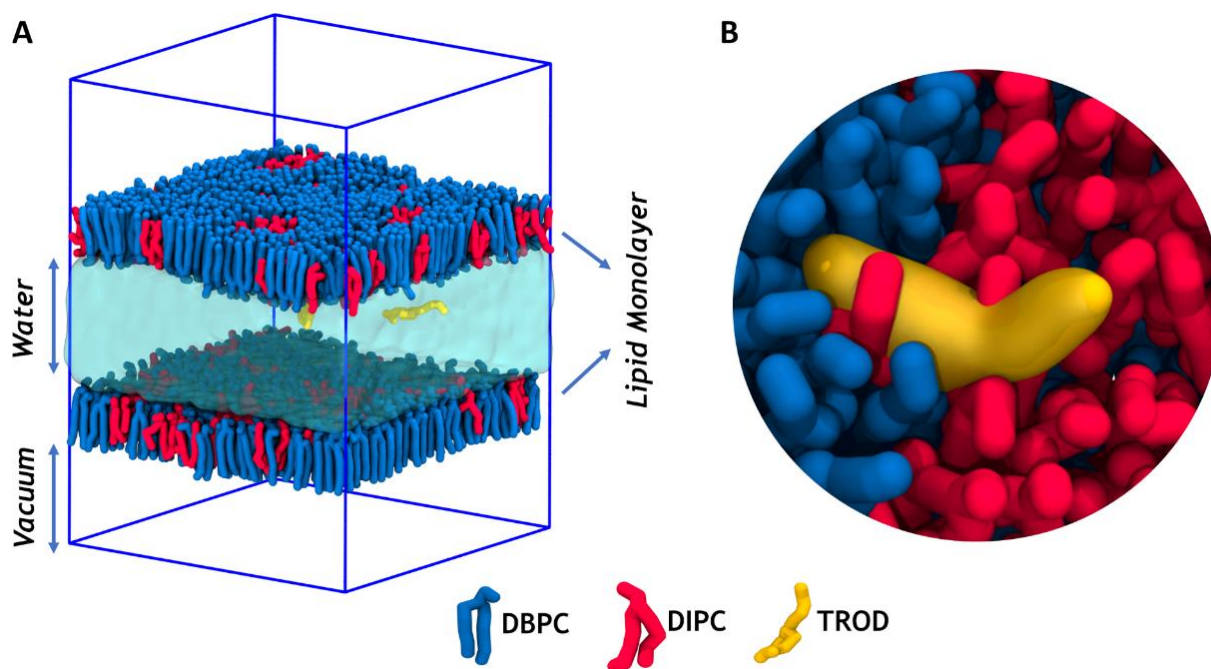


Figure 24: (a) Starting molecular system configuration composed by two monolayers of DBPC (blue molecules) and DIPC (red molecules) in a mole ratio of 3:1, a water slab (cyan color) and vacuum interfaces. Initially the Trodusquimine molecules (yellow color) are localized in water environments. (b) Final configuration of Trodusquimine localized at the phase separation boundary of lipid monolayer.

Both the computational and experimental approaches consistently depict trodusquimine as localizing along the periphery of the Lo domains. This behavior can be attributed to the preferential adsorption of trodusquimine at the line tension regions of the Lo domain borders, which is facilitated by the favorable interactions between the sterol moiety of trodusquimine and the exposed hydrophobic segments of the lipids. Furthermore, the trodusquimine localization along the border of the Lo domains has important implications for the stability and dynamics of lipid rafts. By interacting preferentially with the exposed hydrophobic regions at the interface between the two phases, trodusquimine influences the overall line tension and subsequently impact the size, stability, and organization of lipid rafts^{73,168}.

Conclusions

In conclusion, our study has shed light on the interaction of trodusquimine with lipid rafts. The dynamic nature and small size of lipid rafts have posed significant challenges to experimental investigations, prompting the use of MD simulations to provide a more comprehensive understanding of their behavior and interactions. In this study, we utilized MD simulations to investigate the interaction of trodusquimine molecules with a biphasic membrane exhibiting phase separation. Our computational analysis involved the examination of five distinct

molecular systems, including a phospholipid bilayer composed of 400 lipids and the same bilayer in the presence of 1 to 5 trodusquimine molecules. We initially conducted MD simulations on double layers, which allowed us to model the complex behavior of lipid membranes. Subsequently, we extended our simulations to monolayers, enabling us to assess the applicability of the double-layer model to monolayer systems.

The MD simulations revealed that trodusquimine retains its ability to enter and interact with the membrane despite the evident phase separation. Notably, the simulations demonstrated that trodusquimine exhibits a preference for associating with the boundary between the Lo and Ld phases. This preference suggests that trodusquimine may modulate the coexistence of the liquid-gel and liquid-disorder phases, ultimately altering the physicochemical properties and overall stability of the lipid membrane.

Our findings indicate that trodusquimine ability to influence the equilibrium between these phases is potentially due to its interaction with the hydrophobic surface exposed to the solvent of DBPC in the Lo state along the phase separation boundary. As the concentration of trodusquimine in the simulation increases, its interactions with the hydrophobic surface of DBPC continue to rise, covering up to 6% of the total hydrophobic surface. These insights into trodusquimine interactions with lipid rafts not only enhance our understanding of their role in cellular biology but also contribute to the development of novel therapeutic strategies against neurodegenerative diseases. Our study highlights the potential of *in silico* methodologies to complement experimental findings and provide a more detailed and controlled analysis of lipid raft behavior and interactions. These observations and the understanding of the basic physiochemical driving forces for lipid mixing may have important implications for understanding the protective role of trodusquimine against the binding of aberrant misfolded proteins, as well as its potential involvement in numerous signal transduction pathways in cells that are associated with lipid rafts.

Finally, it is essential to highlight that the employment of a simplified lipid raft model might not wholly mirror the multifaceted nature of lipid rafts. The integration of a more comprehensive and accurate lipid raft model in future studies promises to refine our understanding of lipid raft dynamics and their interactions with trodusquimine.

3.5. Enhanced sampling for free energy landscape of membrane insertion of aminosterols

Introduction

As life expectancy continues to rise, neurodegenerative diseases (NDs) are becoming an increasingly significant burden on society, affecting approximately 15% of the worldwide population¹⁶⁹, and hence a growing challenge for healthcare systems¹⁷⁰. NDs encompass a diverse group of progressive disorders, such as Alzheimer's disease (AD), and Parkinson's disease (PD), which primarily impact neuronal function within the nervous system, resulting in degeneration, eventual cell death, and severe cognitive and motor deficits. These relentless conditions affect essential functions like movement, respiration, speech, and cognition. Despite distinct symptoms, several common pathological features such as accumulation and aggregation of misfolded proteins, neuroinflammation, and oxidative stress, are shared. Furthermore, the underlying process of protein misfolding, its intermediate states, end-products, and principal characteristics show remarkable similarities in various NDs, despite the differences in the specific composition of the implicated protein aggregates³⁷. Currently, there are no definitive cures for NDs, and available treatments primarily aim to alleviate symptoms and improve the quality of life for affected individuals. In June 2021, the FDA granted conditional approval to Aduhelm (aducanumab), the first antibody targeting the neurotoxic amyloid-beta (A β) peptide against specific cases of AD. However, Aduhelm has been linked to adverse side effects and as is typical for monoclonal antibody therapies, it carries a substantial financial burden.

In this context, a promising approach to combat NDs involves the development of anti-amyloid molecules, which are designed to modulate the aggregation kinetics of toxic misfolded proteins that contribute to the neurodegenerative progression⁶⁴. In recent years, a novel class of aminosterol compounds with antimicrobial activity, initially discovered and identified in the digestive tracts of dogfish sharks (*Squalus acanthias*), has generated optimism in the search for effective treatments for NDs including AD and PD. Among the most studied aminosterols are squalamine and trodusquemine that have demonstrated promising biological activities beyond their antimicrobial properties, including anti-inflammatory, antiviral, and anticancer effects.

Squalamine was initially observed to inhibit lipid-induced nucleation of α -synuclein, which is associated with PD progression, resulting in reduced concentrations of α -synuclein oligomeric

aggregates⁶⁷. Furthermore, the restoration of the locomotor abilities of *C. elegans* overexpressing human α -synuclein when treated with squalamine was observed⁸¹. Currently, squalamine is undergoing Phase 2 clinical trial for PD constipation and in Phase 1 for PD dementia⁷³. Trodusquemine, effective at lower doses compared to squalamine^{72,77}, it has been demonstrated to traverse the blood-brain barrier following intraperitoneal administration⁷⁸. Furthermore, trodusquemine promotes the regeneration of injured tissues and organs, while leaving uninjured tissues unaffected, as evidenced in studies conducted on zebrafish and mice⁷⁹. Recent in vivo investigations have shown the trodusquemine ability of preventing hippocampal neuron loss and spatial memory deficits, reducing plaque size, and inhibiting neuronal death in AD mouse model⁸⁰. Subsequently, trodusquemine has been proposed to promote the secondary nucleation of A β ₄₂⁷², and to inhibit the lipid-mediated nucleation and fibril amplification of α -synuclein⁷¹. Indeed, trodusquemine does not alter the physicochemical properties of oligomers at physiological concentrations; rather, it influences their assembly kinetics^{72,77} by displacing the oligomers from the cell membrane, increasing the lipid self-diffusion and the mechanical resistance force, and reducing the natural membrane negative charge⁷³.

In this regard, both aminosterols have been demonstrated to interact with the neuronal cell membrane, thereby offering protection against the toxicity induced by oligomers responsible of AD and PD. Examining the molecular interactions between the aminosterol molecules and the cell membrane contributes to a deeper understanding of the aminosterol molecular mechanism of action. This knowledge is critical for the development of effective pharmacological interventions and the optimization of treatment outcomes. In the drug discovery pipeline, accurately estimating ligand-binding affinity is crucial as it facilitates various steps, including structure-based drug design and lead optimization. Modern advances in experimental methods have made it possible to collect a wealth of information on the membrane interactions of various substances^{171,172}. However, determining the specific mechanisms by which molecules may traverse membranes remains difficult. Molecular dynamics (MD) simulations add quantitative atomic-level insight into chemical interactions with lipid bilayers providing valuable information to experimental data, where the latter is often inaccessible by experimental techniques^{73,98,168,173}.

To provide a thorough understanding of aminosterol-lipid membrane interactions, this section will evaluate the interactions of trodusquemine and squalamine with neuron-like membranes. The goal is to obtain quantitative information to estimate fundamental properties, such as

binding affinity. This ambitious objective is accomplished by combining molecular dynamics simulations and advanced sampling techniques.

Materials and Methods

Martini Force Field

The Martini 2.2p force field¹⁷ was used to represent the molecular components. The concept is based on a four-to-one mapping, in which a single interaction center named ‘bead’, represents four heavy atoms and their associated hydrogens. Each bead has a number of subtypes, allowing for an optimal balance between computational efficiency and chemical representation. The three-bead polarizable water bead was employed to correctly reproduce the orientational polarizability of real water¹⁴ and the ions were represented by a single CG bead. The aminosterols mapping was obtained from an automated approach that preserve the symmetry of the molecules, and correctly reproduce the experimental octanol-water and water-membrane partitioning data²¹. Finally, the bond terms of squalamine and trodusquemine molecules were parametrized employing swarm-cg tool¹⁹ (see Appendix A2 section).

OPES Expanded and OPES MetaD

To enhance the sampling the On-the-fly Probability Enhanced Sampling (OPES) method was employed²⁶. OPES belongs to adaptive bias methods and is a development of metadynamics^{25,29} in which Gaussian kernels are utilized to rebuild the marginal probability distribution along the CVs as opposed to directly creating the bias potential. The equilibrium probability distribution $P(s)$ is estimated on the fly using OPES and the bias is selected to drive the system toward a target distribution. By selecting the proper target distribution, one may generate random samples from a wide range of ensembles, including the well-tempered²⁹ or a generalized ensemble¹⁷⁴.

In this work, two different distributions are targeted. In a first set of simulations, the multithermal distribution is sampled to allow rapid exploration of the thermodynamic states accessible to the system through the OPES expanded method. In detail, a temperature range between 270K and 500K was explored. Next, the well-tempered distribution is sampled using the deep-TICA-1 vector as CV employing the OPES MetaD method. An adaptive kernel width was employed, setting the minimum value that can be achieved to 0.028. The maximum energy barrier that can be overcome was established at 60 kJ mol⁻¹, and the deposition rate was set to

occur every 500 simulation steps. For each molecular system, 8 replicas of 400 ns each were simulated for a total of 3.2 μ s of simulation time. In addition, the replicas shared the same bias potential in order to harvest more transitions between the ligand's states in and out the lipid bilayer.

