

Week 3 (page 1)

Lecture 5

Structural biology, introduction to NNs and prediction

First Section

STRUCTURAL BIOLOGY AND EXPERIMENTAL METHODS OVERVIEW

Experimental prediction of structures. Structure of life, structural biology and information omics, small drugs and macro-molecular targets.

Second Section

INTRO TO NEURAL NETWORKS AND STRUCTURE PREDICTION Protein Folding problem. AlphaFold AI structure prediction. Confidence intervals. Example prediction compared to true structure.

Quiz 5

UNDERSTANDING QUESTIONS

(6 for each session) to answer in Studium

Exercise 5

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

METHOD: ALPHAFOLD, NEURAL NETWORKS, 3D VIEWING
(Narrative here. The exercise itself is on Studium)



Graded material



Extra
Understanding

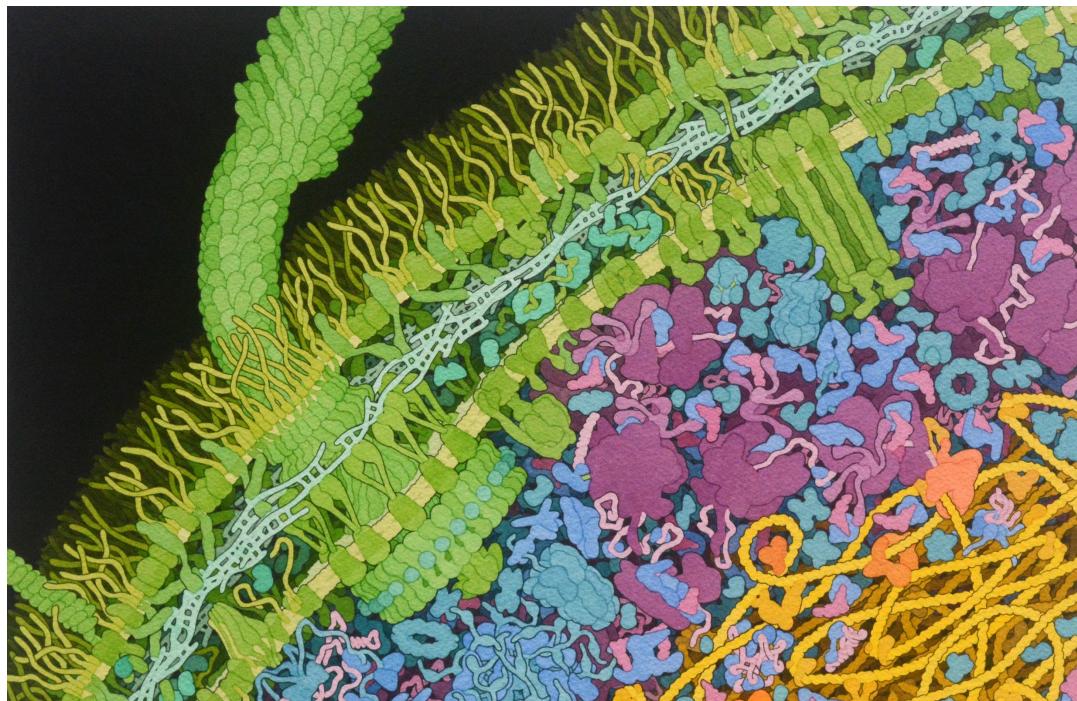


FIGURE: A cartoon of biomolecular structures in a bacterial cell. Note the beautiful machinery of a green complex rotor motor to the left.
Goodsell, D.S. (2016).

Simplified, life consists of four types of biomolecules in highly ordered sequential-chains or other combinations and chemical variations of these types of molecules; amino-acids, nucleic acids, carbohydrates and lipids. Along with ions, smaller organic molecules and water, all life consists of these 4 basic groups of building blocks (Berg, 2005):

Amino-acids
Nucleic acids
Carbohydrates
Lipids

The four groups of structural information of bio-molecules are ordered, oriented and have specific homo-chirality; analogy being like text writing to the right, i.e., in a certain direction (in contrast to otherwise unmeaningful writing letters to the left or writing right and left randomly). Life is composed of bonded atoms, small molecules bonded of tens of atoms, or macro-molecules bonded of hundreds to many thousands of atoms. These macromolecules are built of building blocks connected together. Laws of chemistry and physics governs the machinery of these atomic constructs in cells.



FIGURE: A cartoon of biomolecular structures in a cell, depicting insulin action binding and its triggering of cellular events, by Goodsell, D.S. (2016).



Life being composed of basic building-blocks, that comprises anything between small molecules to larger macro-molecules. All macro-molecules of life is built up by amino-acids, nucleotides, lipids and carbohydrates with chemical variations and combinations of these molecules. There are amazing structures showing beautiful symmetry in cellular biology, that are immensely well ordered in complexity; one can compare it to the specificity and meaning of all the books together in a very large library. The way molecules and macro-molecules assemble in cells is puzzling, how do all these thousands of dynamic molecules and macro-molecules form a dynamic molecular machinery in our cells? Cells are crowded yet extremely well ordered, for instance: in each cell we have about 2 nm wide and 2 meter

long DNA being coiled to fit in the nucleus or center of the cell. The human DNA sequence has 3 billion base-pairs of nucleotides.

