

Lecture 7

AI in biological physics potential and molecular simulation aided with NN

Session 7.1

AI POTENTIAL IN FUNDAMENTAL BIOLOGICAL PHYSICS AND DRUG DISCOVERY

AI potential in discovering fundamental physics and improve our understanding of structure-based drug design and quantum chemistry.

Session 7.2

MOLECULAR DYNAMICS SIMULATIONS AIDED WITH AI

Machine Learning meets simulation of macro-molecular machines, Setting up an MD simulation of a ligand receptor complex, with ML parameters.

UQ 7

Exercise 7

AI FOR PROTEIN STRUCTURE PREDICTION - INVESTIGATING THE CONFIDENCE FOR CYCLOOXYGENASE AND ACTIVE SITE

Methods: AlphaFold. Neural Networks. 3D Viewing



Graded material



Extra
Understanding

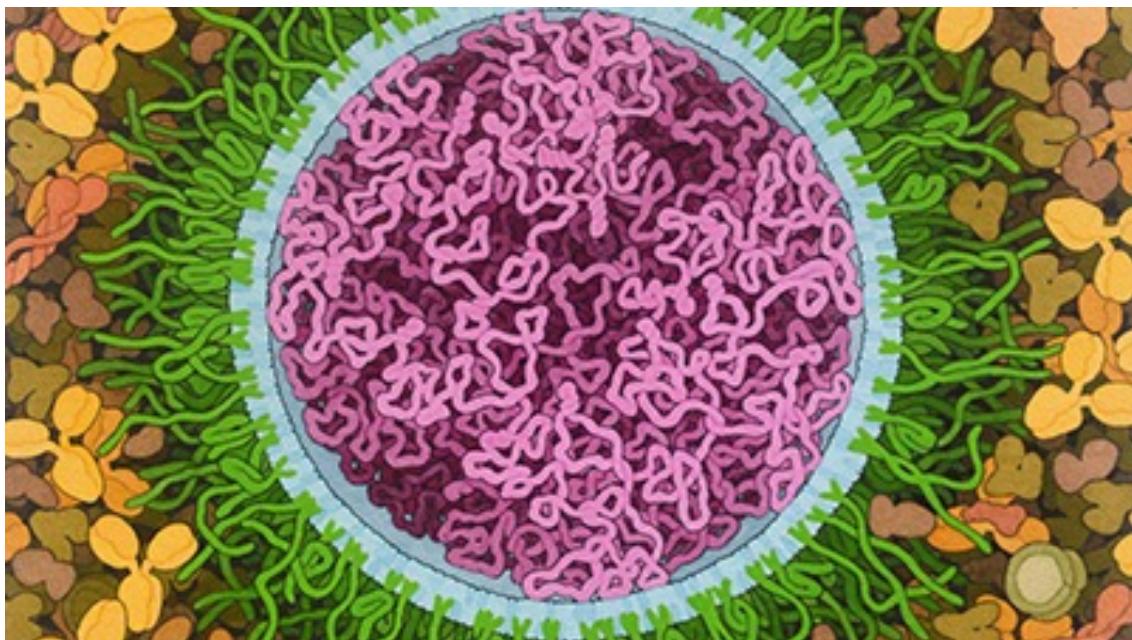


Figure: Schematic showing structure of covid vaccine; where the mRNA strands are embedded in lipids (blue), where lipids are connected to polyethylene glycol chains (green), that help shield the vaccine from the immune system. The surroundings are an illustration of the crowded environment in cells. Painting by Goodsell (2020), reused here ([CC-BY-4.0](#) license).

During the COVID-19 pandemic mRNA vaccines suddenly was on everyone's lips. In these vaccines the developed drug is basically mRNA instructions that hijack the cells machinery to produce a spike protein that can be recognized by our immune system, to train it to recognize the spike protein on corona viruses. In this case it is the understanding of the translation of mRNA to proteins and other information that is utilized in order to create a vaccine for Covid-19.

To understand the process of how

genetic information, in the form of virus copied mRNA, can be translated into a 3D protein spike-protein, by human cells, this understanding is quite an achievement!

Though biochemistry, safety and efficacy would certainly benefit from being better clarified and visualized in general to the public. See also (Goodsell and Burley, 2021).

mRNA: messenger ribonucleic acid (mRNA) is a single-stranded molecule of RNA that corresponds to the genetic sequence of a gene, that is read by a ribosome and converted to a protein. It is made from DNA

during the transcription process.

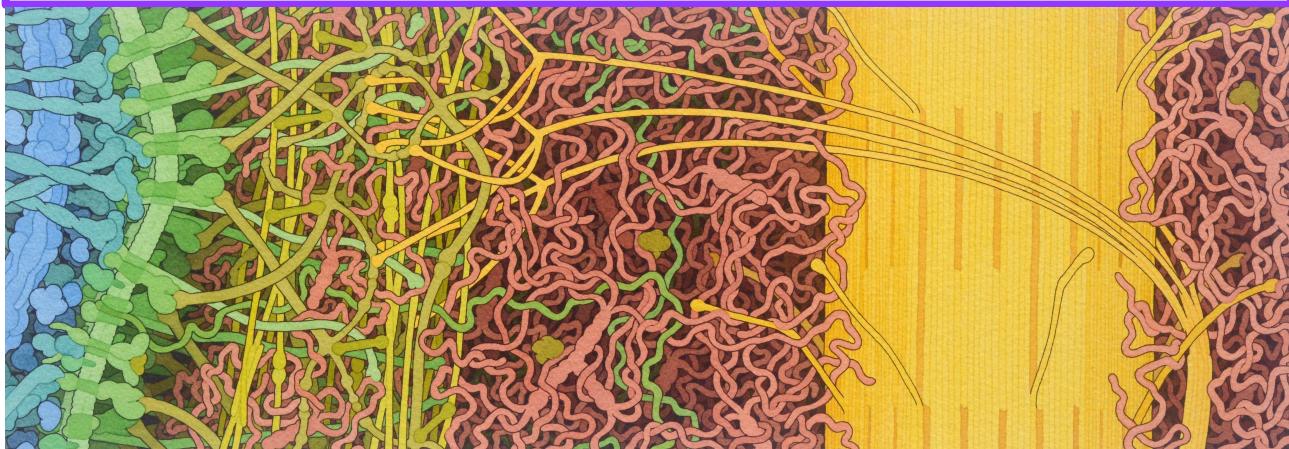


Figure: Collagen and Extracellular Matrix, by Goodsell (2021), an example illustration of the very intricate information flow in cells, how can these chemical macro-molecules and molecules interact and work as a machinery?. Image reused here with other context ([CC-BY-4.0](#) license).

How can the cells even function, why can the structural information in the genome, code for proteins structural information of the proteome. How can all the structural machines distinguish where to bind and how does the very intricate dynamics and mechanics play a role in the machinery?

The estimated number of particles in the universe is an exponential number 10^{80} (Eddington Number, 2022). For a single yeast cell, the interactions of molecules have been estimated to be more complex than an interaction combinatorial higher than the exponential number $10^{79000000000}$. This

Extracellular matrix: is a complex three-dimensional network outside of the cell surfaces, that facilitates movement and interconnect cells, it contains water minerals, and macromolecules such as collagen, enzymes, glycoproteins etc that provide structural and biochemical support to surrounding cells.