Deep-TICA

The deep-TICA CV has been trained on the converged OPES multithermal simulation of each system using the PyTorch library. The first 100 ns were discarded and a set of 180 (trodesquimine system) and 168 (squalamine system) molecular descriptors were evaluated during the remaining simulation time. A feed-forward Neural Network was employed, composed of an input layer with 180 and 168 nodes for trodesquimine and squalamine, respectively. Furthermore, two hidden layers containing 128 and 64 nodes were added. The hyperbolic tangent was used as activation function. The dataset was split into training/validation sets. To optimize the neural network parameters, we used the ADAM optimizer with a learning rate of $1e-3$. The lag time was set equal to 0.01 and in order to prevent overfitting the early stopping with a patience of 10 epochs was employed. Additionally, we scaled the inputs to have a zero mean and a variance of one. The deep-TICA CVs were also adjusted, ensuring their value range fell between -1 and 1. Finally, the trained model was exported as serialized model to be exploited during the OPES MetaD simulations.

Simulations set up

In this study, the trodesquimine and the squalamine molecules were considered with a lipid bilayer. More specifically, a membrane composition of 512 lipids was symmetrically modelled using the python tool insane¹⁵⁸, composed by 59% of 1,2-dioleoyl-sn-glycero-3-phosphocoline (DOPC), 30 % of sphingomyelin (SSM), 10 % of cholesterol (CHL), and 1% of monosialotetrahexosylganglioside 1 (GM1). The lipid bilayer was solvated using the polarizable water model of martini force-field and 150 mM NaCl was added to neutralize the net system charge. Finally, each complex consisted of the lipid bilayer, with the aminosterol molecule in the water environment, for a total of about 20'000 interacting particles. The MD simulations were performed with GROMACS 2021 software package patched with the development version of PLUMED 2.9¹⁷⁵ and the Pytorch library 1.4¹⁷⁶. Each molecular complex was first energetically minimized. To equilibrate each system, 1 ns in NVT ensemble at 310 K and 5 ns in NPT ensemble at 1 bar and 310 K simulations were sequentially performed

starting from each initial minimized structure. The temperatures and pressures of all systems were controlled using v-rescale thermostat⁸⁷ and Parrinello–Rahman method⁸⁸, respectively. The electrostatic interactions were calculated using the particle mesh Ewald method⁹⁰ with a real space cut-off of 11 Å. The cut-off value for van der Waals interactions was set at 11 Å.

Absolute ligand-membrane binding free energy

Compute the absolute ligand-membrane binding free-energy ΔG can be evaluated using the following relation:

$$\Delta G = -RT \ln (K_a) \quad (3.5.1)$$

Where R is the universal gas constant, T is the absolute temperature and K_a is the association constant, which is the inverse of equilibrium constant (K_d).

Consequently, a more favorable binding event, characterized by a lower K_d value, corresponds to a more negative ΔG , indicating a higher degree of spontaneity and stability in the complex formation. Thus, the interplay between ΔG and K_d offers valuable insights into the energetics and specificity of molecular interactions, ultimately shaping the landscape of biophysical and biochemical phenomena.

Results and Discussion

Design of collective variables for aminosterol absorption

Enhanced sampling methods like OPES²⁶, metadynamics²⁹, umbrella sampling²³ and others are based on the use of a set of collective variables (CVs). The CVs are intended to describe slow molecular phenomena as a function of atomic coordinates, reducing the complexity of the molecular system. Two are the primary approaches to constructing an effective CV. The first approach requires a deep understanding of the molecular system, enabling an expert to identify a few representative variables for the system. However, effectively sampling a complex system that describes the interplay between a number of molecular players may require up to hundreds of CVs²⁷. The second approach involves gathering a certain number of transitions between the two metastable states of interest, albeit crudely. Once this is done, specific molecular descriptors that can distinguish between the two states can be calculated.

In this context, OPES multithermal has been utilized to observe a number of transitions between aminosterol in the aqueous environment and aminosterol absorbed in the membrane. By

manipulating the system temperature, multithermal simulations inherently accelerate the system dynamics, reducing the energy barriers separating the metastable states. In Figure 25A, the time evolution of the normal distance between the center of mass of a specific aminosterol and the center of the lipid membrane is depicted as a function of the system potential energy. After approximately 200 ns, the shared bias among the OPES multithermal simulations converges, becoming nearly static (Figure 25B). Particularly when the absolute distance along Z is close to 2 nm, aminosterol is considered to be within the lipid bilayer. Notably, the OPES multithermal simulation enables to observe a certain number of transitions involving aminosterols absorption into the membrane (Figure 25A).

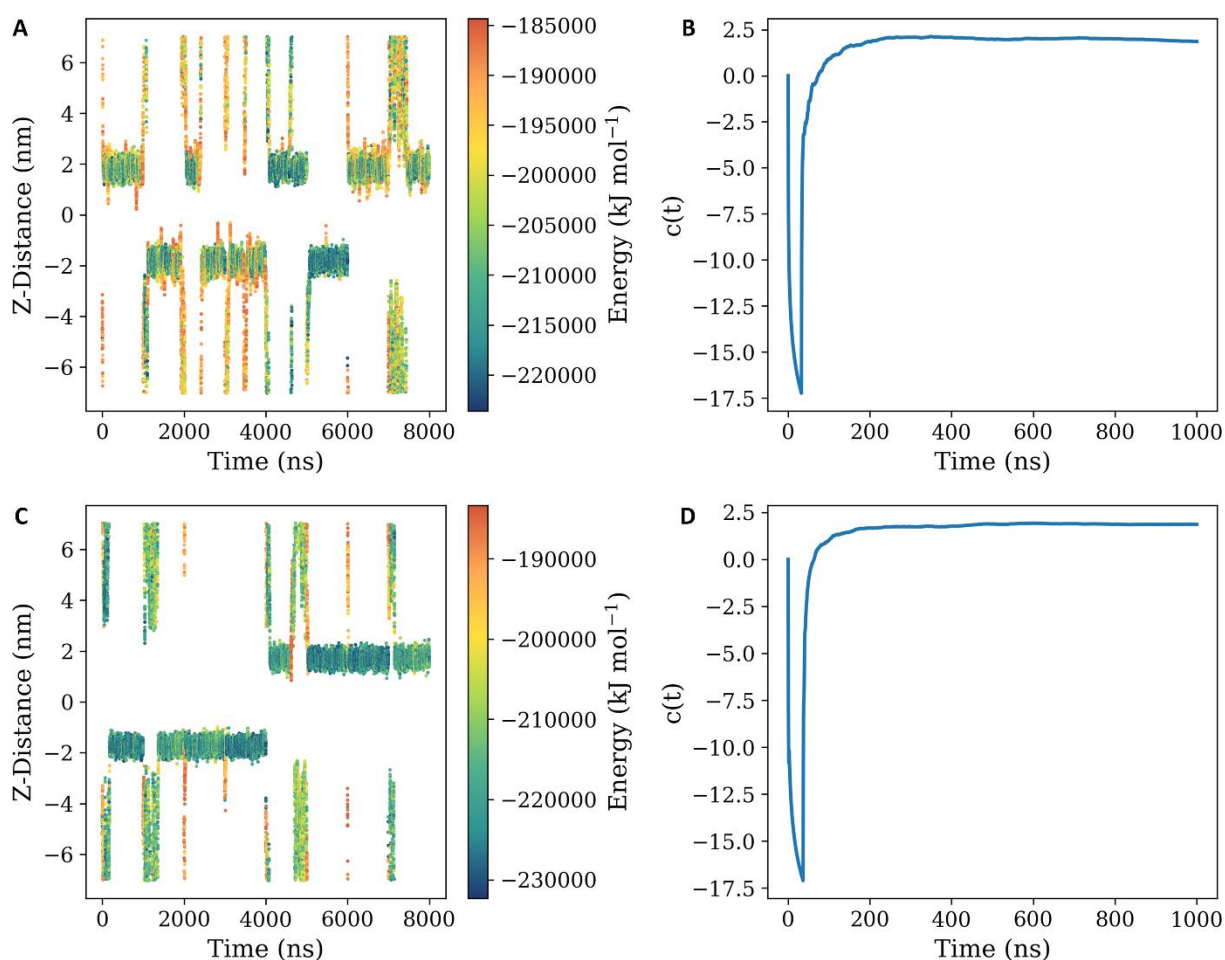


Figure 25: Time evolution of the normal distance between the center of mass of aminosterol (trodusquemine in panel A and squalamine in panel C) and the center of the cell membrane, colored by the potential energy of the molecular system. Time evolution of the OPES multithermal bias shared between 8 replicas that assess the convergence of the simulation of trodusquemine (panel B) or squalamine (panel D) system.

However, in complex molecular systems, this approach might not be sufficient to observe numerous transitions and accurately estimate the free energy profile. Therefore, it becomes essential to construct a CV that aptly describes the phenomenon of aminosterol absorption in the membrane. This not only avoids exploring irrelevant regions of the system but also increases the number of transitions between the metastable states, enabling a more accurate estimation of the free energy profile.

An approach to design an effective CV involves utilizing Time-lagged Independent Component Analysis (TICA), which focused on identifying the most slowly decorrelating modes, through the variational principle^{177,178}. The variational principle leads to a generalized eigenvalue equation when the modes are expressed as a linear combination of descriptors¹⁷⁹. In this context, Deep-TICA³⁵ is a combination of TICA with a hidden layer of a neural network (NN) enables its application beyond just linear combinations of descriptors, significantly improving the variational flexibility of the solution and enhancing its overall quality¹⁸⁰.

Utilizing data from the OPES multithermal enhanced simulations, we aimed to construct a CV capable of distinguishing between the aminosterol in a water environment and the aminosterol inserted into the lipid bilayer. To achieve this goal, each molecular system was evaluated using a series of descriptors. These descriptors were computed based on the minimum distance between each 'bead' in the aminosterol molecule under investigation and the group of 'beads' equal to each other in each phospholipid type, such as DOPC and SSM. DOPC and SSM were chosen because they are present in larger proportions within the lipid membrane and thus serve as suitable representatives for this particular system.

For the two aminosterols studied, a total of 180 descriptors for trodusquemine and 168 descriptors for squalamine have been evaluated. The descriptors were used as input for the Deep-TICA method, which combined the benefits of TICA with the flexibility and adaptability of a neural network (Figure 26).

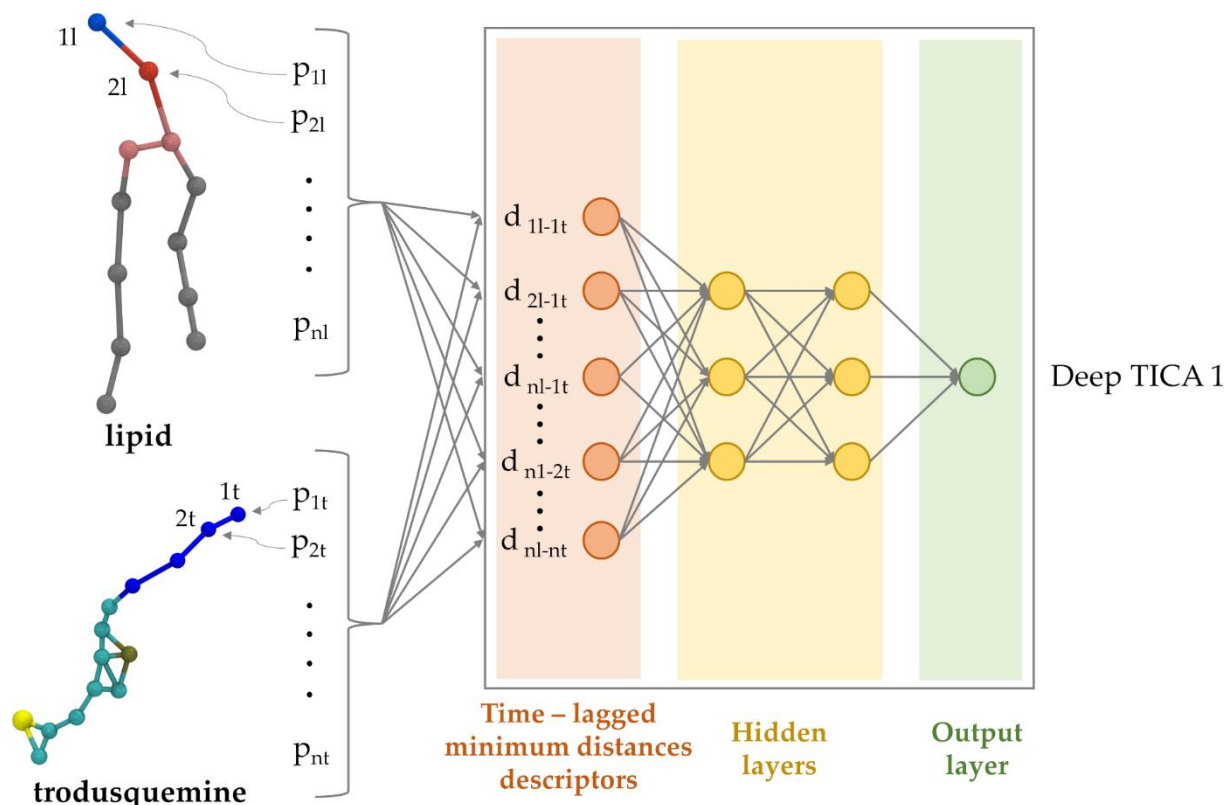


Figure 26: Constructing the Deep-TICA CV. The minimum distance between each 'bead' of the aminosterol and a group containing equal 'beads' from the lipids is calculated for each molecular system. These descriptors serve as the input layer of a neural network with two hidden layers, containing 128 and 64 neurons respectively. The output layer represents the Deep-TICA CV.