Life is dynamic, cellular machinery is puzzlingly dynamic, and macro-molecules are crowded in cells, the cells duplicate, molecular parts including proteins are expressed, recycled continuously, and macromolecules are shipped between various parts of our body and within our cells, our cells can be likened to a human built city but is much more complex. See a simplified but instructive [movie from Harvard](#) (XVIVO Scientific Animation, 2011, 3:12)

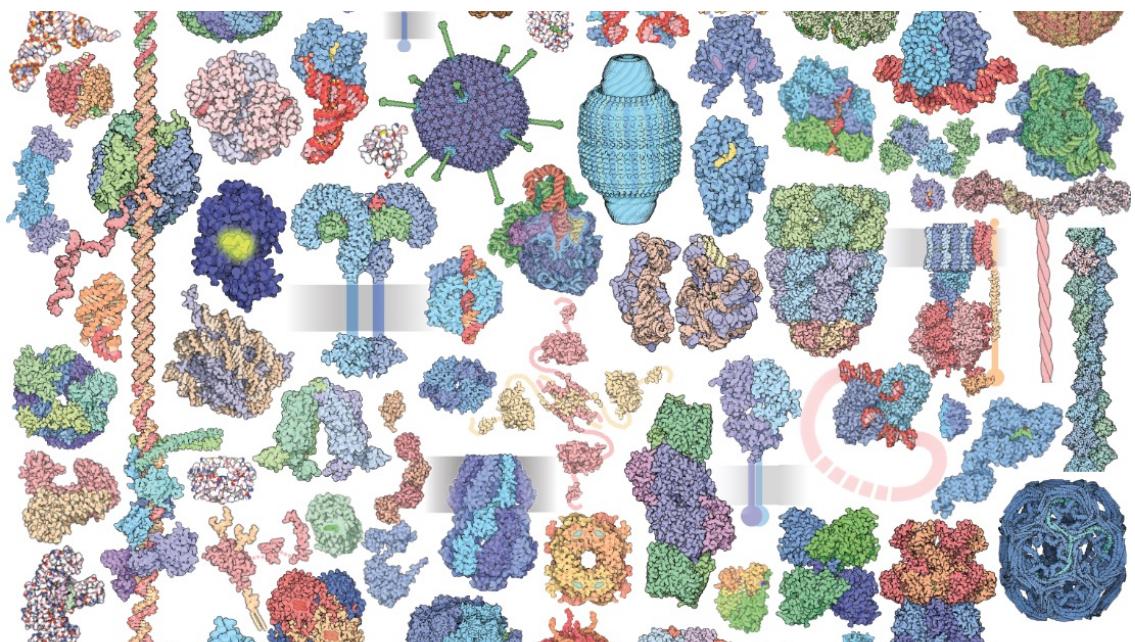


FIGURE: An artistic mosaic of different biochemical structures derived from the Protein Data Bank, by Goodsell, D.S. (n.d.).

Structural Biology is a branch of molecular biology, biochemistry and biophysics that is concerned with the structure of biomolecules (especially proteins which are made up of amino acids), RNA, DNA made up of nucleotides. Also carbohydrates which apart from energy are also bound in proteins, DNA and RNA. And lipids forming membranes, which forms isolating shells of cells and organelles.

Structural biology also relate to how a machinery functions, how parts of cells acquire the structures they have, and how mutations or alterations in their structures affect their function.

Understanding the mechanics, dynamics and chemical reactions of bio-molecules are also part of the field of structural biology.

Solving structures of biomolecules, both need to solve average conformations of relevant biomolecules, but it is also important to understand the macromolecules chemical dynamics, motion and physical mechanics in order to more fully understand its function in our body.

There are different levels of complexity in chemistry and physics of more or less complex molecules one can study and understand. However, for instance knowing only an approximate 3D structure and amino-acid sequence of a protein can give great insight into its function.



MOVIE: Example omics, Glycomics.

Reprinted with permission from Thom Leach and Griffith University. (Institute for Glycomics, 2021)



Knowing the structure, physics, molecular dynamics, chemical reaction mechanics and other quantities of any biomolecule, increases our understanding of biological fields of information, referred to with omics.

Omics (Omics, 2022) aims at the collective characterization and quantification of groups of biological molecules that translate into the structure, function, and dynamics of an organism or organisms. In any omics, the “ome” refers to a totality of some sort. Many different types of omics exist which aims to explain a subtype of biological life. Two examples are:

- Human genome - our complete set of genetic information encoded in DNA.

- Human proteome - the entire complement of proteins in our bodies, including the modifications made to a particular set of proteins, produced by our bodies.

Many other omics fields exist, see for example the [video about glycomics](#) above.

Structural biology, is relevant to Artificial Intelligence in drug discovery. Since we can quantify the chemical structure of life in an ordered way, we can gather data and use AI to further our understanding of life, since the data is intelligible, uniform and predictable, the data can point to patterns in how structure and information is related in cells.

HOW DO SMALL DRUGS WORK WITH TARGET PROTEINS ?



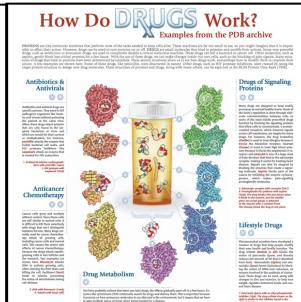
FIGURE: A poster from PDB, showing some smaller drug-molecules and structural macro-molecule targets related to various physiological conditions and diseases.

For instance poster subsection 6,7 “Drugs of Signaling Proteins”

Many drugs are designed to keep bodily processes at normal healthy levels. Much of the body's regulation is done through elaborate communications between cells, so some of the most widely prescribed drugs function by blocking the signaling proteins that allow cells to communicate. G protein coupled receptors, which transmit signals across cell membranes, are targets for many drugs. For instance, the drug loratadine (Claritin) is used to treat allergies because it blocks the histamine receptor (PDB 2RH1).

Signals can also be stopped by blocking the enzymes that create a signaling molecule. Aspirin blocks pain at the source by inhibiting the enzyme cyclooxygenase (PDB 1PTH), which makes pain-signaling prostaglandin molecules.