Exponentantial number: A number multiplied with itself. The number 10^{80} is 10

is indeed a number one cannot even fathom (Tompa and Rose, 2011). It is important to remember that in this vast combinations is order. There is still so much for humanity to learn regarding how a cell or a multicellular creature can have such an complexity, it is immensely spectacular how molecules can order and maintain a stable yet very dynamic cell!

The underlying complexity of structural biomolecular machinery, points to patterns, that ML algorithms may solve eventually with the arrival of more and more structural and biological information.

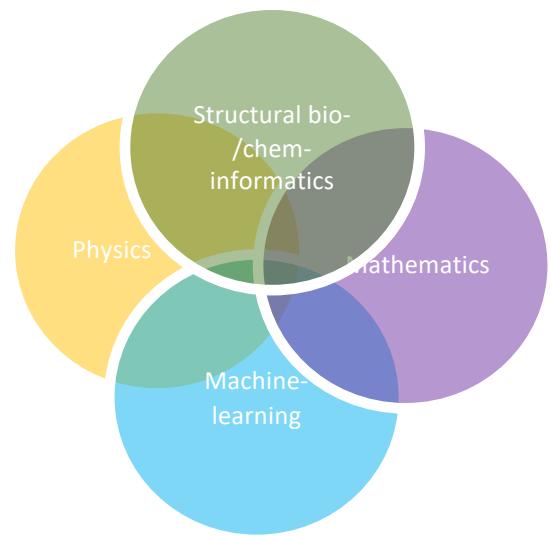
multiplied by itself 80 times, and $10^{79000000000}$ is 10 multiplied by itself 79 billion times.

Interaction combinatorial: The number of potential interactions increases exponentially with e.g. proteome size and other molecules in a cell, for instance 2 proteins have a smaller interaction combinatorial than 5 proteins.

THE SUCCESS OF MACHINE-LEARNING IN STRUCTURE PREDICTION POINTS TO AN UNDERLYING RELATIONSHIP BETWEEN INFORMATION TYPES IN BIOCHEMICAL SYSTEMS ?



Figure: Illustration of intersection of example scientific disciplines which may discover underlying patterns of life. Interestingly, for example research funded by DeepMind presented in the article **Advancing mathematics by guiding human intuition with AI** of a method that can obtain new fundamental results in pure mathematics discovered with the assistance of ML —demonstrating a method by which ML can aid mathematicians in discovering new conjectures and theorems (Davies et al. 2021; Stump 2021).



Even though AI in structural prediction is good at producing accurate results, it does not, at least not for now, explain how, or why, those results happened. Hence in the basic science of structural biology, key problems remain to be solved. The various people contributing to AI structure prediction and elsewhere should be commended for crucial breakthroughs. Yet there is still much work to be done to discover the essential biology, chemistry, mathematics and physics of how and why proteins fold. Text rewritten from *(Artificial intelligence in structural biology is here to stay, 2021)*.

That a ML model can translate or transform between two different information types, as in AlphaFold gene-information to structure-information of a protein, may indicate a mathematical pattern or equation connecting these types of information. Further investigation and study will need to be done to infer the fundamental understanding, by visualising patterns. Possibly it will be discovered and formalised fundamental equations, such as in other fields of physics ? It may be that deep learning methods such as Graph Neural Networks will help to infer such fundamental relationships.

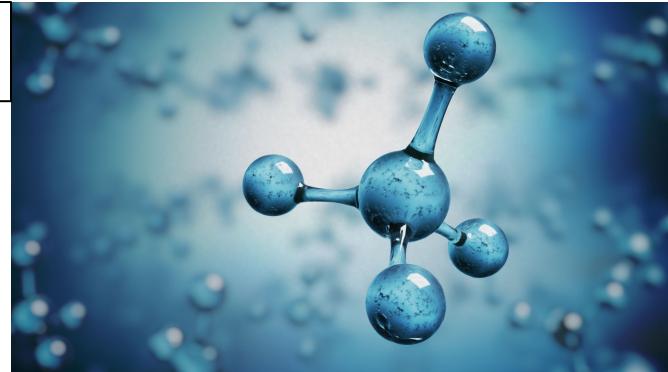
Fundamental equation: In the discipline of physics, there are reliable equations in every field to relate physical measurable quantities to each other and perform calculations and make predictions from them. These have often been discovered by finding patterns in nature.

For example J. C. Maxwell et al.

derived 4 fundamental equations in electromagnetism, preceded by M. Faraday et al., having found patterns and relationships related to experimental electronics and magnetism. A main reason we have the electronic capabilities today in the 21'st century.



Figure: An artistic representation of molecules.



Organic chemical reactions is well described by quantum chemistry (Ball, 2011). Biochemical cells are essentially composed of organic molecules that make up these immensely well-ordered smaller to gigantic organic macro-molecular assemblies. Hence it appears certain that organic chemistry and physical laws governing bio-chemistry applies also to cellular machinery, that there are many hidden quantum mechanical features of biochemical systems to discover; where the structure and dynamics of bio-molecules will be of essence in understanding these effects.

Many quantum effects appear in biochemical systems, for instance in the exchange/tunneling of hydrogen atoms between water molecules and protonable groups in biomolecules . Another important example is chiral-induced spin selectivity of electrons, which affect the biochemistry and physics of the cell, utilizing the ubiquitous handedness or chirality of biomolecular structure in biological systems (Michaeli et al. 2016; Tour 2016; Naaman & Waldeck D., 2017). For example, the components from which proteins are made—the amino acids—are enantiomerically pure in all organisms (Enantiomer, 2022); this

purity is extremely well preserved with high fidelity. Moreover, the enantiopurity of drug medications is known to be essential to their efficacy. And without it, side effects are common and sometimes deadly. The mid 19'th century disaster of drug Thalidomine, was due to not knowing the importance of enantio-purity.

AI is developed for usage in simulating quantum chemistry, moreover, to understand quantum physics and chemistry at a fundamental level. More and more sophisticated understanding will eventually most likely be incorporated in AI of structural biology and biochemistry to increase our understanding of the chemical reactions and mechanics of the structural biomolecules making up our cell's complex machinery.

Since drug-molecules and their target macro-molecules obey the rules of chemistry and quantum effects in how they react with target macro-molecules a better understanding of quantum physics in structural biology will most certainly elucidate understanding and drug discovery in the years to come.

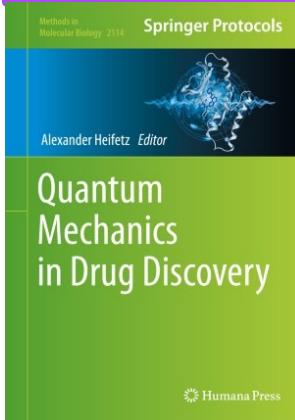


Figure 1: Book devoted to Quantum Mechanics in Drug Discovery (Heifetz, 2020).