In each molecular system, the loss function is optimized with respect to four eigenvalues (see Appendix Figure 38 and Figure 39). In details, a CV was successfully developed, which precisely delineated the distinctions between the aminosterol situated in the aqueous environment and the aminosterol embedded within the lipid bilayer. Figure 27 presents a two-dimensional scatter plot illustrating the relationship between the Z component of the aminosterol dipole moment and the normal distance to the membrane, measured from the aminosterol center of mass to the membrane center. This relationship is depicted as a function of the TICA 1 CV value, which is assessed during the multithermal simulations. In detail, the CV is successful in describing the two characteristic metastable states of the molecule under investigation. Specifically, the two metastable states of interest are located at the extremes of CV TICA 1. When CV TICA 1 is equal to +1, the aminosterol molecule is situated in an aqueous environment. In contrast, when CV TICA 1 is equal to -1, absorption into the lipid bilayer occurs.

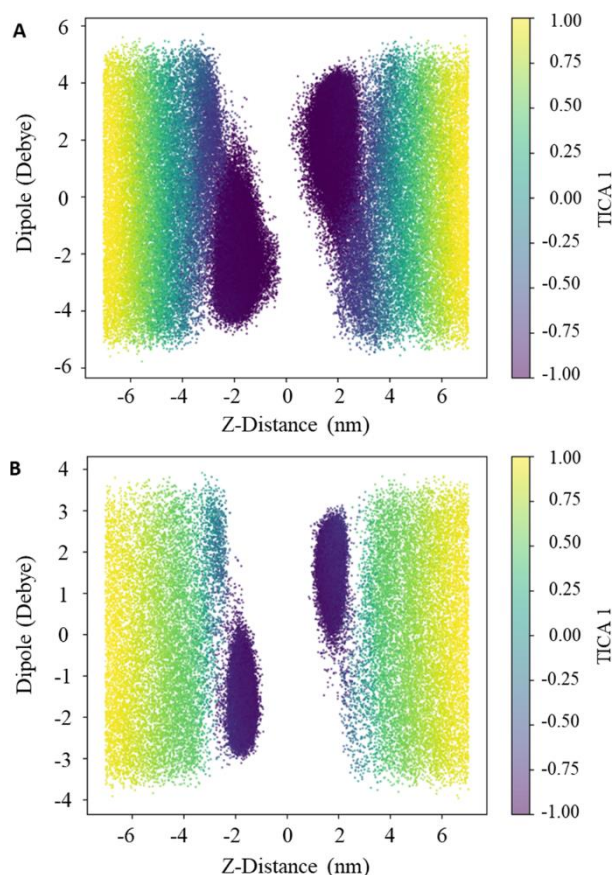


Figure 27: A scatter plot representation of the ability of the developed CV to differentiate the distinct metastable states of aminosterols when interacting with the cell membrane. The plot showcases the Z component of the aminosterol dipole moment and the perpendicular distance between the aminosterol center of mass and the membrane center, colored by the TICA 1 value.

Trodoquamine and squalamine thermodynamic properties prediction

Unbiased MD simulations struggle to build an accurate free-energy profile due to the limited accessible time-scales and the rarity of membrane penetration events. For a substance to enter cells through diffusion, it must overcome a significant free-energy barrier present in the cell membrane. The presence of this barrier acts as a roadblock, limiting the MD simulation's ability to thoroughly sample the large configuration space. This limitation highlights the need for enhanced sampling techniques and constructing suitable CVs to better understand the molecular behavior and overcome these challenges in simulating complex systems.

In this context, we employed OPES MetaD to enhance the trodoquamine and squalamine absorption. In detail, the developed TICA 1 variable was employed as CV. In Figure 28, the free energy surface (FES) is presented in a physically interpretable space characterized as a function of the normal distance from the aminosterol and the membrane center and the normal component to the dipole moment of the aminosterol. Two main states, separated by a free

energy barrier, can be observed. These states correspond to the basins found in water (state A) and lipid (state B) environments. In addition, the free energy difference between the A and B states is greater for the squalamine molecule, indicating a greater affinity for the neuron-like membrane.

This visualization helps to understand the transitions between the metastable states and the energy barriers associated with them, providing valuable insights into the behavior of the molecular system under investigation.

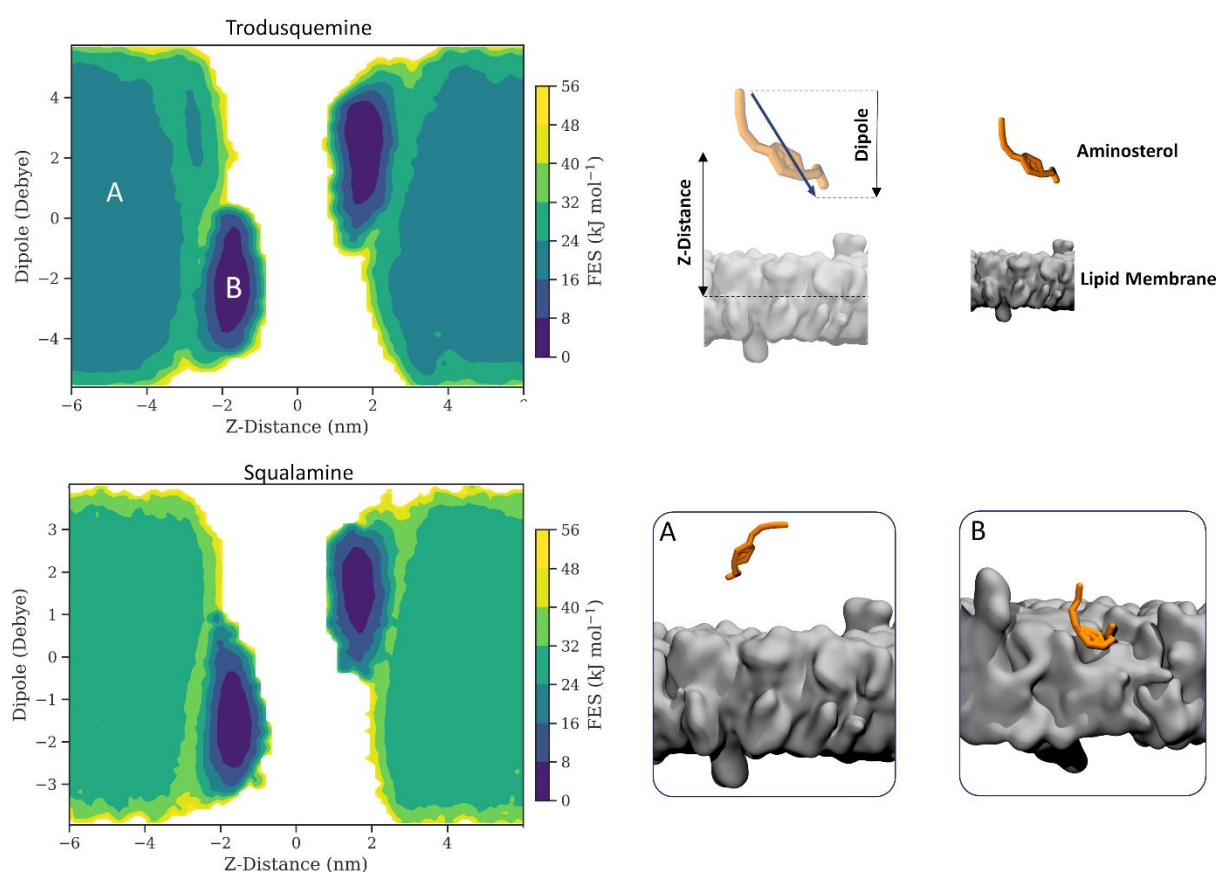


Figure 28: The FES of trodusquimine and squalamine are illustrated as a function of the z-distance from the membrane center and the z-dipole moment. Two FES minima, representing the metastable states, are depicted: A) aminosterol in the water environment, and B) aminosterol absorbed in the lipid bilayer.

In greater detail, the orientation of the molecule in the membrane plays a crucial role in describing the energy minimum. This can be achieved by considering the z-component of the dipole moment of aminosterols, which encompasses information about the molecule's orientation and charge.

However, the dipole moment is not an appropriate variable for understanding aminosterols affinity for the lipid membrane. Instead, the distance along the membrane component can be

considered as well representative variable for the free energy difference between the state A and B. In details Figure 29 shows the FES projection along the normal distance of the aminosterols and the membrane center. Trodusquemine exhibits a ΔG value of -16.77 ± 1.47 kJ mol⁻¹ with the energy minimum located at 1.72 nm. In contrast, squalamine presents a ΔG value -22.78 ± 1.05 kJ mol⁻¹ with the energy minimum positioned at 1.64 nm. Furthermore, an entry energy barrier of approximately 4.7 kJ mol⁻¹ at 2.64 nm for trodusquemine and 1.5 kJ mol⁻¹ at 2.58 nm for squalamine was observed. Squalamine therefore is able to go deeper into the plasma membrane. This ability is intuitively related to its molecular structure, which includes one less amine group than trodusquemine. In fact, a lower charge on the tail results in a lower preference to interact with the aqueous environment.

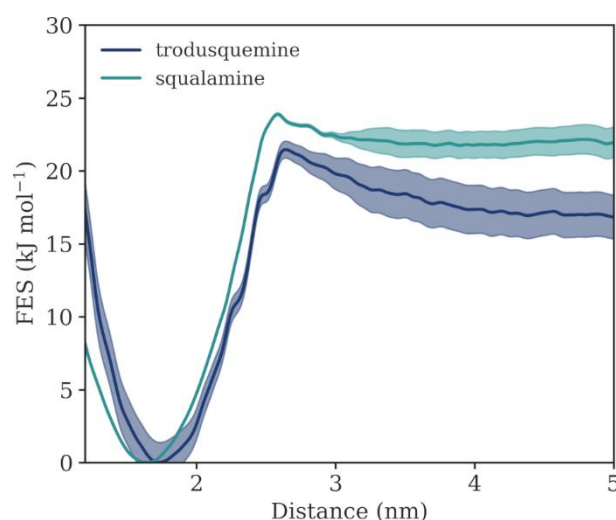


Figure 29: The FES of trodusquemine and squalamine along the normal distance from the membrane center.

However, membrane entry of aminosterols are multidimensional phenomena. The slow entrance movement of trodusquemine and squalamine is better represented by the projection in the 2D plane of the TICA 1 component and the distance along the z-axis, due to the intrinsic definition of the TICA variable (see Figure 30).

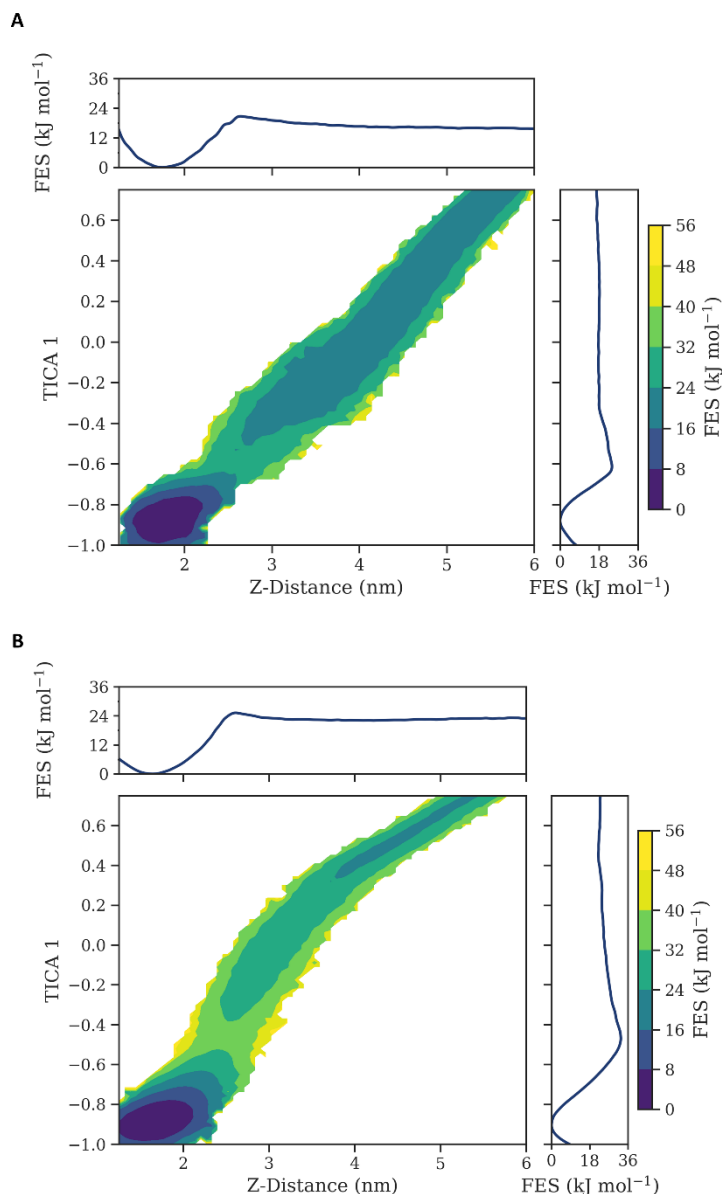


Figure 30: The FES of trodusquimine (panel A) and squalamine (panel B) is illustrated as a function of the z-distance from the membrane center and the TICA 1 CV. The FES projections along the corresponding axis are displayed in the upper and right panels.