Text are from by Goodsell and Voigt (2008), one may [download the entire poster](#) and read more.



Proteins are tiny molecular machines that perform most of the tasks needed to keep cells alive. These machines are far too small to see, so you might imagine that it is impossible to affect their action. However, drugs can be used to turn proteins on or off.

Drugs are small molecules that bind to one specific protein and modify its action. Some very powerful drugs, such as antibiotics or anticancer drugs, are used to completely disable a critical molecular machine. These drugs can kill a bacterial or cancer cell. Other molecules, such as aspirin, gently block less-critical proteins for a few hours. With the use of these drugs, we can make changes inside our own cells, such as the blocking of pain signals. Many structures of drugs that bind to proteins have been determined by scientists. These atomic

structures allow us to see how drugs work, and perhaps how to modify them to improve their action. A few examples are shown in this article (Goodsell and Voigt, 2008). Some of these drugs, like penicillin, were discovered in nature. Other drugs, such as HIV protease inhibitors, were created by using the target protein structure to design new drug molecules. These structures of proteins and drugs, along with many others, can be explored at RCSB PDB and other related databases. Text from Goodsell and Voigt (2008)

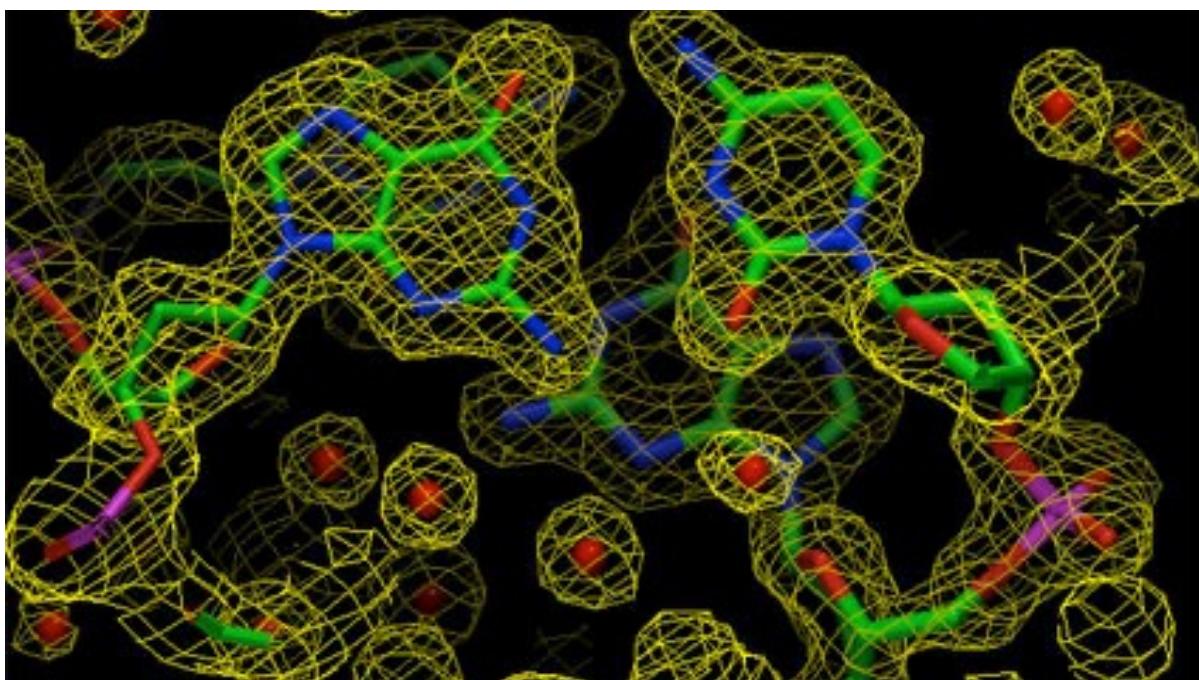


FIGURE: The experimental electron density from a structure of DNA is shown here, along with the atomic model that was generated based on the data. The contours surround regions with high densities of electrons, which correspond to the atoms in the molecule.

Several methods are currently used, to determine the structure of a protein: including X-ray crystallography, NMR spectroscopy, and electron microscopy. Each method has advantages and disadvantages. In each of these methods, the scientist uses many pieces of information to create the final atomic model. Primarily, the scientist has some kind of experimental data about the structure of the molecule. For X-ray crystallography, this is the X-ray diffraction pattern. For NMR spectroscopy, it is information on the local conformation and distance between atoms that are close to one another. In electron microscopy, it is an image of the overall shape of the molecule.

When looking at PDB entries, it is always good to be a bit critical. Keep in mind that the structures in the PDB archive are determined using a balanced mixture of experimental

observation and knowledge-based modeling. It often pays to take a little extra time to confirm for yourself that the experimental evidence for a particular structure supports the model as represented and the scientific conclusions based on the model.

The text and image for this slides are from an PDB-101 article ('Methods for Determining Atomic Structures', no date) where you may study further if interest.