Figure 2: Article from DeepMind, with title "Simulating matter on the quantum scale with AI", (Kirkpatrick J. et al., 2021). AI shows potential to aid other fields such as quantum chemistry and structure-based drug design.

It is becoming increasingly popular with the application of quantum mechanical (QM) methods in drug discovery. Improvements in computing power, have led to significant improvements in QM algorithm development and new applications of methods. QM methods are being routinely applied to many problems in computer aided drug discovery problems, for example characterizing protein-ligand, protein-water-ligand and protein-protein interactions, giving estimates of binding affinities, ligand energies and bioactive conformations, moreover they are used in refinement of molecular geometries, scoring docked protein-ligand poses, describing molecular similarity, structure-activity relationship analysis, and ADMET prediction. Methods of QM are increasingly incorporated and even replacing classical molecular mechanics methods because there is no need to use MM approximations when QM calculations with much higher accuracy can be performed at similar speed. *Rewritten from Preface in book (Heifetz, 2020)*

Functional: A functional is a function of a function. In DFT the functional is

The Schrödinger equation governs the behaviour of quantum mechanical particles making up any molecule, and also organic and biomolecular molecules relevant to drug discovery. Knowing the probability for any electron to be at each position (i.e., the electron density) in molecules is sufficient to exactly compute all chemistry. In density functional theory a functional converts the electron probability or density to interaction energies, to better understand chemical reactions. It has already been shown that by expressing a functional as a neural network and incorporating exact properties into training data, it has been learned functionals free from important systematic errors: resulting in a better description of a broad class of chemical reactions. *Text from (Kirkpatrick J. et al., 2021)*.

In addition, even other approaches such as FermiNet shows great results of applying neural networks to solve the Schrödinger equation accurately in chemistry (Pfau D., et al., 2020).

the electron density which is a function of space and time.

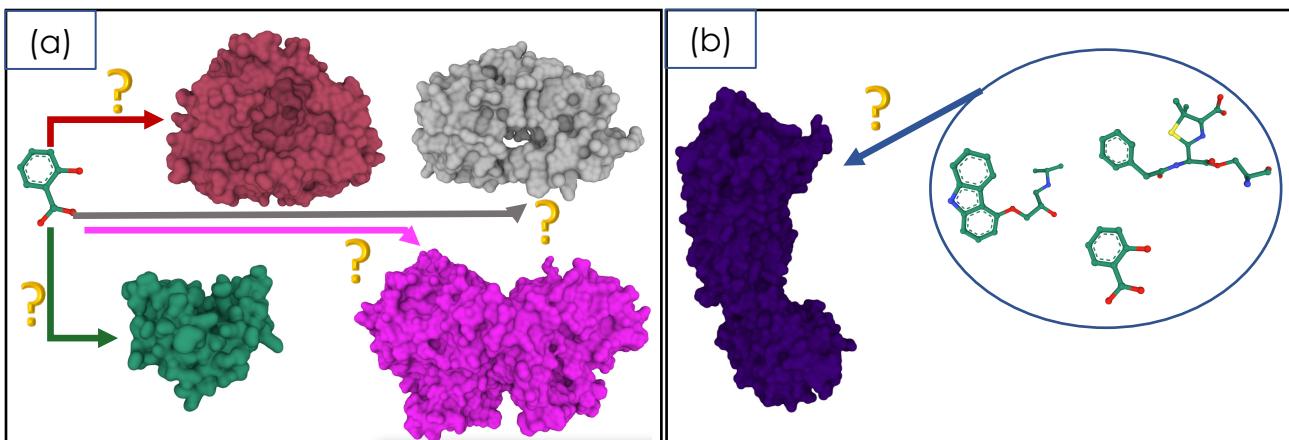


Figure 1: Molecular docking processes, two types. (a) Inverse Docking, this process involves e.g. finding the optimal protein target for a given ligand. (b) Classical molecular docking, involving e.g. finding the optimal ligand for a given protein.

Protein and ligand structure in this picture from PDB entries 1pws, 2hr1, 1bp2, 1hxb, 1pth. Structure renderings made in RCSB 3D viewer, Mol (Sehnal et al., 2021). Figure inspired from (Crampon et al., 2022).

It is an aim to reduce both cost and time associated with drug development for researchers and pharmaceutical industries. To underline only the most relevant drug candidates, molecular docking can be used as a filter. In addition molecular docking can be used to detect potential drug side-effects and toxicities of molecules. Molecular docking uses 3D structures for predicting the favoured orientation of a ligand bound to a macromolecule when forming a stable complex.

Commonly the ligand is the smallest molecule, yet the designation is depending on the system under study and aim of study. The ligand (or drug) is usually a small molecule, and the target is a macromolecule (e.g., a protein or RNA). The pairing can also cover a wider range of docking possibilities: protein to DNA; protein to RNA; sugar to protein, peptide to protein; etc.

Covering an important selection of

existing methods, protein to small compounds (protein to ligand molecular docking) is the most common.

For drug discovery thorough testing towards a comprehensive ligand library or virtual screening is often required. In Figure 1a one ligand is tested against multiple targets (Figure 1a), it can also be a fundament in research to find novel targets for a ligand (in the case of drug repurposing) or to characterize potential side effects, such as toxicity of a molecule. Nevertheless detecting the optimally bound ligand among a ligand database is an important use for molecular docking (Figure 1b).

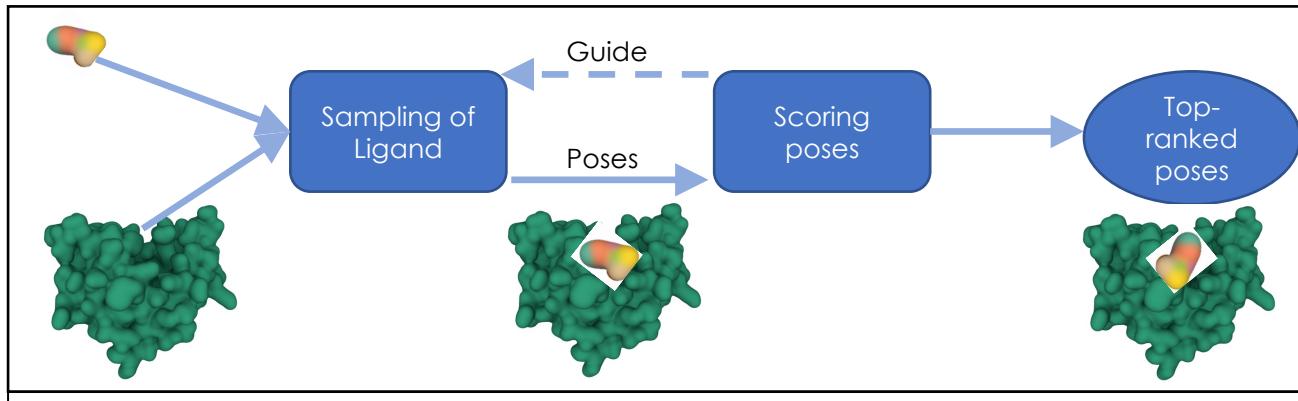


Figure: A simplification of a workflow of the general molecular docking process with sampling and scoring subprocesses.