In this case, the TICA 1 CV allows for a more accurate estimation of the FES (see Figure 31). In detail, the ΔG values are similar to the previous results with the z-distance as CV, obtaining 16.71 ± 1.1 kJ mol⁻¹ for trodusquimine and 22.75 ± 0.6 kJ mol⁻¹ for squalamine. This is consistent with the fact that the free-energy difference between the state A and the state B depends exclusively on the free-energy value at the two states, independently from the path that connects one state to the other. However, the entry barrier is approximately 7 kJ mol⁻¹ for trodusquimine and 10 kJ mol⁻¹ for squalamine. Consequently, TICA 1 CV facilitates enhanced

sampling among the metastable states of aminosterols and more effectively represents the energy barriers encountered along the membrane uptake pathway.

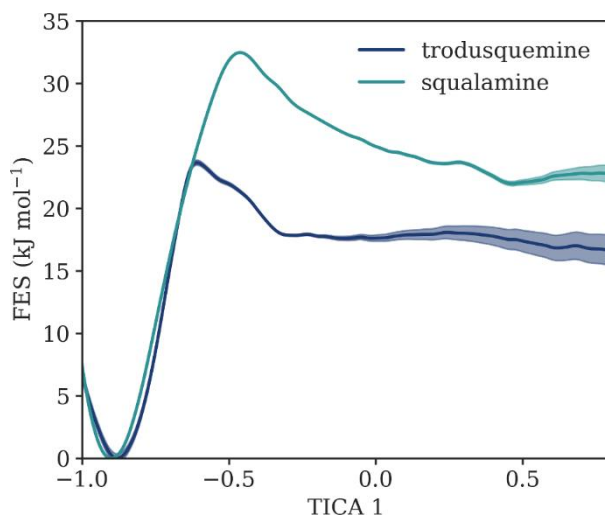


Figure 31: The FES of trodusquemine and squalamine along the TICA 1.

Furthermore, the advanced computational framework employed successfully delineates the membrane insertion process while also demonstrating exceptional proficiency in accurately estimating binding affinity within a reasonable simulation time. Binding thermodynamics properties as dissociation constant can be evaluated also by numerous experimental techniques. In this context, the binding affinity of trodusquemine and squalamine was also experimentally evaluated through Dynamic Light Scattering (DLS) and binding assay of fluorescently-labelled aminosterols techniques on large unilamellar vesicles (LUVs). The experimental data, acquired through collaboration with our experimental partners and currently in our possession, are unpublished at this time (see Appendix A4). The computational binding affinity values are in line with the experimental trend in which squalamine appears to have a higher binding affinity than trodusquemine (Table 5). Furthermore, the binding affinity values obtained by OPES MetaD are in agreement with the experimental ones if we take into consideration that the experiments were carried out at temperatures of 298 K and thus lower, than the computational 310 K. The excellent agreement with experimental data obtained achieved in this study is not commonly attained, considering the many efforts made in the past to determine the binding path and estimate binding energy values¹⁸¹. Considering, OPES MetaD is an MD-based protocol, the results are contingent upon the accuracy of the force-field parameters employed for both the membrane and the ligand, as well as the system setup, including factors such as the ligands' protonation states^{181,182}. Among computational techniques, metadynamics has proven to be highly effective in investigating both bound and unbound molecular phenomena. Nonetheless,

in the binding processes examined previously, only qualitative estimations of protein-ligand binding free energy have typically been achievable. Indeed, when the ligand exits the binding target, accurately sampling the free-energy profile within a reasonable computational timeframe often presents a challenge¹⁸¹.

Table 5: Trodusquemine and squalamine binding affinity with the neuron-like lipid bilayer. Comparison between MD enhanced sampling, experimental DLS and binding assay of fluorescently-labelled aminosterols data. Two fluorophores are considered: BODIPY TMR-X-labelled (BODIPY) or Alexa Fluor 594 (A594). All measures are expressed as kJ mol^{-1} .

	OPES MetaD	Exp. DLS	Exp. BODIPY	Exp. A594
ΔG_{TROD}	-16.77 ± 1.47	-25.42 ± 1.00	-25.17 ± 0.52	-21.16 ± 1.86
ΔG_{SQUA}	-22.78 ± 1.05	-28.07 ± 1.71	-26.61 ± 0.66	-23.4 ± 0.75

Conclusions

Two aminosterol compounds, trodusquemine and squalamine, derived from dogfish sharks, have emerged as promising candidates for combating NDs through the modulation of toxic misfolded protein aggregation kinetics. In details, squalamine and trodusquemine have demonstrated remarkable ability to inhibit lipid-induced nucleation of α -synuclein and $A\beta_{42}$, respectively. Furthermore, these molecules have shown promising results in preclinical studies and are currently undergoing clinical trials, with squalamine in Phase 2 for PD constipation and Phase 1 for PD dementia, while trodusquemine has exhibited potential in preventing hippocampal neuron loss and spatial memory deficits in an AD mouse model.

In this study, the combination of enhanced MD simulations, and advanced dimensionality reduction techniques, provides valuable insights into the interactions of aminosterols with neuron-like membranes. Previous studies highlighted the aminosterols influences on the kinetics of toxic oligomers^{72,77}, by displacing them from the cell membrane, increasing the lipid self-diffusion and the mechanical resistance force, and reducing the natural membrane negative charge⁷³. The results highlight the complex, multidimensional nature of aminosterol membrane insertion, with two main metastable states corresponding to lipid and aqueous environments. The free energy difference between these states suggests a greater affinity of squalamine for the neuron-like membrane. Furthermore, the use of TICA in conjunction with neural networks, through the innovative Deep-TICA approach, has allowed for the generation of an effective CV

to describe aminosterols absorption in the membrane. The study demonstrates the effectiveness of the advanced computational framework employed in accurately estimating binding affinity, which is vital for drug discovery. The free energy values obtained from OPES MetaD align with unpublished experimental data, lending further credibility to the findings. Overall, this research contributes to a deeper understanding of aminosterol-lipid membrane interactions, paving the way for the development of innovative treatments for NDs such as Alzheimer's and Parkinson's diseases. In conclusion, the computational framework employed in this research hold great promise not only for furthering our knowledge of aminosterol compounds but also for advancing the broader field of drug discovery in the fight against NDs and other complex disorders.

Conclusions and Future Perspectives

The present PhD thesis provides a comprehensive understanding of the molecular mechanism of action of trodusquemine, an aminosterol isolated from the dogfish shark *Squalus acanthias*, which has shown promise as a potential therapeutic agent against Alzheimer's disease. Using computational molecular modeling, this work sheds light on the self-aggregation behavior of trodusquemine and its interactions with the plasma membrane. The findings highlight the amphiphilic nature of trodusquemine, where the hydrophobic sterols represent the centre of the nucleation block, and the significant role of phosphate buffer in stabilizing its assembly. Furthermore, the trodusquemine ability to be incorporated in the hydrophilic portion of the cell membrane, modulates the physico-chemical properties of neuron-like membranes, increasing their resistance to the toxicity action of amyloid beta oligomers. The use of both coarse-grained and all-atoms models in this work highlights the importance of employing complementary multiscale modeling techniques to accurately capture the complex molecular interactions and dynamics involved in drug-membrane interactions. In detail, this study develops a computational pipeline that employs a bottom-up approach, which maps the coarse-grained model from all-atom data and automates the parameterization of bond terms, showcasing the effectiveness of multiscale modeling techniques. The coarse-grained investigation additionally uncovered that trodusquemine exhibits a predilection for associating with the separation boundary in a biphasic cell membrane. This interaction may modulate the coexistence of the liquid-gel and liquid-disorder phases, ultimately affecting the lipid raft's physicochemical properties and overall stability. Finally, this PhD thesis utilizes a combination of classical and enhanced sampling techniques to investigate the intricate, multidimensional process of aminosterol membrane insertion. The integration of dimensionality reduction techniques enhances by machine learning, improves the accuracy of the molecular dynamics simulations, allowing for the identification of crucial molecular interactions and the prediction of metastable states associated with the insertion process. These methods facilitate the identification of two primary metastable states corresponding to lipid and aqueous environments, which are separated by a free-energy barrier. Moreover, the findings demonstrate the effectiveness of the advanced computational framework employed in accurately estimating the aminosterols binding affinity for neuron-like membrane, which is crucial for drug discovery.

In summary, this PhD thesis has successfully elucidated the molecular mechanism of action of trodusquemine, paving the way for the development of novel therapeutic approaches for

Alzheimer's disease. The constructed computational platform can be further utilized in the development of new aminosterol-based therapies, with the potential to improve the treatment of Alzheimer's disease and other neurodegenerative disorders. Drug development for Alzheimer's disease has been an arduous endeavor, as evidenced by the lack of new FDA-approved medicines since 2003. However, this work represents a significant step forward in understanding the molecular basis of aminosterol action. Elucidating the molecular mechanism of action of drugs at the molecular level, may provide novel opportunities for overarching therapeutic approaches such as modulating misfolded proteins aggregation, or blocking common aggregation-induced cellular toxicity pathways. These findings could contribute to the enhancement of the drug development pipeline and ultimately lead to the discovery of effective treatments for Alzheimer's disease and other related disorders.

Appendix

A1 - Trodusquemine aggregation

Table A1: Summary of the all-atom MD simulation systems investigated to evaluate the trodusquemine aggregation mechanism in presence of counter ions or phosphate buffers.

# TROD	# Cl⁻	# PO₄¹⁻	# PO₄²⁻	Interacting Particles	Replicas	Total Simulation Time (ns)
2	6	-	-	15700	5	500
3	9	-	-	24800	5	500
4	12	-	-	37100	5	1000
6	18	-	-	49200	5	1000
8	24	-	-	64800	5	2000
2	-	2	2	15700	5	500
3	-	3	3	24800	5	500
4	-	4	4	37100	5	1000
6	-	6	6	49200	5	1000
8	-	8	8	64800	5	2000

A2 - Trodusquimine and squalamine bond terms optimization.

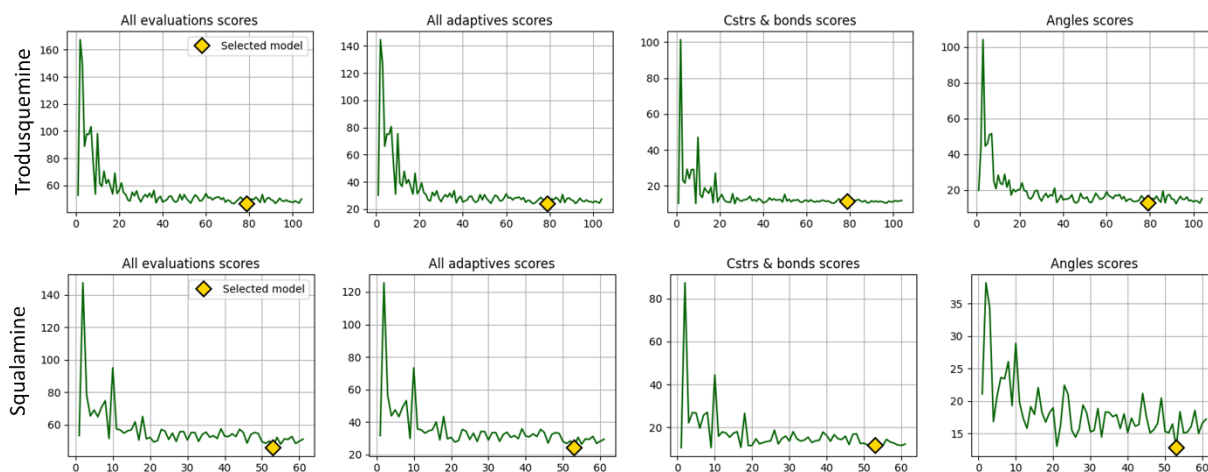


Figure 32: Total optimization scores over iterations.