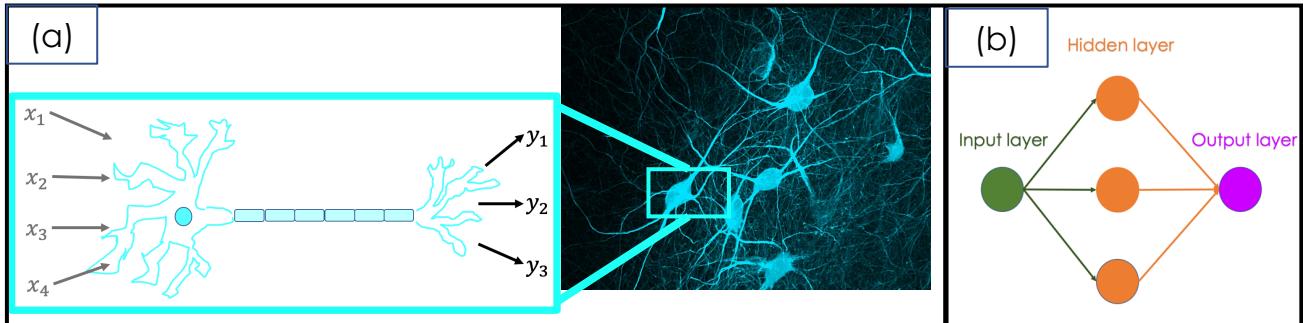


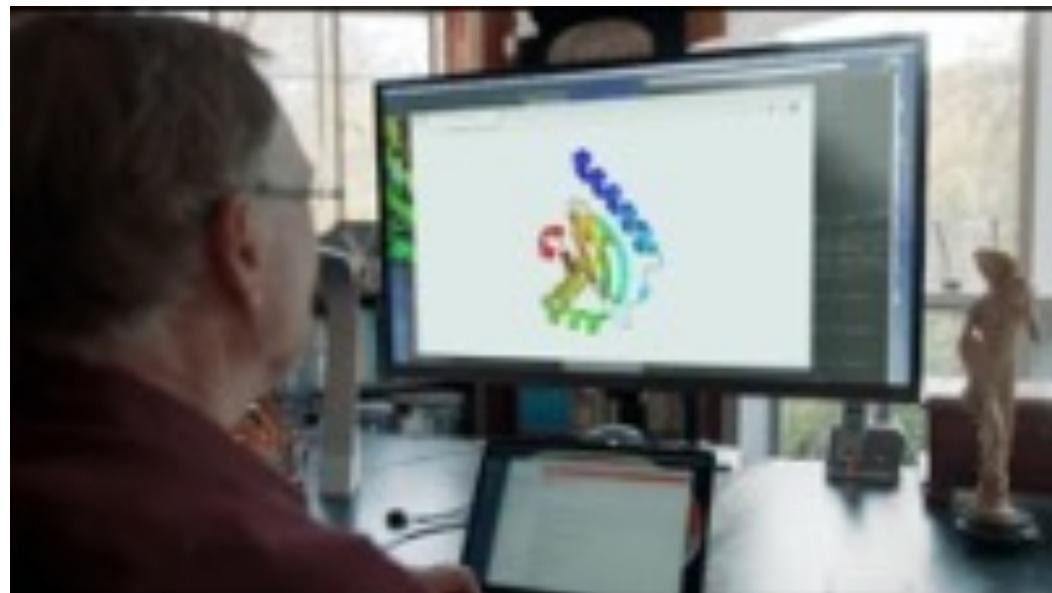
FIGURE: (a) A simplified cartoon of a biological neuron, with a signal flow from inputs at dendrites to outputs at axon terminals (to other neurons), x and y variables denoting a mathematical analogue of a function. Neural network of multiple neurons in middle. (b) A simplified schematic graph of a feedforward ANN, where the information between neurons or nodes (coloured spheres) go only in one direction. In contrast to a Recurrent Neural Network, that allows connections between neurons in the same or previous layers. Each circle denote an artificial neuron. The network's architecture is based on connected neurons in an input layer, a hidden layer or layers, and an output layer. In a typical design, each connection between neurons carries a weight. The weights are varied during the training phase as the network learns how to connect input and output data, before being tested on unseen instances of data. Connections between neurons can also be called edges. Figure, real neurons image from Yung et al. (2013), adapted from original ([CC BY-NC-SA 4.0](#)).

In a general sense, a neural network (NN) can mean a biological network made up of biological neurons; or an Artificial Neural Network (ANN), used for solving problems related to artificial intelligence. The connections of the biological neurons are modeled in ANN as weights between nodes. Being composed of artificial neurons, ANNs are conceptually derived from biological neurons. Where each artificial neuron has inputs and produces outputs which can be sent to multiple other neurons. The inputs can be the feature values of a sample of external data, such as numbers, images, documents, genomic sequences, or 3D structure data. Inputs can also be the outputs of other neurons. The outputs of the final output neuron or neurons of the NN accomplish a task, such as recognizing an object in an image, or predicting 3D protein structures. Where ANNs can be trained via a dataset, they can be used for predictive modelling, adaptive control, and any applications. Self-learning resulting from experience can occur within networks, which can derive conclusions from a complex and seemingly unrelated set of information (Neural

Network 2022).

Hence ANNs, is another mathematical model used for pattern recognition and machine learning. While the structure and function of the human brain is an inspiration to ANN, it is enormously simpler in design and in no way simulates higher brain function. (Mitchell, 2014).

There are very many different architectures of ANNs, yet that share these neural network characteristics. Some main groups of architectures relevant for drug discovery is deep neural networks (DNN), Recurrent Neural Network (RNN), Convolutional Neural Networks (CNN) and Graph Neural Networks. Yet many variations and combinations exist are relevant also for drug discovery, just as example: Transformers, Graph attention networks and Graph Convolutional Networks. (Patel et al., 2020). In this course we refer to ANNs as neural networks (NN) or deep learning (DL).



MOVIE: [Link to Interview](#), from the producers of AlphaFold at Google DeepMind. (DeepMind, 2020a)

AlphaFold is an initiative, used for predicting structure with machine-learning. AlphaFold is a machine-learning algorithm for protein structure prediction that has now been used to obtain hundreds of thousands of protein models.

Imagine a website where you could download a reliable three-dimensional model of your protein of interest. In recent years, this was just a dream. Now such structure prediction has become reality, at least for many monomeric proteins. (Cramer, 2021).