To acquire the binding affinity and conformation of a ligand, experiments of molecular docking can be used.

Sampling is one main process, that is to say, generating a group of conformations from a starting rigid structure of a ligand, supposed to find all possible conformations. An important evaluation of a molecular docking method is its capacity to explore (or sample) this theoretical conformational space of a starting rigid ligand

Scoring is another main process, evaluating what is called a pose (i.e. an orientation of ligand to the target), here also evaluating the particular binding affinity (or strength of binding) of the pose of ligand and target. The estimation of binding affinity in a pose is made by a scoring function (SF) which is a mathematical function. There can be significant correlation between sampling and scoring, even if they introduced separately, since the scoring function often guides the sampling method.

Some main challenges for any workflow of a molecular docking process (Figure)

are molecular flexibility and the realistic portraying of binding binding; it is a trade-off between accuracy of sampling and scoring versus reasonable computing time. A summary of the problems associated with molecular docking, current challenges, and approaches to address them without ML is covered in Crampon et al. (2022), moreover they revealed that machine-learning and particularly deep learning (or neural networks) has shown to be able to help handle these challenges.

The sampling challenge is understudied and still unsolved in ML methodology. Whereas the scoring challenge is, unquestionably the most studied problem. Actually, ML scoring methods are interesting in terms of scoring function space exploration; wherein many ML methods has been developed, most of which surpass classical methods.

The review of Crampon et al. (2022) list the molecular docking software with or without ML approaches in tables, and it is a good read if you are interested in further study of this field.

Scoring function: is a mathematical function used to approximately predict (or score) the binding affinity between two molecules (say a ligand and a target protein) after docking.

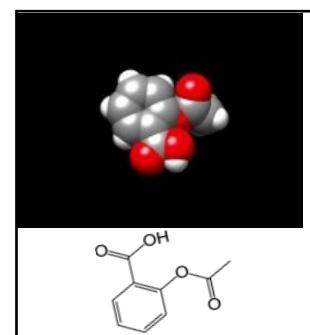
Scoring function space: is a concept which purpose is to find the mathematical model that best

predicts (or scores) the binding affinity of two molecules (say a ligand and target protein, from a set of possible mathematical functions. The concept is similar to molecular chemical space (all possible molecules adhering to a set of construction principles).



Figure: Below: Skeletal formula of Aspirin.

Above: Video of classical molecular dynamics of aspirin. Here aspirin is shown in 3D all-atom showing the hydrogens in white, oxygen red, and carbons in black, oriented in the same direction as the skeletal formula. The vibrations and rotations of bonds are apparent, where only the single covalent bonds represented by one line in the skeletal formula is able to rotate the connecting moieties. Video made with software UCSF Chimera, reused with permission, [Link to movie](#) or see reference (Timothy Siniscalchi, 2020).



Classical or Newtonian mechanics, aka classical Molecular dynamics (MD), is a simulation technique where the mathematical/physical model used to simulate molecular behaviour is called a force-field. For an example of molecular dynamics see Figure 1 (<https://youtu.be/Ow79mbbv7tY>).

There are many examples of studies (Glaser, 2015; Lei et al., 2015), that have investigated the structure and binding mechanisms of aspirin, using either separately or in combination Quantum Mechanical (QM) and classical aka molecular mechanical (MM) methods, that attempts to aid our understanding of this molecule and its chemistry.

What is fascinating is that classical Molecular Dynamics starting from phospholipase and aspirin molecules in isolation can find and reproduce the same binding area for aspirin, as in the X-ray structure 1OXR, where it predicts the binding pocket. This would be a more difficult task for aspirin binding to cyclooxygenase, where aspirin creates a covalent

bond to a serine residue, in the active site, requiring quantum mechanics to describe this.

Classical molecular Dynamics is a technique that first solvates the systems with water and relevant ions, then energetically relaxes the system under physiological conditions. Then simulating for a time relevant for the chemical interactions taking place. MD is a statistical technique meaning that it may not produce the exact same interactions each time, and will need confirmation from repeating simulations or experiment, depending on the size and complexity of the system under study.

MD simulation lengths has increase from picoseconds in the late 1970s to microseconds in the 2020s. The size of biochemical systems possible to simulate has also increased, the greater the size of biochemical system in a simulation the less time one can calculate in the simulation.

Force-field: A computational method in chemistry modelling estimating the forces between atoms within molecules and between molecules. "Force-field" refers to the mathematical form and sets of parameters used to calculate the potential energy of a system of atoms or groups of atoms. Parameters for a force-field may be derived from experiments in physics and chemistry, calculations in

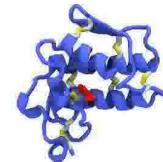
quantum mechanics or both.

Pico-second: is an time-unit, designated by the prefix pico- and the SI unit second, equal to: $10^{-12} = 1/(1\ 000\ 000\ 000\ 000) = 0.000\ 000\ 001$ (one trillionth) of a second.

Mill-second: $10^{-3} = 1/(1\ 000) = 0.001$ of a second.

Figure: [Link to movie](#) showing a Molecular Dynamics simulation of aspirin's and a target macro-molecule phospholipase A2 starting from PDB 1OXR, but with aspirins initially relocated randomly in simulation-box outside of binding pocket.

Where this particular binding causes a pharmacological effect of interest. Showing that also water and an calcium ion molecules plays a role, in addition to structural and dynamical features of the binding pocket, (Michael Kuiper, 2015).



Among the first MD simulations including all atoms of a biomolecule was achieved in 1977 of a protein of 58 amino-acid residues which was simulated for 9.2 picoseconds; from 2020's it can be attained up to milliseconds simulations of solvated proteins of similar size, (Salo-Ahen et al. 2021).

Classical MD simulations will be able to in this decade reach sampling of larger time and size-scales, producing petabytes of simulation data at current force-field accuracy.

As it is, classical MD will still be in the regime of low-throughput, high-latency predictions with average accuracy. However considering the progress of AI, it has been envisioned that ML will be able to solve both the accuracy and shorten the computation-time-to-prediction-problem by training predictive models using expensive simulation data.

It is mainly the development of computer and model development that has enabled the rise in MD simulation time across the years. Hence the continuing development

and quantum computing etc, may also play a significant role in the simulation field.

Considering model development alone, the harmonising classical and quantum mechanical simulations with ML methods such as ANN; have the potential to drastically reshape the way we make predictions in computational structural biology and drug discovery.

Since high accuracy and computationally expensive simulation models can be obtained at shorter intervals; this can be used to train simulation models with high predictivity of other biochemical systems to achieve longer simulation times.

An example of ML implemented insofar in MD is replacing expensive QM calculations to predict dihedral energies given a NN already trained on data from QM simulations. Text rewritten (Peres et al., 2018).