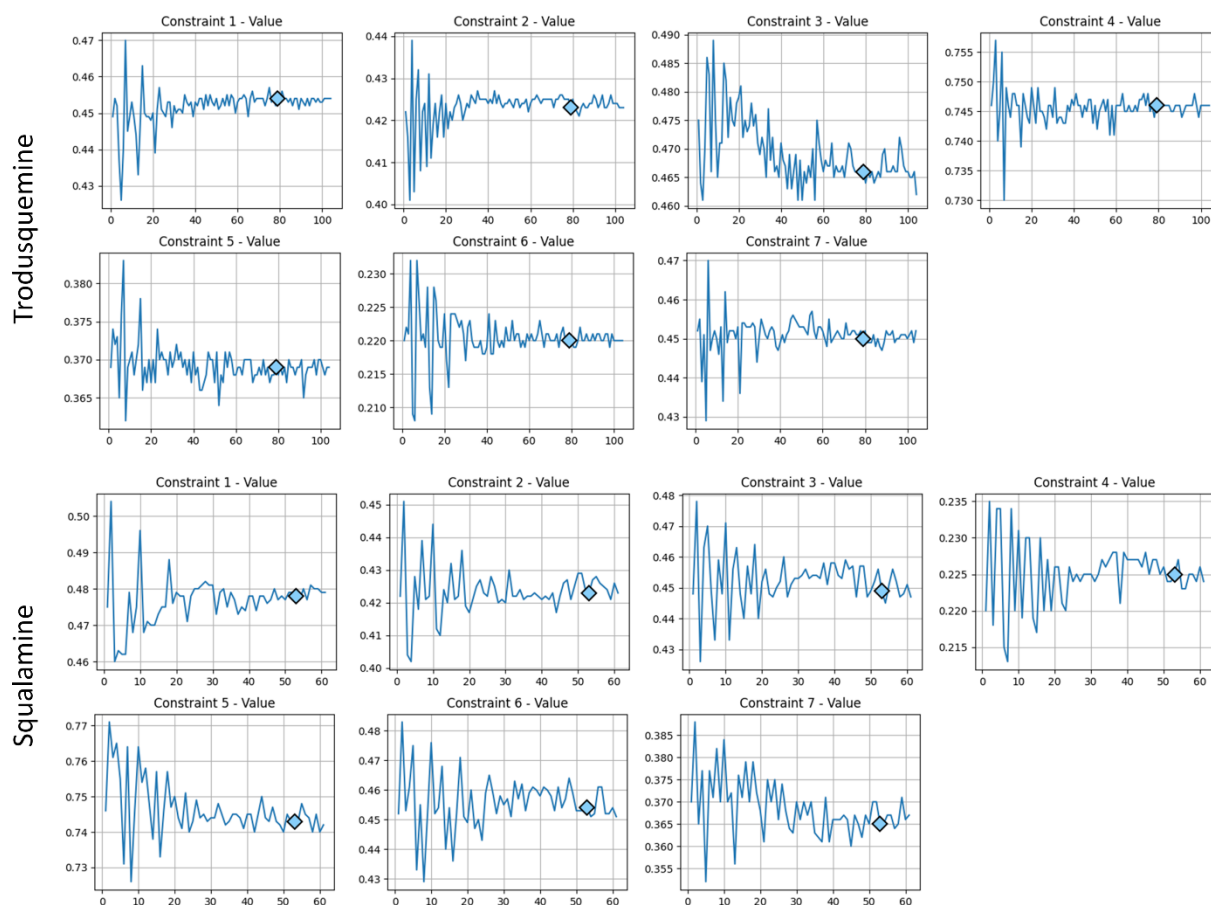


Figure 33: Constraints optimization over iterations.

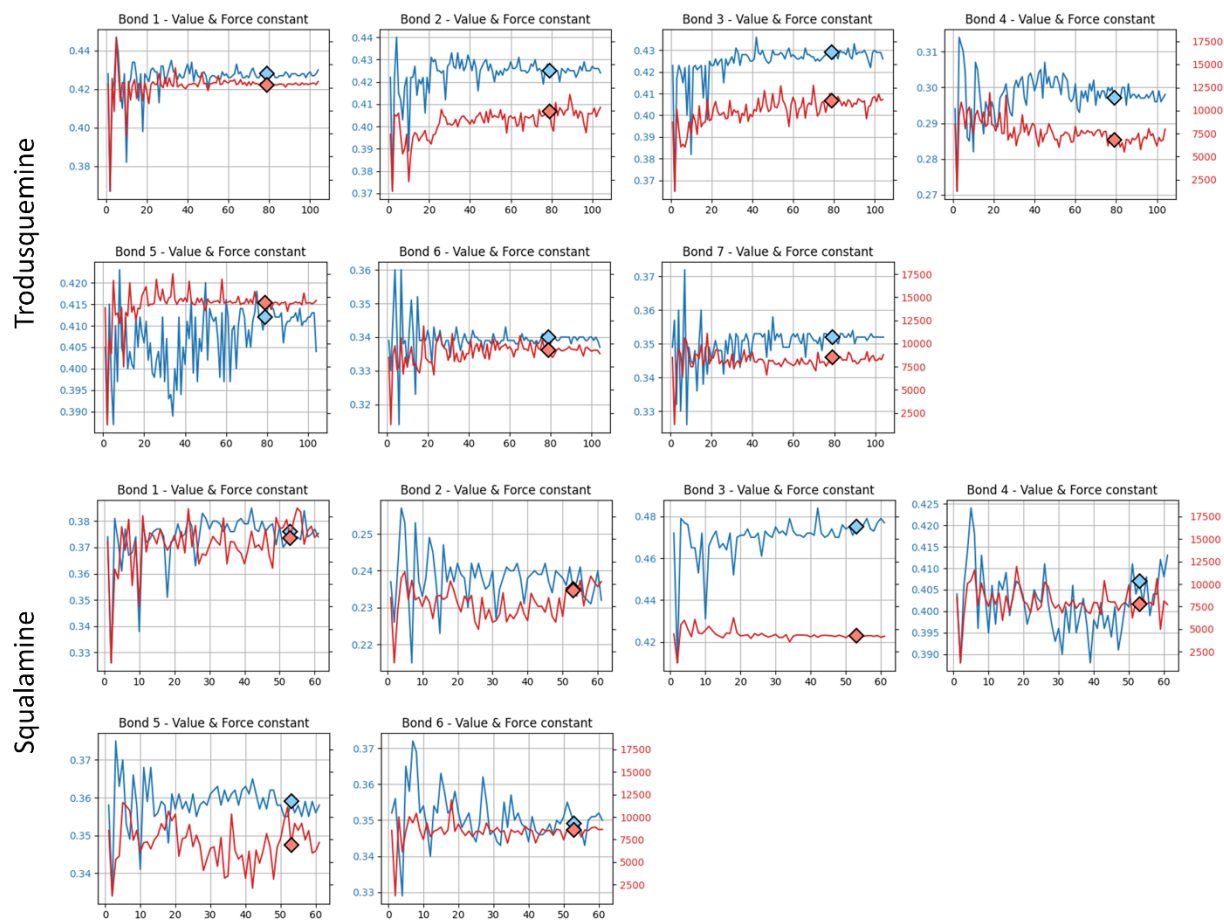


Figure 34: Bonds optimization over iterations.

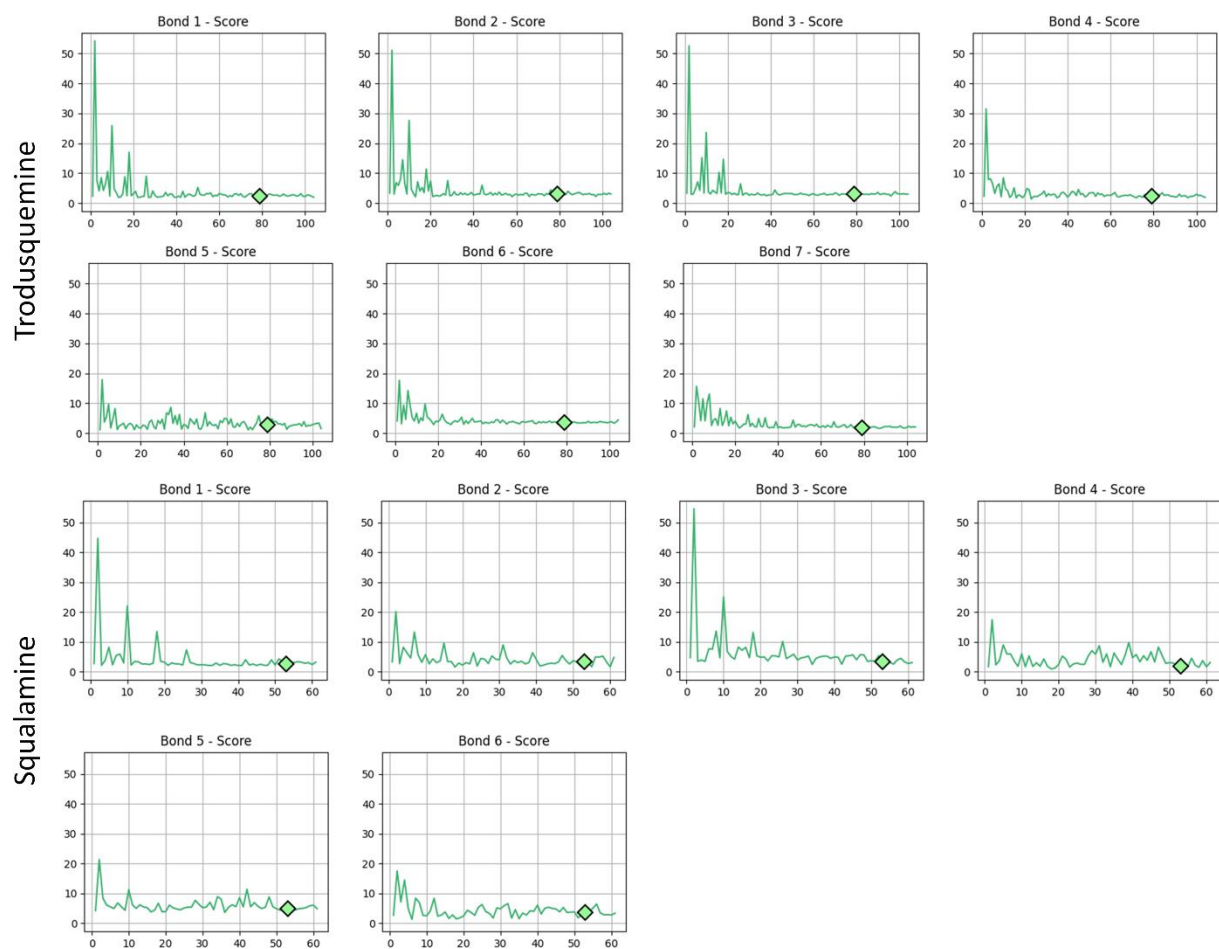


Figure 35: Summary of the bond scores over iterations.

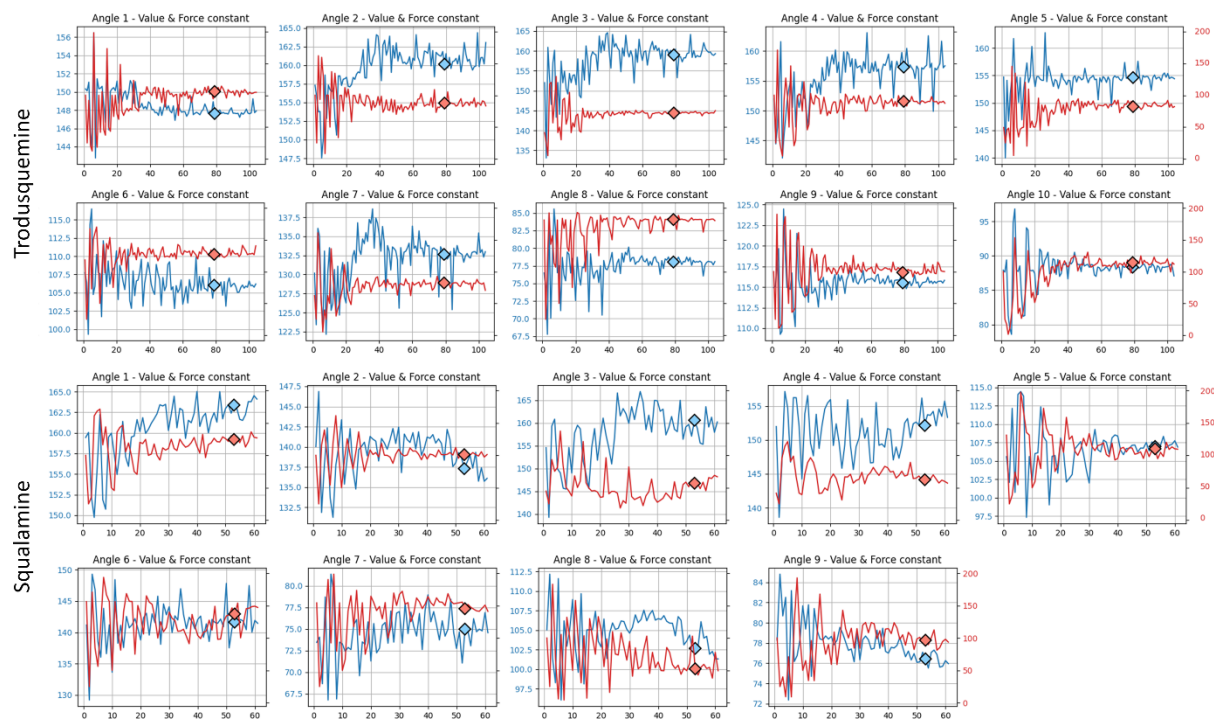


Figure 36:Angles optimization over iterations.