AlphaFold are utilising methods of deep learning which draws upon inspiration from previous insights and fields of biology, physics and machine learning. AlphaFold is trained on the structural data of the PDB and other data-bases.

AlphaFold incorporates empirical knowledge about protein structure

into a deep-learning algorithm. The algorithm also makes use of information from evolutionary conservation in the form of multiple-sequence alignment. The resulting protein models being often as accurate as experimentally determined structures. (AlphaFold, 2021a)

AlphaFold and similar initiatives have reached a revolutionary advance for protein structure predictions, however the implications for drug discovery are, for now, more incremental (Mullard 2021). As time goes by and people continuing to build upon this kind of predicting ability, can however have a significant impact on the drug discovery field.

THE ALPHAFOLD NEURAL NETWORK METHOD

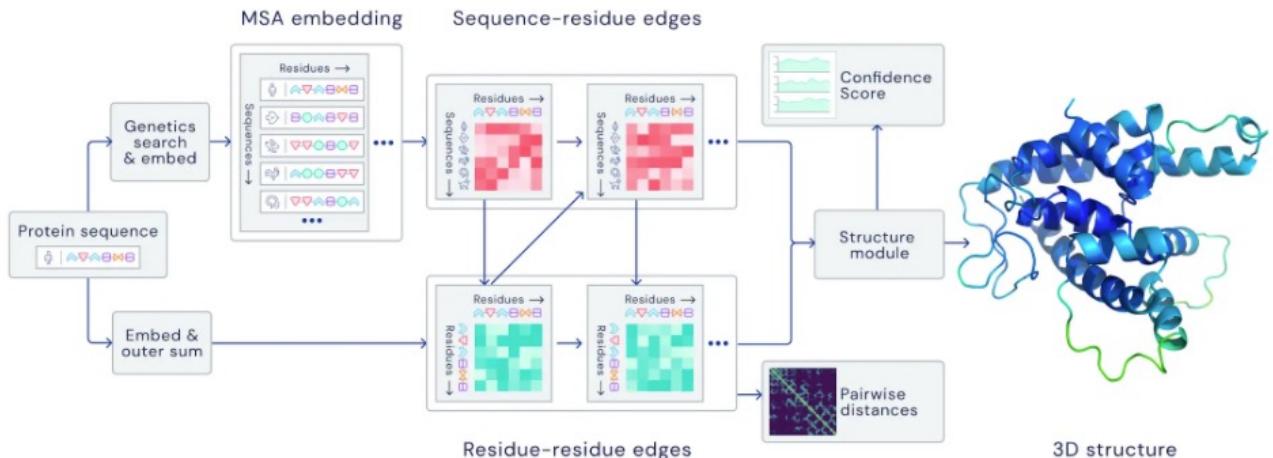


FIGURE: An overview of the neural network main model architecture. Operating over species related protein amino-acid sequences as well as amino acid residue pairs; passing information between both representations iteratively to generate a structure. The schematic and figure text are from AlphaFold Team (2021a).

The AlphaFold neural network consists of two main stages. Stage 1 takes as input the amino acid sequence and a multiple sequence alignment (MSA). The goal is to learn a rich "pairwise representation" being informative about which residue pairs are close in 3D space.

Stage 2 uses this "pair-wise representation" to directly produce atomic coordinates by treating each residue as a separate object, predicting the rotation and translation necessary to place each residue, and ultimately assembling a structured chain. The design of the network draws on our intuitions about protein physics and geometry, for example, in the form of the updates applied and in the choice of loss.

Interestingly, it is produced a 3D

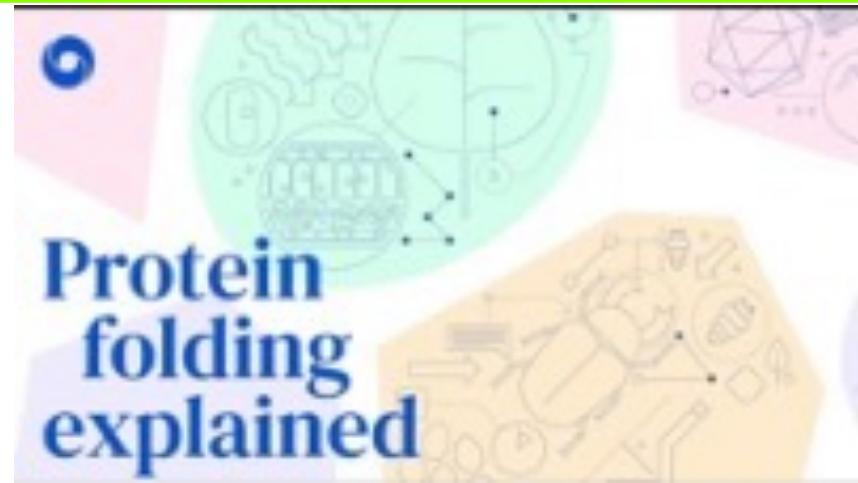
Residue: In chemistry or molecular biology, a residue refers to a single unit that makes up a polymer, such as an amino acid in a polypeptide or protein. For example a protein consisting of 400

structure based on the representation at intermediate layers of the network. Typically a hypothesis emerges after the first few layers followed by a lengthy process of refinement, although some targets require the full depth of the network to arrive at a good prediction.

Text and further information can be found from AlphaFold Team (2021b)

In this course we will not go into the intricate detail of the AlphaFold architecture but it can be understood from other resources, e.g. Jumper (2021), in how AlphaFold utilized NN and types of algorithms used. AlphaFold is continuously under development, on this page AlphaFold 2 is described.

amino acid residues. The residues are numbered from the N-terminal to the C-terminal in the protein.