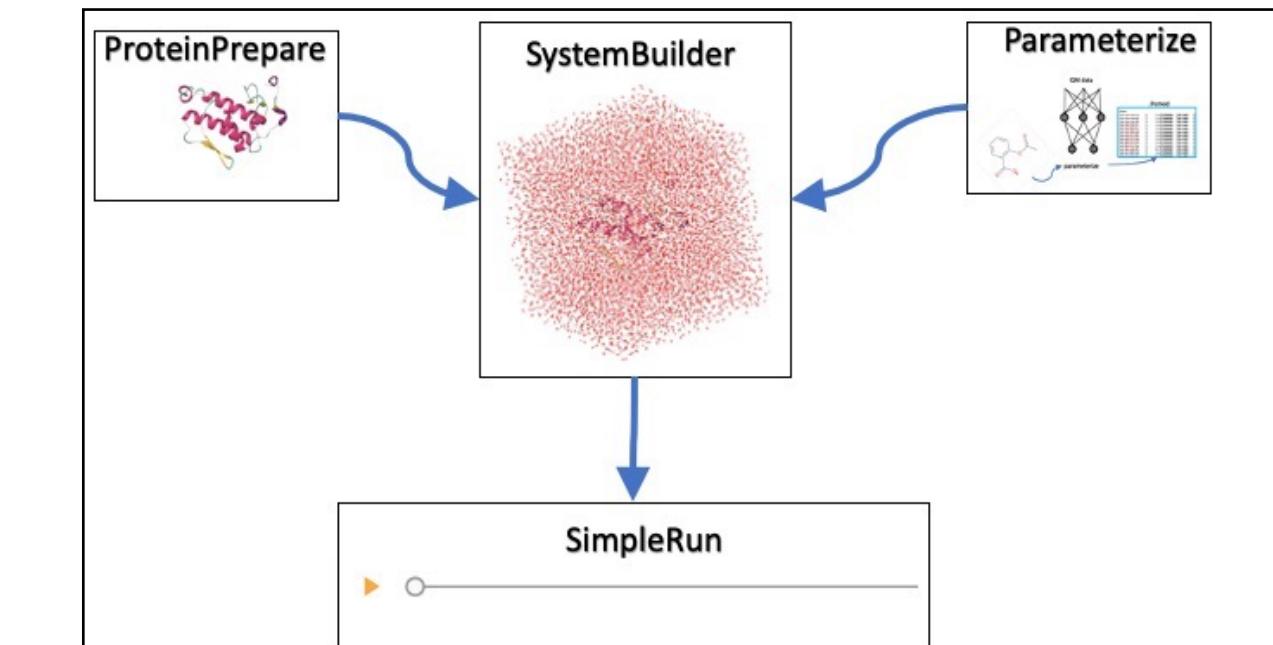


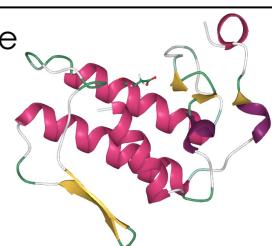
Figure 1: PlayMolecule molecular dynamics workflow. Applications used at the web-interface of PlayMolecule to prepare and run an simulation of dynamics a macro-molecule and ligand.

For MD, there are a multitude of simulation software and code-frameworks for doing these types of simulations, however similar in core methods. Other examples are GROMACS an open-source MD software. There are also a variety of force-fields, for example AMBER, or parameters of which the physics simulation model is adjusted from. Using ML learned parameters is a newer occurrence in this field.

PlayMolecules MD workflow (see Figure 1) consists of a set of

applications that have the capability to confer information amongst them — that is to say the output of one application can be used as input for another application. In order to build and run an MD simulation, we must use a number of applications in succession. We will describe them in order here, in accordance with a tutorial, which one may read also if interest (Martinez G., 2019b).

Figure 2: A protein phospholipase A2, prepared by playmolecule ProteinPrepare. Obtained by inputting structure PDB 1OXR.



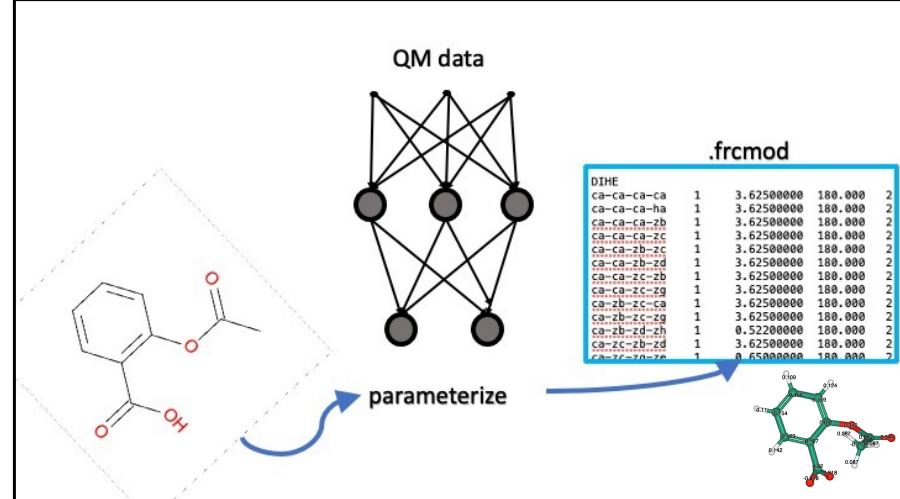
Essentially, by using ProteinPrepare, one inputs an initial protein PDB structure and obtains a protonated and optimized structure that one is able to use in SystemBuilder at a later step, see Figure 2 on previous page.

This application allows the user to prepare a protein extracted from RSCB Protein Data Bank (PDB), make it ready to build a system and run molecular dynamics simulations. Preparation process includes the titration of the protonation states

using python code *PROPKA*. The first can be manually overidden by examining the protein in the 3D webGL viewer. Several graphs can aid with information to optimize the protonation states at a certain pH.

Moreover the addition of missing atoms and overall optimization of the hydrogen bond network is made using python code *PDB2PQR*. More detail if interest at tutorial (Martinez G., 2019a).

Figure 3: Schematic of how the app Parameterize, can generate parameters from QM data by using Neural networks. Made in ppt using inspiration from (Martinez G., 2019b; Galvelis et al. 2019)



In Figure 3 it is illustrated the program-unit Parameterize by which we can obtain force-field parameters that will be used for simulating how a ligand acts in an MD simulation (Martinez G., 2019b).

Parameterize provide AMBER-compatible force-field parameters.

Parameterize is fairly complex, but in essentially it is taking care of:

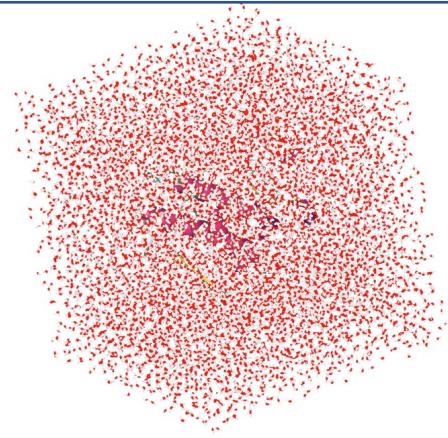
- Computing the partial charge of the molecule atoms.
- Assigning general parameters

(using the GAFF2 forcefield).
iii. Optimizing dihedral parameters.

Step iii here is important because general force-fields like GAFF2 are not really accurate enough when guessing the dihedral angle terms.