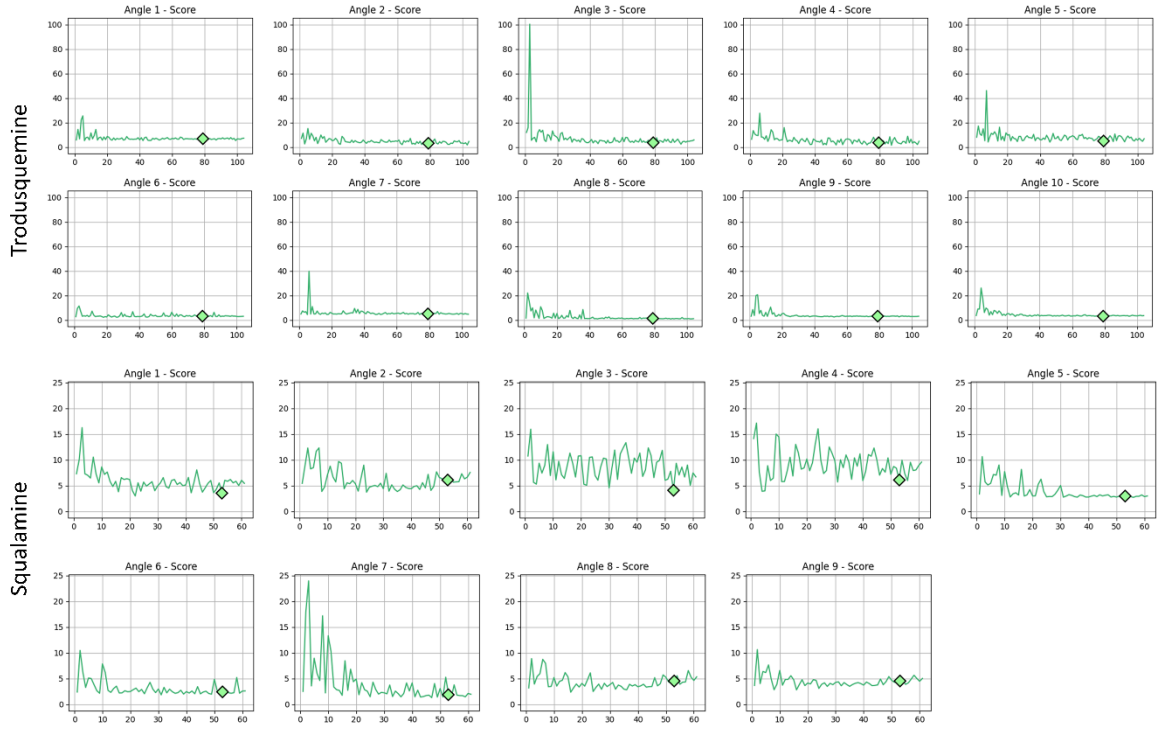


Figure 37: Summary of angles over iterations.

A3 – TICA training and validation

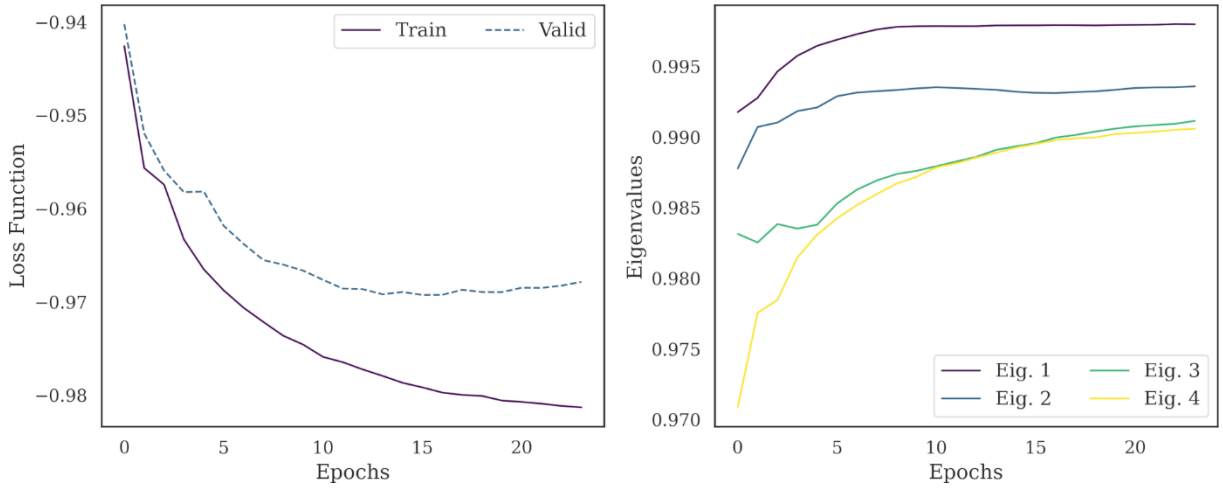


Figure 38: Loss function for TICA training (left panel) and eigenvalue of first TICA vectors (right panel) vs epochs of trodoquimine system.

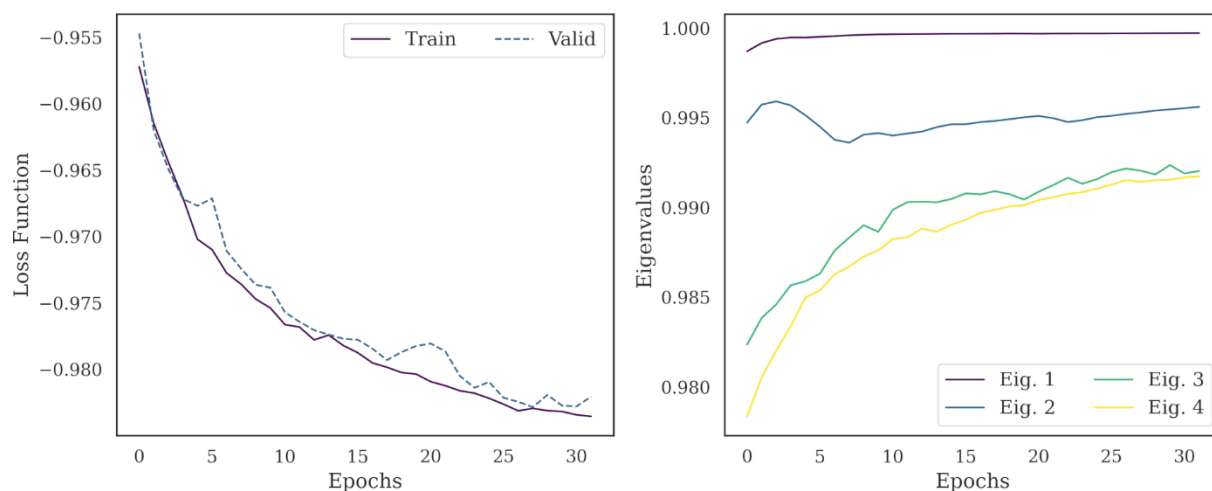


Figure 39: Loss function for TICA training (left panel) and eigenvalue of first TICA vectors (right panel) vs epochs of squalamine system.

A4 – Aminosterols experimental binding affinity

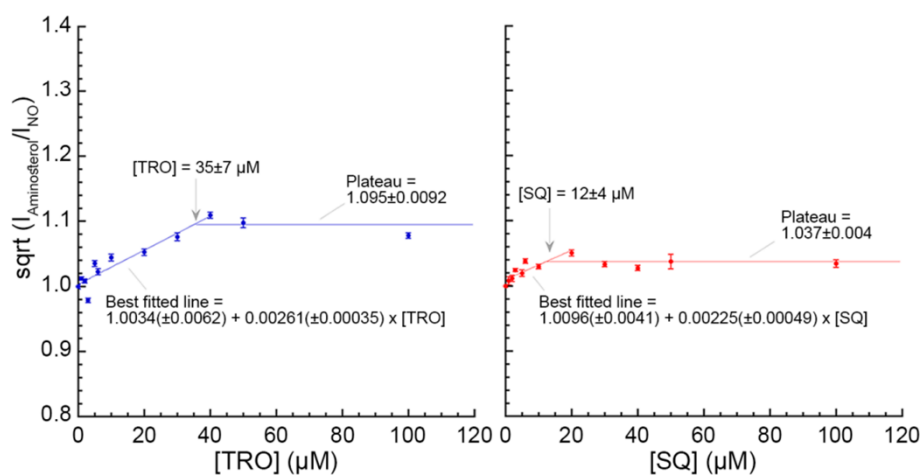


Figure 40: Dynamic Light Scattering of LUVs in the presence of a saturating concentration of trodusquemine (left panel) and squalamine (right panel).

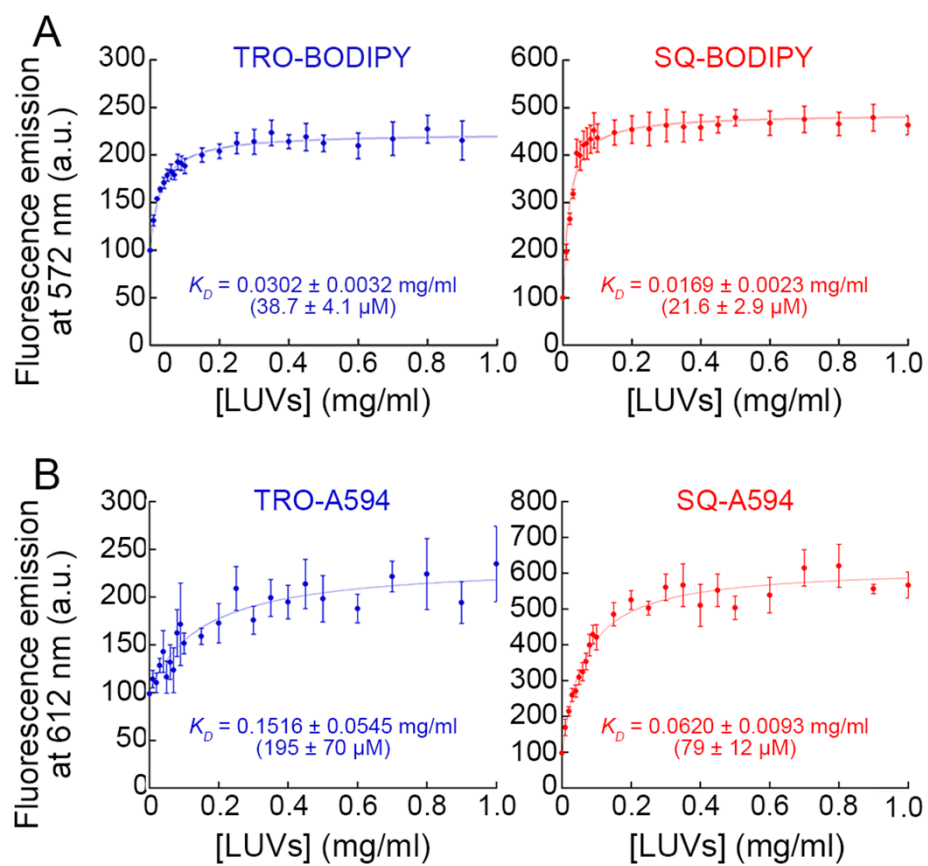


Figure 41: Binding assay of fluorescently-labelled aminosterols and LUVs; A) BODIPYTM TMR-X-labelled or B) Alexa Fluor[®] 594-labelled aminosterols.

Bibliography

1. Valverde, J. R. Molecular Modelling: Principles and Applications. *Brief Bioinform* 2, 199–200 (2001).
2. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J Chem Phys* 76, 637–649 (1982).
3. Sanbonmatsu, K. Y. & Tung, C.-S. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *J Struct Biol* 157, 470–480 (2007).
4. Riniker, S., Allison, J. R. & Gunsteren, W. F. van. On developing coarse-grained models for biomolecular simulation: a review. *Phys Chem Chem Phys* 14, 12423–12430 (2012).
5. Kmiecik, S. *et al.* Coarse-Grained Protein Models and Their Applications. *Chem Rev* 116, 7898–7936 (2016).
6. Monticelli, L. *et al.* The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* 4, 819–834 (2008).
7. López, C. A. *et al.* Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J Chem Theory Comput* 5, 3195–3210 (2009).
8. Rossi, G., Monticelli, L., Puisto, S. R., Vattulainen, I. & Ala-Nissila, T. Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter* 7, 698–708 (2010).
9. Uusitalo, J. J., Ingólfsson, H. I., Akhshi, P., Tieleman, D. P. & Marrink, S. J. Martini Coarse-Grained Force Field: Extension to DNA. *J Chem Theory Comput* 11, 3932–3945 (2015).
10. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & Vries, A. H. de. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J Phys Chem B* 111, 7812–7824 (2007).
11. Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chem Soc Rev* 42, 6801–6822 (2013).
12. Jarin, Z., Newhouse, J. & Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J Chem Theory Comput* 17, 1170–1180 (2021).
13. 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.
14. 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).