MOVIE: [Link to movie](#), Protein folding explained (DeepMind, 2020b)

How the information from a gene from the genome is translated to its corresponding protein in the proteome: is a process entailing much other complex macro-molecules and chemical machinery in the cell. The gene or DNA information is mapped by the cell to messenger RNA (mRNA) information, mRNA which is then translated to amino-acid chains that the cell then folds into proteins.

The protein folding problem is how one can by algorithms and computation understand how one can translate the gene or the aminoacids of a protein to its 3D structure. This has been a grand challenge since around the 1970s. An estimate of the time required to fold a typical protein on its own from its amino-acid sequence is that it would require more than billions of years. Yet the cell can do it in very short time!

AlphaFold can map entire genomes to the proteome, so now basically a crude solution of the protein folding problem has been solved by machine-learning, and much work and inspiration is under progress to

refine the prediction quality. (AlphaFold Team, 2021a)

AlphaFold, employs Neural Networks to predict 3D structure of whole proteomes using genomes information of organisms, for instance their model can predict the 3D structure of Myoglobin (the iron and oxygen binding protein in our muscles) from experimental 3D experimental structures amino-acid sequences compared between homolog proteins in different species.

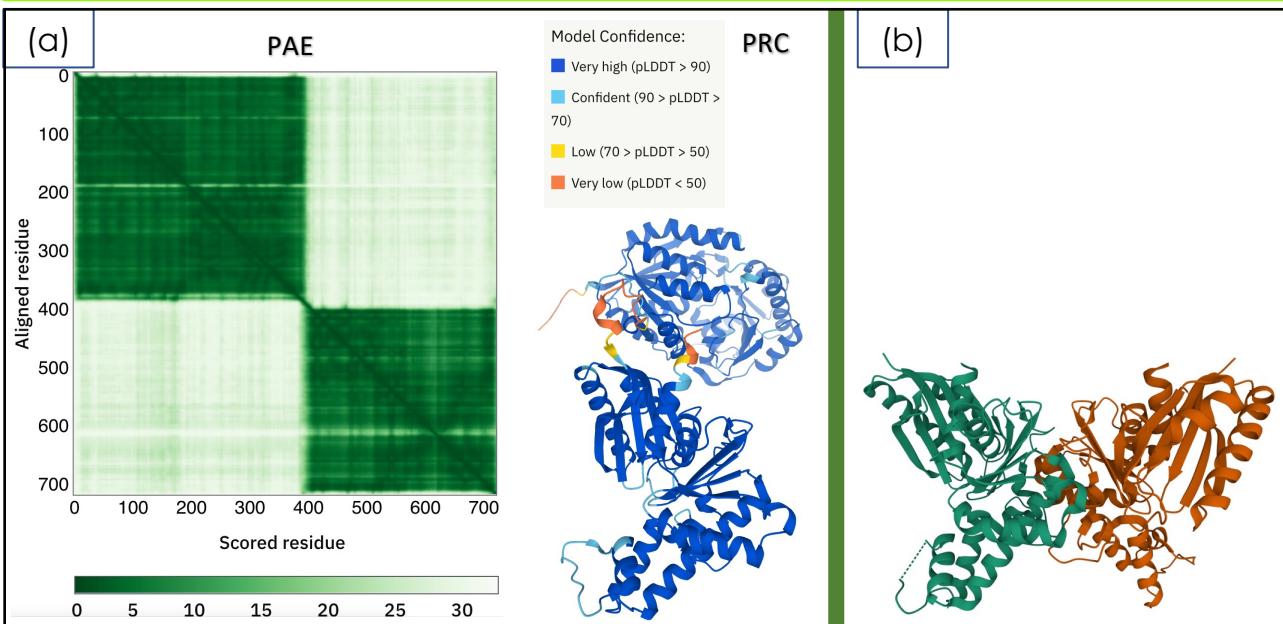


Figure: The structure " Bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase". (a) The Confidence measures of an AlphaFold prediction PAE revealing improper interdomain fold, whereas PRC indicates high confidence on most predicted residue conformations, excepting a few. (b) An X-ray crystal structure in PDB entry 2YI1, showing proper interdomain fold (green and orange domains. AlphaFold prediction made 6 April 2022 at <https://www.alphafold.ebi.ac.uk/entry/Q9Y223> .

Two confidence measures are used for the structure obtained in the AlphaFold databank for anyone to have a measure of their accuracy. To understand the following error metrics, one can overlap the predicted structure (the folded chain of amino-acid residues) against a true known structure, and comparing the same residues in between.

Per residue confidence (PRC): The first is pLDDT (predicted IDDT-Ca), a per-residue measure of local confidence on a scale from 0 - 100. Where pLDDT can vary vastly along a chain, enabling the model to reveal high confidence on structured domains but low confidence on the linkers between domains. There are indications that some regions with low pLDDT may be unstructured in isolation; either intrinsically disordered or structured only in the context of a larger

complex. Regions with pLDDT < 50 should not be interpreted except as a possible disorder prediction.

Predicted Aligned Error (PAE), a second metric reporting AlphaFold's expected position error at residue x, in Ångströms, when the predicted and true structures are aligned on residue y. This is useful for assessing confidence in global features, especially domain packing. The green parts of matrix indicate good congruence, whereas the pale-green regions indicate lower congruence.

At the end of any entry at AlphaFold (e.g. entry on Figure select Predicted aligned error tutorial for an further explanation of PAE concept).