Parameterize can obtain dihedral terms in different ways, i.e. Dihedral scanning and parameter fitting using the ANI-1x NN potential.

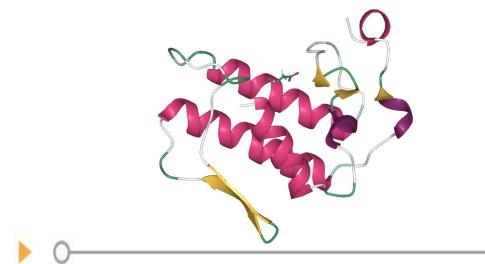
Figure 4: SystemBuilder, a built box of ligand aspirin, receptor phospholipase A2, water and ions.



The next step in the MD workflow is to build the biochemical system of protein, ions and water-molecules. Building a system means solvating (with water), ionizing and obtaining force-field parameters for the whole MD system. We can do so by using the SystemBuilder application, which is built on top of the Python framework HTMD.

In this case as shown in Figure 4 , we included the protein previously prepared with ProteinPrepare and the ligands are parameterized with Parametrize. For instance, in our case we selected the protein of PDB entry 1OXR, included ions and the ligands we identified as aspirin and calcium.

Figure 5: An illustration or snapshot from calculated Molecular Dynamics trajectory or to say a movie of the chemical dynamics of phospholipase A2.



After the system is built in SystemBuilder, we can proceed to the SimpleRun application (Figure 5), which uses an MD engine ACEMD to run an MD simulation.

The SimpleRun application essentially follows this three-step process:

- i. **Minimization.** Forces and potential energy are minimized, structural clashes are solved, etc.
- ii. **Equilibration**, that calibrates the system to the right pressure and

temperature. Here MD is run for few nanoseconds with some constraints so that the system doesn't blow up).

- iii. **Production run.** Here we calculate a trajectory of the dynamics with MD without any constraints, where each step consists of a snap-shot of the dynamical movement.

Figure 6: PlexView application at Playmolecule.

Plexview (Figure 6), makes 2D diagrams of protein-ligand interactions, it calculates the interactions from geometrical considerations.

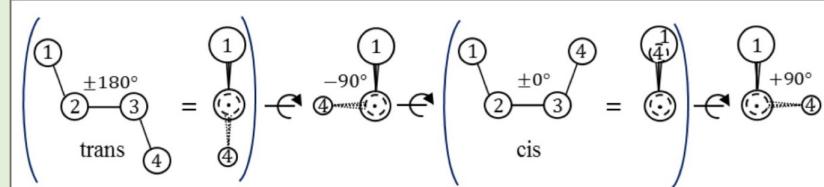
The currently supported interactions include:

- Hydrogen bonds.
- Pi-pi stacking (including face-to-

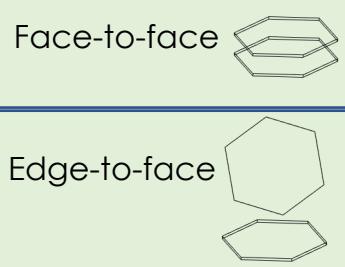
face and edge-to-face).

The ligand structure needs to be already protonated to get correct results from PlexView. Otherwise hydrogen bond/donor interactions might be missing. If not already protonated, the protein will be automatically protonated by PlexView at pH 7. (Text from web-application at PlayMolecule).

Dihedral angle Definition of a dihedral angle of bonded atoms . From left to right: clock-wise +90° rotations of atom 4 relative to atom 1, where the central bond of atoms 2-3 is the rotating axis.
Reprinted figure from (Wittler et al. 2019).



Pi-stacking: also called π - π stacking, in biochemistry it refers to attractive, noncovalent interactions between aromatic rings, since they contain pi bonds (a type of a covalent bond). For example benzene aromatic rings can pi-stack in e.g. these orientations:



Workflow: Some 3 basic components are of input, transformation, output of a workflow diagram. Every step within a workflow may be assigned as input, transformation and output. Workflows are applicable in many fields such as Neural Network prediction, molecular dynamics and other types of programming driven handling of data, such as related to simulation, processing and calculation.

UQ 5 - Understanding Questions to answer at Studium.

Session 7.1

6 UQ



Session 7.2

6 UQ





Exercise 7 (Optional not graded)

AI TO AID MD SIMULATIONS. EXAMPLE PREPARE AND RUN RECEPTOR PHOSPHOLIPASE A2 AND LIGAND ASPIRIN WITH ML PARAMETERS

Methods: Neural Networks, Molecular Dynamics Preparation and Simulation.

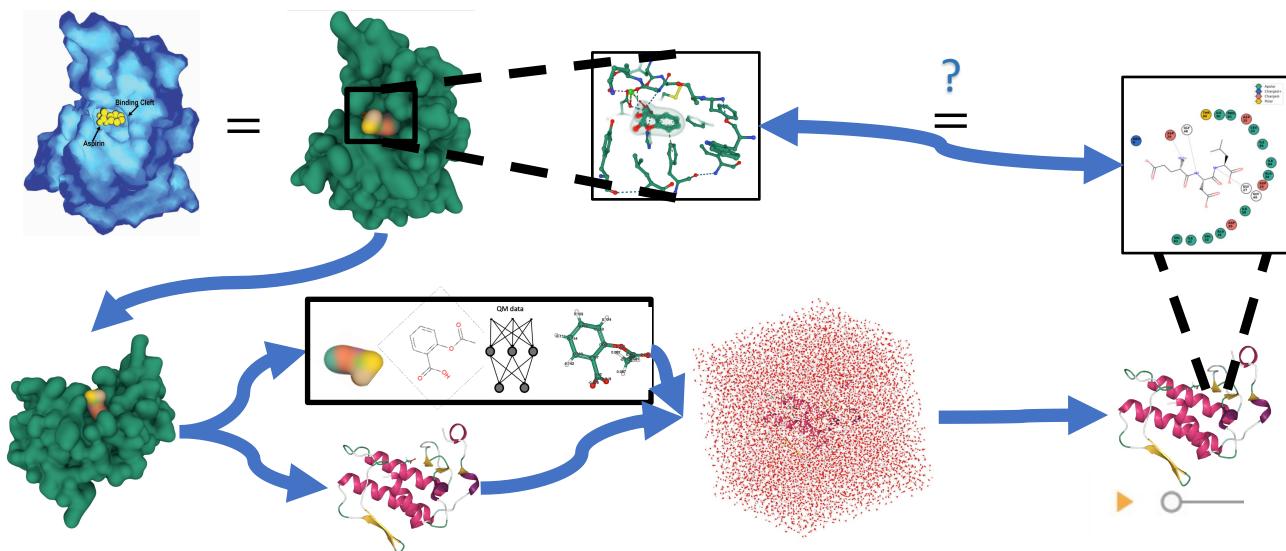


Figure: Abstract image for EXERCISE 7: We will here and in following pages, present some narrative for this exercise, but the exercise itself is on Studium.

NARRATIVE STORY

First we present some overall knowledge of this drug-target macromolecule protein phospholipase A2, some information from a crystal structure with structural details of aspirin bound.