15. Michalowsky, J., Schäfer, L. V., Holm, C. & Smiatek, J. A refined polarizable water model for the coarse-grained MARTINI force field with long-range electrostatic interactions. *J Chem Phys* 146, 054501 (2017).
16. Michalowsky, J., Zeman, J., Holm, C. & Smiatek, J. A polarizable MARTINI model for monovalent ions in aqueous solution. *J Chem Phys* 149, 163319 (2018).
17. Jong, D. H. de *et al.* Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* 9, 687–697 (2013).
18. Graham, J. A., Essex, J. W. & Khalid, S. PyCGTOOL: Automated Generation of Coarse-Grained Molecular Dynamics Models from Atomistic Trajectories. *J Chem Inf Model* 57, 650–656 (2017).
19. Empereur-Mot, C. *et al.* Swarm-CG: Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. *Acs Omega* 5, 32823–32843 (2020).
20. Bereau, T. & Kremer, K. Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules. *J Chem Theory Comput* 11, 2783–2791 (2015).
21. Potter, T. D., Barrett, E. L. & Miller, M. A. Automated Coarse-Grained Mapping Algorithm for the Martini Force Field and Benchmarks for Membrane–Water Partitioning. *J Chem Theory Comput* 17, 5777–5791 (2021).
22. Camilloni, C. & Pietrucci, F. Advanced simulation techniques for the thermodynamic and kinetic characterization of biological systems. *Adv Phys X* 3, 1477531 (2018).
23. Marsili, S., Barducci, A., Chelli, R., Procacci, P. & Schettino, V. Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations. *J Phys Chem B* 110, 14011–14013 (2006).
24. Comer, J. *et al.* The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J Phys Chem B* 119, 1129–1151 (2015).
25. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc National Acad Sci* 99, 12562–12566 (2002).
26. Invernizzi, M. & Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *J Phys Chem Lett* 11, 2731–2736 (2020).
27. Chen, M., Yu, T.-Q. & Tuckerman, M. E. Locating landmarks on high-dimensional free energy surfaces. *Proc National Acad Sci* 112, 3235–3240 (2015).
28. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314, 141–151 (1999).
29. Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett* 100, 020603 (2007).
30. Park, S., Khalili-Araghi, F., Tajkhorshid, E. & Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *J Chem Phys* 119, 3559–3566 (2003).

31. Schlick, T. Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules. *F1000 Biology Reports* 1, 51 (2009).
32. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys Rev Lett* 78, 2690–2693 (1997).
33. Brandt, S., Sittel, F., Ernst, M. & Stock, G. Machine Learning of Biomolecular Reaction Coordinates. *J Phys Chem Lett* 9, 2144–2150 (2018).
34. Ribeiro, J. M. L., Bravo, P., Wang, Y. & Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J Chem Phys* 149, 072301 (2018).
35. Bonati, L., Piccini, G. & Parrinello, M. Deep learning the slow modes for rare events sampling. *Proc National Acad Sci* 118, e2113533118 (2021).
36. Dobson, C. M. Protein folding and misfolding. *Nature* 426, 884–890 (2003).
37. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat Neurosci* 21, 1332–1340 (2018).
38. Hardesty, B. & Kramer, G. Folding of a nascent peptide on the ribosome. *Prog Nucleic Acid Re* 66, 41–66 (2000).
39. Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struc Biol* 11, 114–119 (2001).
40. Dobson, C. M. The structural basis of protein folding and its links with human disease. *Philosophical Transactions Royal Soc Lond Ser B Biological Sci* 356, 133–145 (2001).
41. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annu Rev Biochem* 86, 1–42 (2016).
42. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci* 4, 49–60 (2003).
43. Goedert, M. Alzheimer’s and Parkinson’s diseases: The prion concept in relation to assembled A β , tau, and α -synuclein. *Science* 349, 1255555 (2015).
44. Scheltens, P. *et al.* Alzheimer’s disease. *Lancet* 397, 1577–1590 (2021).
45. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer’s disease at 25 years. *Embo Mol Med* 8, 595–608 (2016).
46. Roher, A. E. *et al.* Morphology and Toxicity of A β -(1-42) Dimer Derived from Neuritic and Vascular Amyloid Deposits of Alzheimer’s Disease*. *J Biol Chem* 271, 20631–20635 (1996).
47. Bloom, G. S. Amyloid- β and Tau: The Trigger and Bullet in Alzheimer Disease Pathogenesis. *Jama Neurol* 71, 505–508 (2014).
48. Nguyen, P. H., Sterpone, F. & Derreumaux, P. Aggregation of disease-related peptides. *Prog Mol Biol Transl* 170, 435–460 (2020).

49. Cummings, J. *et al.* Alzheimer's disease drug development pipeline: 2022. *Alzheimer's Dementia Transl Res Clin Interventions* 8, e12295 (2022).
50. Robert, H. *et al.* Donepezil and Memantine for Moderate-to-Severe Alzheimer's Disease. *New Engl J Med* 366, 893–903 (2012).
51. Hyde, C. *et al.* Evolution of the evidence on the effectiveness and cost-effectiveness of acetylcholinesterase inhibitors and memantine for Alzheimer's disease: systematic review and economic model. *Age Ageing* 42, 14–20 (2013).
52. Bieschke, J. Natural Compounds May Open New Routes to Treatment of Amyloid Diseases. *Neurotherapeutics* 10, 429–439 (2013).
53. Kroth, H. *et al.* Discovery and Structure Activity Relationship of Small Molecule Inhibitors of Toxic β -Amyloid-42 Fibril Formation*. *J Biol Chem* 287, 34786–34800 (2012).
54. Colletier, J.-P. *et al.* Molecular basis for amyloid- β polymorphism. *Proc National Acad Sci* 108, 16938–16943 (2011).
55. Cohen, S. I. A. *et al.* Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. *Proc National Acad Sci* 110, 9758–9763 (2013).
56. Habchi, J. *et al.* An anticancer drug suppresses the primary nucleation reaction that initiates the production of the toxic A β 42 aggregates linked with Alzheimer's disease. *Sci Adv* 2, e1501244 (2016).
57. Demuro, A. *et al.* Calcium Dysregulation and Membrane Disruption as a Ubiquitous Neurotoxic Mechanism of Soluble Amyloid Oligomers* \diamond . *J Biol Chem* 280, 17294–17300 (2005).
58. Arispe, N. & Doh, M. Plasma membrane cholesterol controls the cytotoxicity of Alzheimer's disease A β P (1–40) and (1–42) peptides. *Faseb J* 16, 1526–1536 (2002).
59. Abramov, A. Y., Ionov, M., Pavlov, E. & Duchon, M. R. Membrane cholesterol content plays a key role in the neurotoxicity of β -amyloid: implications for Alzheimer's disease. *Aging Cell* 10, 595–603 (2011).
60. Yu, X. & Zheng, J. Cholesterol Promotes the Interaction of Alzheimer β -Amyloid Monomer with Lipid Bilayer. *J Mol Biol* 421, 561–571 (2012).
61. Fantini, J. *et al.* Bexarotene Blocks Calcium-Permeable Ion Channels Formed by Neurotoxic Alzheimer's β -Amyloid Peptides. *Acs Chem Neurosci* 5, 216–224 (2014).
62. Graham, W. V., Bonito-Oliva, A. & Sakmar, T. P. Update on Alzheimer's Disease Therapy and Prevention Strategies. *Annu Rev Med* 68, 413–430 (2017).
63. Poewe, W. & Mahlknecht, P. Pharmacologic Treatment of Motor Symptoms Associated with Parkinson Disease. *Neurol Clin* 38, 255–267 (2020).
64. Ladiwala, A. R. A., Dordick, J. S. & Tessier, P. M. Aromatic Small Molecules Remodel Toxic Soluble Oligomers of Amyloid β through Three Independent Pathways*. *J Biol Chem* 286, 3209–3218 (2011).

65. Muscat, S. *et al.* The Impact of Natural Compounds on S-Shaped A β 42 Fibril: From Molecular Docking to Biophysical Characterization. *Int J Mol Sci* 21, 2017 (2020).
66. Limbocker, R. *et al.* Squalamine and trodusquemine: two natural products for neurodegenerative diseases, from physical chemistry to the clinic. *Nat Prod Rep* 39, 742–753 (2021).
67. Perni, M. *et al.* A natural product inhibits the initiation of α -synuclein aggregation and suppresses its toxicity. *Proc National Acad Sci* 114, E1009–E1017 (2017).
68. Galvagnion, C. *et al.* Lipid vesicles trigger α -synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol* 11, 229–234 (2015).
69. Lantz, K. A. *et al.* Inhibition of PTP1B by Trodusquemine (MSI-1436) Causes Fat-specific Weight Loss in Diet-induced Obese Mice. *Obesity* 18, 1516–1523 (2010).
70. Smith, A. M. *et al.* The protein tyrosine phosphatase 1B inhibitor MSI-1436 stimulates regeneration of heart and multiple other tissues. *Npj Regen Medicine* 2, 4 (2017).
71. Perni, M. *et al.* Multistep Inhibition of α -Synuclein Aggregation and Toxicity in Vitro and in Vivo by Trodusquemine. *Acs Chem Biol* 13, 2308–2319 (2018).
72. Limbocker, R. *et al.* Trodusquemine enhances A β 42 aggregation but suppresses its toxicity by displacing oligomers from cell membranes. *Nat Commun* 10, 225 (2019).
73. Errico, S. *et al.* Making biological membrane resistant to the toxicity of misfolded protein oligomers: a lesson from trodusquemine. *Nanoscale* 12, 22596–22614 (2020).
74. Ricke, K. M. *et al.* Neuronal Protein Tyrosine Phosphatase 1B Hastens Amyloid β -Associated Alzheimer's Disease in Mice. *J Neurosci* 40, 1581–1593 (2020).
75. Limbocker, R. *et al.* Trodusquemine displaces protein misfolded oligomers from cell membranes and abrogates their cytotoxicity through a generic mechanism. *Commun Biology* 3, 435 (2020).
76. Rao, M. N. *et al.* Aminosterols from the Dogfish Shark *Squalus acanthias*. *J Nat Prod* 63, 631–635 (2000).
77. Perni, M. *et al.* Multistep Inhibition of α -Synuclein Aggregation and Toxicity in Vitro and in Vivo by Trodusquemine. *Acs Chem Biol* 13, 2308–2319 (2018).
78. Lantz, K. A. *et al.* Inhibition of PTP1B by Trodusquemine (MSI-1436) Causes Fat-specific Weight Loss in Diet-induced Obese Mice. *Obesity* 18, 1516–1523 (2010).
79. Smith, A. M. *et al.* The protein tyrosine phosphatase 1B inhibitor MSI-1436 stimulates regeneration of heart and multiple other tissues. *Npj Regen Medicine* 2, 4 (2017).
80. Ricke, K. M. *et al.* Neuronal Protein Tyrosine Phosphatase 1B Hastens Amyloid β -Associated Alzheimer's Disease in Mice. *J Neurosci* 40, 1581–1593 (2020).
81. Perni, M. *et al.* A natural product inhibits the initiation of α -synuclein aggregation and suppresses its toxicity. *Proc National Acad Sci* 114, E1009–E1017 (2017).

82. Ingólfsson, H. I. *et al.* Lipid Organization of the Plasma Membrane. *J Am Chem Soc* 136, 14554–14559 (2014).
83. Seghezze, S., Diaspro, A., Canale, C. & Dante, S. Cholesterol Drives A β (1–42) Interaction with Lipid Rafts in Model Membranes. *Langmuir* 30, 13934–13941 (2014).
84. Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* 12, 405–413 (2016).
85. Kim, S. *et al.* CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J. Comput. Chem.* 38, 1879–1886 (2017).
86. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J Chem Phys* 81, 3684–3690 (1984).
87. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* 126, 014101 (2007).
88. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52, 7182–7190 (1981).
89. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472 (1997).
90. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J Chem Phys* 103, 8577–8593 (1995).
91. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminformatics* 4, 17 (2012).
92. Vanommeslaeghe, K. & MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J Chem Inf Model* 52, 3144–3154 (2012).
93. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 14, 71–73 (2017).
94. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software* 1, 19–25 (2015).
95. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graphics* 14, 33–38 (1996).
96. Khelashvili, G., Kollmitzer, B., Heftberger, P., Pabst, G. & Harries, D. Calculating the Bending Modulus for Multicomponent Lipid Membranes in Different Thermodynamic Phases. *J Chem Theory Comput* 9, 3866–3871 (2013).
97. Johner, N., Harries, D. & Khelashvili, G. Implementation of a methodology for determining elastic properties of lipid assemblies from molecular dynamics simulations. *Bmc Bioinformatics* 17, 161 (2016).
98. Grasso, G. *et al.* Cell penetrating peptide modulation of membrane biomechanics by Molecular dynamics. *J Biomech* 73, 137–144 (2018).