(Text from AlphaFold Team, 2021b)

ANOTHER PREDICTION EXAMPLE OF ADRENERGIC GPCR

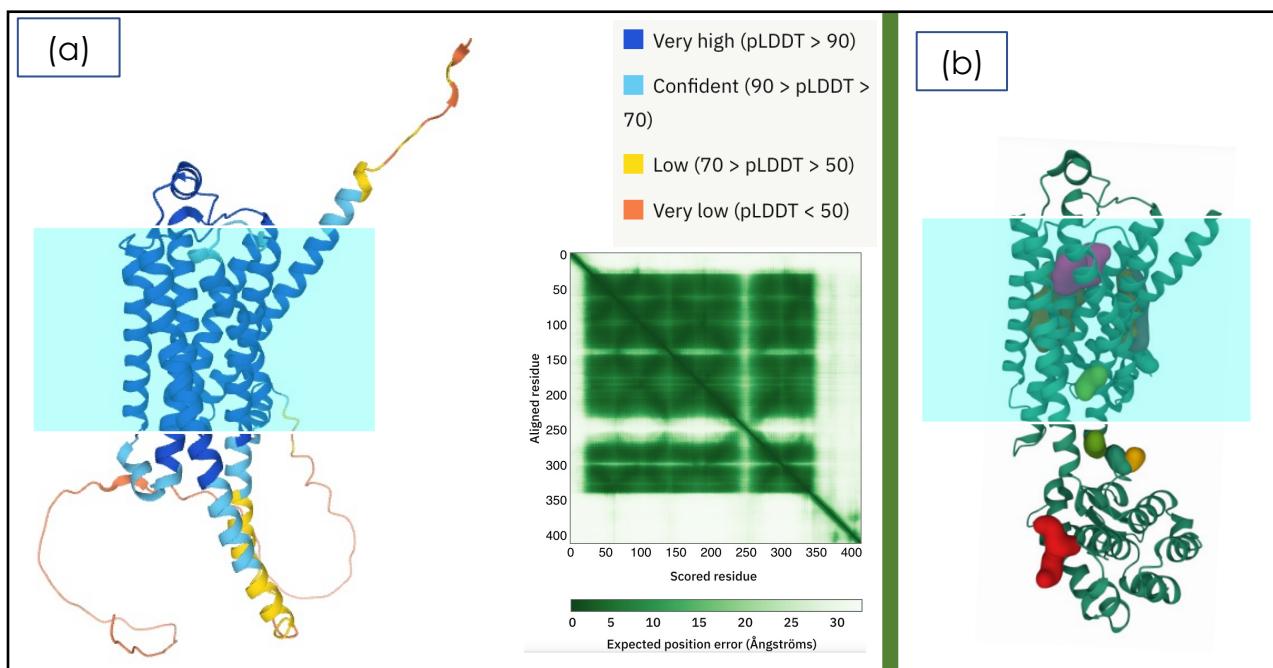


FIGURE: Experimental structure and prediction from AlphaFold of receptor, B2-adrenergic G protein-coupled receptor. This is otherwise an membrane-embedded protein in biological cells, structure of which here is being estimated with experimental methods and AlphaFold structure prediction. The light blue shade area, signifies the cells lipid membrane that this receptor embed in.

(a) AlphaFold prediction with confidence measures PCE and PAE.

(a) The receptor structure entry PDB 2RH1, an structure obtained by X-ray crystallography. Many ligands are seen bound, ligand carazolol in magenta at top of molecule. 3D View of protein from link <https://www.rcsb.org/3d-view/2RH1>

The same orientation of protein in (a) and (b) are shown.

AlphaFold prediction were obtained in January 2022

<https://www.alphafold.ebi.ac.uk/entry/P07550>

In FIGURE(b) is the PDB structure entry 2RH1, where a ligand in Magenta(at top), the small drug Carazolol which is one of a large class of beta-blockers that bind to the adrenergic receptor, making it useful to treat heart disease.

In FIGURE(a) is the AlphaFold prediction of the same receptor protein. For quantitative comparison, we made the assumption here that the true structure used by AlphaFold are very similar to that of PDB 2RH1.

The blue parts of protein (same value as in Model Confidence) have very high confidence, we can compare visually that they indeed look the same. Whereas the parts with very low to confident pLDDT have less similarity.

The PAE matrix (at bottom left) of this

protein reflects these domain, and inter-domain similarities and differences as described here, atleast seen here for the N-, and C- terminal has low per residue confidence and high predicted aligned error.

AlphaFold is trained upon structures in PDB databases however, as obtained in this date, it does not perfectly predict the receptor structure, where here the X-ray crystal structure is a better representation of this, in biological cells, membrane-bound protein. In structural biology one should always compare with literature at hand, for example the report explaining the PDB structure (Cherezov, 2007).

Session 5.1

6 UQ



Session 5.2

6 UQ





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

METHOD: ALPHAFOLD, NEURAL NETWORKS, 3D VIEWING

(Narrative here. The exercise itself is on Studium)