- **NARRATIVE TASK 2**
We will use the same structure and prepare for MD simulation, with PlayMolecule applications.

- **NARRATIVE TASK 1**

Here we will investigate the 3D structure using the inherent 3D viewer at the protein data bank. We will investigate the macromolecule, in particular the binding interactions of aspirin.



Aspirin induces its anti-inflammatory effects through its specific binding to phospholipase A₂: Crystal structure of the complex formed between phospholipase A₂ and aspirin at 1.9 Å resolution



Journal of Drug Targeting

Figure 1: An article (Singh et al., 2005) describing a structural target of aspirin located in the PDB entry 1OXR.

ISSN: 1061-186X (Print) 1029-2330 (Online) Journal

The article (Singh et al., 2005) describes one structural target for the drug aspirin, it also describes the active site that binds aspirin.

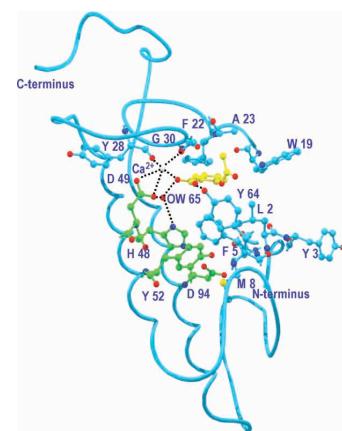
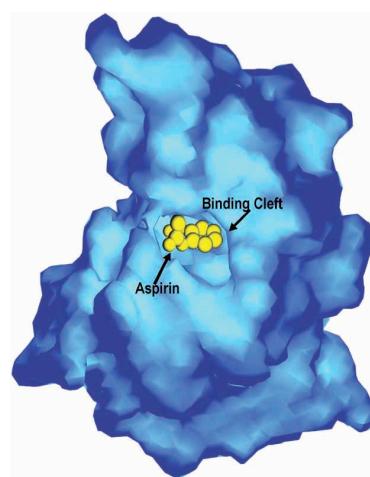
A clear binding pocket with aspirin bound was solved as a crystal structure, now located in PDB 1OXR.

Where the structural and kinetic data, clearly indicated that aspirin binds with a dissociation constant ($K_D = 6.4 \cdot 10^{-6}$ M).

The protein structure in PDB 1OXR is from cobra (or snake), however homologous in structure to the bovine (or cow) structure in PDB 1BP2.

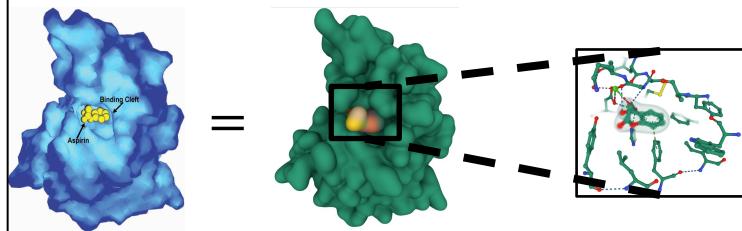
A lot has been understood about the physiological effects of aspirin, useful information. However, in general in biochemistry of systems the processes could be much more elucidated and clarified of how aspirin and other molecules affects cascade signalling pathways in the cell, to be understood about the atomic interactions moreover with the target enzymes, for example by understanding the elaborate chemistry and dynamics of these protein structures is a not complete picture yet.

Figure 2: From article (Singh et al., 2005), showing aspirin bound to a hollow pocket in phospholipase A2; (all-atom surface cartoon at left); (ribbon of amino-acid chain with atoms of aminoacids and aspirin in the hollow pocket at right).





Graphical Abstract for Question 1: The protein phospholipase structure with aspirin bound to a target binding pocket, solved in PDB 1OXR. The green structures are obtained directly from www.rcsb.org for entry 1OXR, and with visualizing it in the 3D-viewer.



In this exercise we will merely investigate and obtain the ligand interactions of aspirin in its binding pocket, from the crystal structure.

To be aware however, bound ligands in structure-determined proteins may have minor or larger change in binding interactions, than what is most likely in a biological setting where there is other particular temperature and chemical environment.

This affects the interpretation which interactions are most important when considering why aspirin binds in a physiological relevant biological conditions and systems.

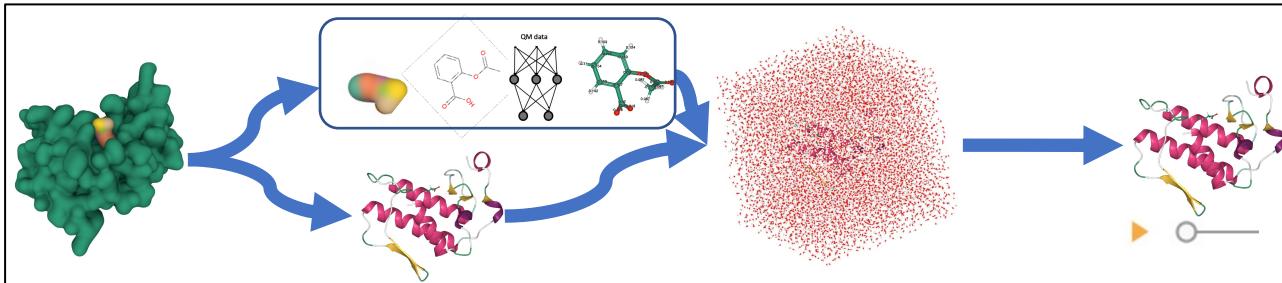
Moreover this protein active site is faced towards a phospholipid membrane, which is not reflected in the crystal structure here, this is nevertheless possible to simulate by molecular dynamics though in this exercise we will not regard the inclusion of the membrane, however key insights into aspirin binding to

PLA2 can nevertheless be obtained even though it does not perfectly reflect biological conditions.

Moreover macro-molecules are incredibly dynamic, but the structure predictions from X-ray or AlphaFold may often be considered to be a good average picture of the dynamic interaction between ligand and receptor.

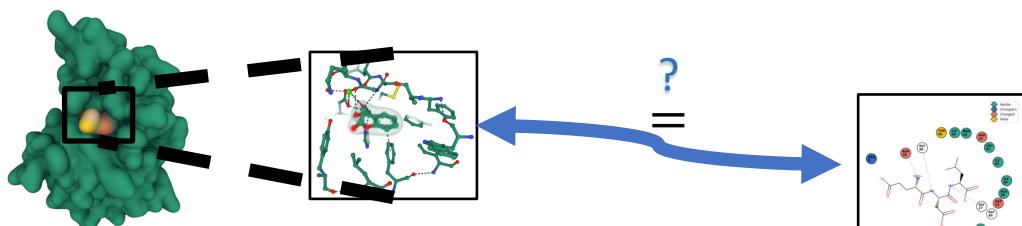
The binding interactions can still be accurate and trustworthy as a starting point for structure-based drug design.

Solvent such as water (Red dots in picture) or ions are included in structures, the calcium ion (green) is a mediator between aspirin, water and the binding pocket.



Graphical Abstract for Question 2: MD workflow of simulating PLA2 with ligand aspirin bound, using ML parameters for the ligand, starting from PDB 1OXR.