99. Khelashvili, G., Pabst, G. & Harries, D. Cholesterol Orientation and Tilt Modulus in DMPC Bilayers. *J Phys Chem B* 114, 7524–7534 (2010).
100. Watson, M. C., Penev, E. S., Welch, P. M. & Brown, F. L. H. Thermal fluctuations in shape, thickness, and molecular orientation in lipid bilayers. *J Chem Phys* 135, 244701 (2011).
101. Kozlovsky, Y. & Kozlov, M. M. Stalk Model of Membrane Fusion: Solution of Energy Crisis. *Biophys J* 82, 882–895 (2002).
102. Fošnarič, M., Iglič, A. & May, S. Influence of rigid inclusions on the bending elasticity of a lipid membrane. *Phys Rev E* 74, 051503 (2006).
103. Carrer, D. C., Schmidt, A. W., Knölker, H.-J. & Schwille, P. Membrane Domain-Disrupting Effects of 4-Substituted Cholesterol Derivatives. *Langmuir* 24, 8807–8812 (2008).
104. Chiantia, S., Ries, J., Kahya, N. & Schwille, P. Combined AFM and Two-Focus SFCS Study of Raft-Exhibiting Model Membranes. *Chemphyschem* 7, 2409–2418 (2006).
105. Khelashvili, G., Kollmitzer, B., Heftberger, P., Pabst, G. & Harries, D. Calculating the Bending Modulus for Multicomponent Lipid Membranes in Different Thermodynamic Phases. *J Chem Theory Comput* 9, 3866–3871 (2013).
106. Sanbonmatsu, K. Y. & Tung, C.-S. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *J Struct Biol* 157, 470–480 (2007).
107. Gupta, C., Sarkar, D., Tieleman, D. P. & Singharoy, A. The ugly, bad, and good stories of large-scale biomolecular simulations. *Curr Opin Struc Biol* 73, 102338 (2022).
108. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314, 141–151 (1999).
109. Souza, P. C. T. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods* 18, 382–388 (2021).
110. Ingólfsson, H. I., Arnarez, C., Periole, X. & Marrink, S. J. Computational ‘microscopy’ of cellular membranes. *J Cell Sci* 129, 257–268 (2016).
111. Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G. & Schulten, K. Discovery Through the Computational Microscope. *Structure* 17, 1295–1306 (2009).
112. Ingólfsson, H. I. *et al.* The power of coarse graining in biomolecular simulations. *Wiley Interdiscip Rev Comput Mol Sci* 4, 225–248 (2014).
113. Abellón-Ruiz, J. *et al.* Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nat Microbiol* 2, 1616–1623 (2017).
114. Yen, H.-Y. *et al.* PtdIns(4,5)P₂ stabilizes active states of GPCRs and enhances selectivity of G-protein coupling. *Nature* 559, 423–427 (2018).
115. Eerden, F. J. V., Melo, M. N., Frederix, P. W. J. M., Periole, X. & Marrink, S. J. Exchange pathways of plastoquinone and plastoquinol in the photosystem II complex. *Nat Commun* 8, 15214 (2017).

116. Vögele, M., Köfinger, J. & Hummer, G. Hydrodynamics of Diffusion in Lipid Membrane Simulations. *Phys Rev Lett* 120, 268104 (2018).
117. D’Agostino, M., Risselada, H. J., Lürick, A., Ungermann, C. & Mayer, A. A tethering complex drives the terminal stage of SNARE-dependent membrane fusion. *Nature* 551, 634–638 (2017).
118. Jeena, M. T. *et al.* Mitochondria localization induced self-assembly of peptide amphiphiles for cellular dysfunction. *Nat Commun* 8, 26 (2017).
119. Jiang, Z. *et al.* Subnanometre ligand-shell asymmetry leads to Janus-like nanoparticle membranes. *Nat Mater* 14, 912–917 (2015).
120. Maingi, V. *et al.* Stability and dynamics of membrane-spanning DNA nanopores. *Nat Commun* 8, 14784 (2017).
121. Frederix, P. W. J. M. *et al.* Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nat Chem* 7, 30–37 (2015).
122. Bochicchio, D., Salvalaglio, M. & Pavan, G. M. Into the Dynamics of a Supramolecular Polymer at Submolecular Resolution. *Nat Commun* 8, 147 (2017).
123. Joshi, S. Y. & Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol Simulat* 47, 786–803 (2021).
124. Huber, F. *et al.* Emergent complexity of the cytoskeleton: from single filaments to tissue. *Adv Phys* 62, 1–112 (2013).
125. Pak, A. J. & Voth, G. A. Advances in coarse-grained modeling of macromolecular complexes. *Curr Opin Struc Biol* 52, 119–126 (2018).
126. Kovermann, M., Rogné, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q Rev Biophys* 49, e6 (2016).
127. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & Vries, A. H. de. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J Phys Chem B* 111, 7812–7824 (2007).
128. Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chem Soc Rev* 42, 6801–6822 (2013).
129. Reith, D., Pütz, M. & Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* 24, 1624–1636 (2003).
130. Moore, T. C., Iacovella, C. R. & McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J Chem Phys* 140, 224104 (2014).
131. Tschöp, W., Kremer, K., Batoulis, J., Bürger, T. & Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* 49, 61–74 (1998).
132. Milano, G., Goudeau, S. & Müller-Plathe, F. Multicentered Gaussian-based potentials for coarse-grained polymer simulations: Linking atomistic and mesoscopic scales. *J. Polym. Sci. B Polym. Phys.* 43, 871–885 (2005).

133. Unni, S. *et al.* Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J. Comput. Chem.* 32, 1488–1491 (2011).
134. Gambacorta, A., Gliozzi, A. & Rosa, M. D. Archaeal lipids and their biotechnological applications. *World J Microbiol Biotechnology* 11, 115–131 (1995).
135. Dhasaiyan, P. & Prasad, B. L. V. Self-Assembly of Bolaamphiphilic Molecules. *Chem. Rec.* 17, 597–610 (2017).
136. Huang, Z. *et al.* Structure–activity relationship studies of symmetrical cationic bolasomes as non-viral gene vectors. *J Mater Chem B* 4, 5575–5584 (2016).
137. Jeftić, J. & Benvegnu, T. Self-assembling Biomaterials. 113–156 (2018) doi:10.1016/b978-0-08-102015-9.00007-1.
138. Richter, S. *et al.* Effects of Common Buffer Systems on Drug Activity: The Case of Clerocidin. *Chem Res Toxicol* 17, 492–501 (2004).
139. Ramakrishnan, M. *et al.* Probing Cocaine-Antibody Interactions in Buffer and Human Serum. *Plos One* 7, e40518 (2012).
140. Velinova, M., Sengupta, D., Tadjer, A. V. & Marrink, S.-J. Sphere-to-Rod Transitions of Nonionic Surfactant Micelles in Aqueous Solution Modeled by Molecular Dynamics Simulations. *Langmuir* 27, 14071–14077 (2011).
141. Mirzaei, S. *et al.* Enhancement of The Stability of Human Growth Hormone by Using Tris(hydroxymethyl)aminomethane: Molecular Docking and Experimental Analysis. *Cell J Yakhteh* 22, 406–414.
142. Kim, S. J., Hahn, S. K., Kim, M. J., Kim, D. H. & Lee, Y. P. Development of a novel sustained release formulation of recombinant human growth hormone using sodium hyaluronate microparticles. *J Control Release* 104, 323–335 (2005).
143. Taha, M. & Lee, M.-J. Interactions of TRIS [tris(hydroxymethyl)aminomethane] and related buffers with peptide backbone: Thermodynamic characterization. *Phys Chem Chem Phys* 12, 12840–12850 (2010).
144. Marrink, S. J. *et al.* Computational Modeling of Realistic Cell Membranes. *Chem Rev* 119, 6184–6226 (2019).
145. Yang, Y., Lee, M. & Fairn, G. D. Phospholipid subcellular localization and dynamics. *J Biol Chem* 293, 6230–6240 (2018).
146. Niemelä, P. S., Ollila, S., Hyvönen, M. T., Karttunen, M. & Vattulainen, I. Assessing the Nature of Lipid Raft Membranes. *Plos Comput Biol* 3, e34 (2007).
147. Levental, I., Levental, K. R. & Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol* 30, 341–353 (2020).
148. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* 357, (2017).

149. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat Commun* 8, 275 (2017).
150. Grassi, S. *et al.* Lipid rafts and neurodegeneration: structural and functional roles in physiologic aging and neurodegenerative diseases Thematic Review Series: Biology of Lipid Rafts. *J Lipid Res* 61, 636–654 (2020).
151. Cheng, H. *et al.* Mechanisms of Disease: new therapeutic strategies for Alzheimer’s disease—targeting APP processing in lipid rafts. *Nat Clin Pract Neuro* 3, 374–382 (2007).
152. Lemkul, J. A. & Bevan, D. R. Lipid composition influences the release of Alzheimer’s amyloid β -peptide from membranes. *Protein Sci* 20, 1530–1545 (2011).
153. Hartmann, T., Kuchenbecker, J. & Grimm, M. O. W. Alzheimer’s disease: the lipid connection. *J Neurochem* 103, 159–170 (2007).
154. Sonnino, S. *et al.* Lipid Rafts in Neurodegeneration and Neuroprotection. *Mol Neurobiol* 50, 130–148 (2014).
155. Nyholm, T. K. M., Lindroos, D., Westerlund, B. & Slotte, J. P. Construction of a DOPC/PSM/Cholesterol Phase Diagram Based on the Fluorescence Properties of trans-Parinaric Acid. *Langmuir* 27, 8339–8350 (2011).
156. Thallmair, S., Javanainen, M., Fábíán, B., Martinez-Seara, H. & Marrink, S. J. Nonconverged Constraints Cause Artificial Temperature Gradients in Lipid Bilayer Simulations. *J Phys Chem B* 125, 9537–9546 (2021).
157. He, S. & Maibaum, L. Identifying the Onset of Phase Separation in Quaternary Lipid Bilayer Systems from Coarse-Grained Simulations. *J Phys Chem B* 122, 3961–3973 (2018).
158. 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).
159. Fragneto, G., Delhom, R., Joly, L. & Scoppola, E. Neutrons and model membranes: Moving towards complexity. *Curr Opin Colloid In* 38, 108–121 (2018).
160. Baoukina, S. & Tieleman, D. P. Simulations of lipid monolayers. *Methods Mol Biology* 924, 431–444 (2012).
161. Heberle, F. A. & Feigenson, G. W. Phase Separation in Lipid Membranes. *Csh Perspect Biol* 3, a004630 (2011).
162. Risselada, H. J. & Marrink, S. J. The molecular face of lipid rafts in model membranes. *Proc National Acad Sci* 105, 17367–17372 (2008).
163. Simakova, O. & Arispe, N. J. The Cell-Selective Neurotoxicity of the Alzheimer’s A β Peptide Is Determined by Surface Phosphatidylserine and Cytosolic ATP Levels. Membrane Binding Is Required for A β Toxicity. *J Neurosci* 27, 13719–13729 (2007).

164. Kawarabayashi, T. *et al.* Dimeric Amyloid β Protein Rapidly Accumulates in Lipid Rafts followed by Apolipoprotein E and Phosphorylated Tau Accumulation in the Tg2576 Mouse Model of Alzheimer's Disease. *J Neurosci* 24, 3801–3809 (2004).
165. Gyls, K. H., Fein, J. A., Yang, F., Miller, C. A. & Cole, G. M. Increased cholesterol in A β -positive nerve terminals from Alzheimer's disease cortex. *Neurobiol Aging* 28, 8–17 (2007).
166. Martín, V. *et al.* Lipid Alterations in Lipid Rafts from Alzheimer's Disease Human Brain Cortex. *J Alzheimer's Dis* 19, 489–502 (2010).
167. Rushworth, J. V. & Hooper, N. M. Lipid Rafts: Linking Alzheimer's Amyloid- β Production, Aggregation, and Toxicity at Neuronal Membranes. *Int J Alzheimer's Dis* 2011, 603052 (2011).
168. Barletti, B. *et al.* Reorganization of the outer layer of a model of the plasma membrane induced by a neuroprotective aminosterol. *Colloids Surfaces B Biointerfaces* 222, 113115 (2022).
169. Feigin, V. L. *et al.* The global burden of neurological disorders: translating evidence into policy. *Lancet Neurology* 19, 255–265 (2019).
170. Kins, S., Schäfer, K.-H. & Endres, K. Drug development for neurodegenerative diseases. *Biol Chem* 403, 1–1 (2022).
171. Cardenas, A. E., Shrestha, R., Webb, L. J. & Elber, R. Membrane Permeation of a Peptide: It Is Better to be Positive. *J Phys Chem B* 119, 6412–6420 (2015).
172. Swift, R. V. & Amaro, R. E. Back to the Future: Can Physical Models of Passive Membrane Permeability Help Reduce Drug Candidate Attrition and Move Us Beyond QSPR? *Chem Biol Drug Des* 81, 61–71 (2013).
173. Trofimenko, E. *et al.* Genetic, cellular, and structural characterization of the membrane potential-dependent cell-penetrating peptide translocation pore. *Elife* 10, e69832 (2021).
174. Invernizzi, M., Piaggi, P. M. & Parrinello, M. Unified Approach to Enhanced Sampling. *Phys Rev X* 10, 041034 (2020).
175. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 185, 604–613 (2014).
176. Bonati, L., Rizzi, V. & Parrinello, M. Data-Driven Collective Variables for Enhanced Sampling. *J Phys Chem Lett* 11, 2998–3004 (2020).
177. Pérez-Hernández, G., Paul, F., Giorgino, T., Fabritiis, G. D. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* 139, 015102 (2013).
178. Naritomi, Y. & Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J Chem Phys* 134, 065101 (2011).
179. Wu, H. *et al.* Variational Koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *J Chem Phys* 146, 154104 (2017).

180. Ansari, N., Rizzi, V. & Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. *Nat Commun* 13, 5438 (2022).
181. Raniolo, S. & Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat Protoc* 15, 2837–2866 (2020).
182. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690 (2010).