1. Basically here we will input the structure from PDB 1OXR to prepare a protein structure to be used for MD simulation.
2. From PDB 1OXR we can also obtain a .sdf file to be used in training parameters from the playmolecule application Parametrize.



Graphical Abstract for Question 3: Comparing the ligand receptor interactions between the MD simulation, and the crystal PDB structure, using PlexView for the target simulation and 3D Viewer of PDB 1OXR.

Please follow the instructions at Exercise 7 at Studium to complete the Questions with answers.

References (Harvard citation style)

- Artificial intelligence in structural biology is here to stay, 2021. *Nature*, 595(7869), 625–626. Available at: <https://doi.org/10.1038/d41586-021-02037-0> (Accessed 06 April 2022)
- Ball, P., 2011. Physics of life: The dawn of quantum biology. *Nature News*, 474(7351), pp.272-274. Available at: <https://doi.org/10.1038/474272a> (Accessed 22 April 2022)
- Davies, A., et al., 2021. Advancing mathematics by guiding human intuition with AI. *Nature*, 600(7887), pp.70-74. Available at: <https://doi.org/10.1038/s41586-021-04086-x> (Accessed: 12 April 2022)
- Eddington Number, 2022. Wikipedia. Available at: https://en.wikipedia.org/wiki/Eddington_number (Accessed: 05 April 2022)
- Enantiomer, 2022. Wikipedia. Available at: <https://en.wikipedia.org/wiki/Enantiomer> (Accessed: 05 April 2022)
- Galvelis et al. 2019. A scalable molecular force field parameterization method based on density functional theory and quantum-level machine learning. *Journal of chemical information and modeling*, 59(8), pp.3485-3493.
- Glaser, R., 2001. Aspirin. An ab initio quantum-mechanical study of conformational preferences and of neighboring group interactions. *The Journal of organic chemistry*, 66(3), pp.771-779. <https://doi.org/10.1021/jo001241s>
- Goodsell, D.S., 2020. SARS-CoV-2 mRNA Vaccine 2020, Molecular Landscapes by David S. Goodsell. Available at: <https://pdb101.rcsb.org/sci-art/goodsell-gallery> (Accessed: 05 April 2022)
- Goodsell, D.S., 2021. Collagen and Extracellular Matrix, 2021, Molecular Landscapes by David S. Goodsell. Available at: <https://pdb101.rcsb.org/sci-art/goodsell-gallery> (Accessed: 05 April 2022)
- Goodsell, D.S. and Burley, S.K., 2022. RCSB Protein Data Bank resources for structure-facilitated design of mRNA vaccines for existing and emerging viral pathogens. *Structure*, 30(1), pp. 55-68.
- Goodsell, D.S., 2016. Insulin action 2016, Molecular Landscapes by David S. Goodsell. Available at: <https://pdb101.rcsb.org/sci-art/goodsell-gallery> (Accessed: 05 April 2022)
- Heifetz, A. ed., 2020. *Quantum mechanics in drug discovery*. New York: Humana Press.
- Available at: <https://link.springer.com/book/10.1007/978-1-0716-0282-9> (Accessed: 12 April 2022)
- Lei et al. 2015. Mechanistic Insights into a Classic Wonder Drug Aspirin. *Journal of the American Chemical Society*, 137(1), pp.70-73. <https://doi.org/10.1021/ja5112964>
- Martinez G., 2019a. How to prepare a protein structure using PlayMolecule® ProteinPrepare [TUTORIAL]. Available at: <https://medium.com/playmolecule/how-to-prepare-a-protein-structure-using-playmolecule-proteinprepare-tutorial-d69bd0a7c5b6> (Accessed: 05 April 2022)
- Martinez G., 2019b. Building and running a molecular dynamics (MD) simulation of a protein-ligand system on the PlayMolecule® web platform [TUTORIAL]. Available at: <https://medium.com/playmolecule/building-and-running-a-molecular-dynamics-md-simulation-of-a-protein-ligand-system-on-the-a852ddb3eb87> (Accessed: 05 April 2022)
- Michael Kuiper, 2015. aspirin dock v1sc1. 16 June 2015. Available at: <https://youtu.be/8XPlvNBNxdk> (Accessed: 05 April 2022)
- Michaeli, K., et al., 2016. The electron's spin and molecular chirality—how are they related and how do they affect life processes?. *Chemical Society Reviews*, 45(23), pp.6478-6487. Available at: <https://doi.org/10.1039/C6CS00369A> (Accessed 22 April 2022)
- Naaman R., Waldeck D., 2017, Spin in Quantum biology, *Inference*, 3(2). Available at: <https://inference-review.com/article/spin-in-quantum-biology>

Tompa, P. and Rose, G.D., 2011. The Levinthal paradox of the interactome. *Protein Science*, 20(12), pp.2074-2079. Available at: <https://doi.org/10.1002/pro.747> (Accessed 22 April 2022)

Tour, J., 2016. Chiral-Induced Spin Selectivity, *Inference*, 2(4). Available at: <https://inference-review.com/article/chiral-induced-spin-selectivity> (Accessed 22 April 2022)

Kirkpatrick J. et al., 2021. Simulating matter on the quantum scale with AI. DeepMind. Available at: <https://deepmind.com/blog/article/Simulating-matter-on-the-quantum-scale-with-AI> (Accessed: 05 April 2022)

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S.K., Koča, J. and Rose, A.S., 2021. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, 49(W1), pp.W431-W437.

Pérez, A., et al., 2018. Simulations meet machine learning in structural biology. *Current opinion in structural biology*, 49, pp.139-144.

Phospholipase A2, 2022. Wikipedia. Available at: https://en.wikipedia.org/wiki/Phospholipase_A2 (Accessed: 05 April 2022)

Pfau D., et al., 2020. FermiNet: Quantum Physics and Chemistry from First Principles. DeepMind. Available at: <https://www.deepmind.com/blog/ferminet-quantum-physics-and-chemistry-from-first-principles> (Accessed: 05 April 2022)

Timothy Siniscalchi, 2020. How to Create Simple Molecular Dynamics Simulations in Chimera. 28 October. Available at: <https://www.youtube.com/watch?v=AfcxEauzYak> (Accessed: 05 April 2022)

Salo-Ahen et al. 2021. Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. *Processes*, 9(1), p.71.

Singh, R.K. et al. ,2005. Aspirin induces its anti-inflammatory effects through its specific binding to phospholipase a2: Crystal structure of the complex formed between phospholipase a2 and aspirin at 1.9 Å resolution. *Journal of drug targeting*, 13(2), pp.113-119. Available at: <https://doi.org/10.1080/10611860400024078> (Accessed: 21 April 2022)

Stump, C., 2021. Artificial intelligence aids intuition in mathematical discovery. *Nature News and views*. Available at: <https://www.nature.com/articles/d41586-021-03512-4> (Accessed: 12 April 2022)

Wittler, H.P.A., 2019. A Molecular Dynamics Analysis of Insulin. Available at: <http://hdl.handle.net/1959.9/568798> (Accessed: 13 April 2022)