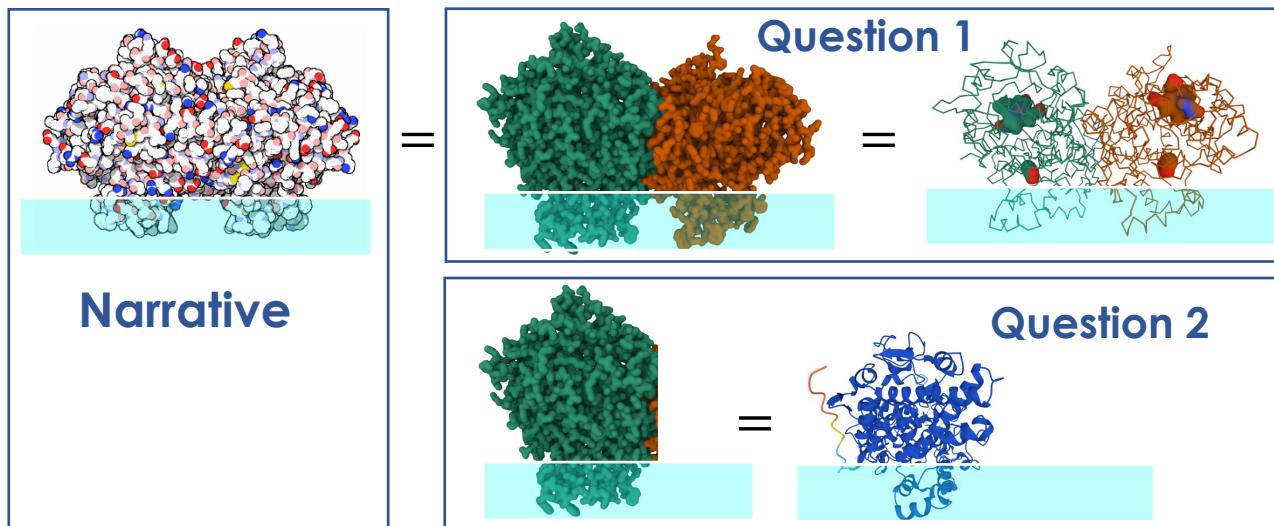


FIGURE: Abstract for the exercise, when going through the steps of exercise at Studium this abstract will be clear.

- **NARRATIVE INFORMATION**
First we present some overall knowledge of this drug-target macromolecule protein cyclooxygenase, the drug effect and action of ligand aspirin.
- **NARRATIVE Question 1**
Here we will investigate the 3D structure using the inherent 3D viewer at the protein data bank. We will investigate the ligand binding inhibition site, and retrieve the closest amino-acids of binding site of salicyc acid (a similar molecule to aspirin with basically the same binding mechanism).

- **NARRATIVE Question 2**
Here we will use the machine-learned structure prediction

databank AlphaFoldDB to predict the 3D structure and look closer at the residues of ligand binding, and also investigate the confidence measures and quality. How well does the alphafold prediction compare to the real known structure.

Please follow the instructions at Exercise 5 at Studium to complete the TASKs with answers.

EXERCISE 5 - NARRATIVE INFORMATION

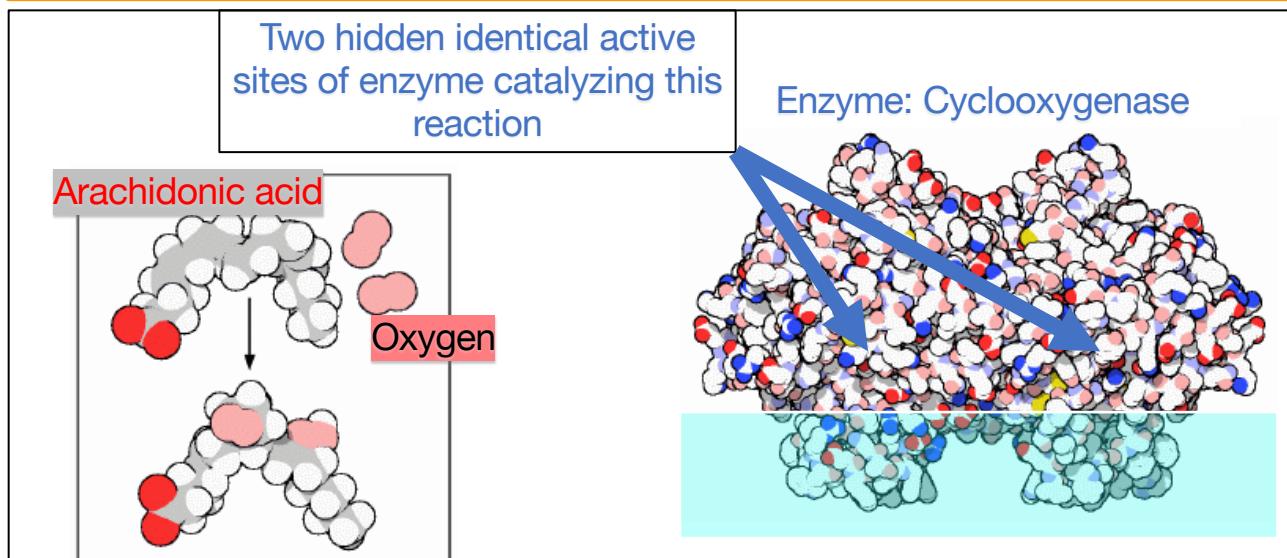


FIGURE: Part explanation of action of aspirin, showing the enzyme which aspirin inhibits. The light blue shade area, signifies the cells bilayer phospholipid membrane that Cyclooxygenase embed in. Cyclooxygenase RCSB PDB structure represented is entry 1PTH, from article Loll et al. (1995).

Enzyme protein Cyclooxygenase performs the first step in the creation of prostaglandins from a common fatty acid, as shown in the box. It adds two oxygen molecules to arachidonic acid. Aspirin inhibits this process. Aspirin blocks the production of prostaglandins, which is a form of lipids that function as key hormones to carry local messages. Unlike most hormones, which are produced in specialized glands and then delivered throughout the body by the blood, prostaglandins are created by cells and then act only in the surrounding area before they are broken down. Prostaglandins, control many of these neighborhood processes, including the constriction of muscle cells around blood vessels, aggregation of platelets during blood clotting, and constriction of the uterus during labor. Prostaglandins also

deliver and strengthen pain signals and induce inflammation. These many different processes are all controlled by different prostaglandins, but all created from a common precursor molecule.

COX-1 and COX-2

Our bodies actually build two different cyclooxygenases (termed COX-1 and COX-2) for different purposes. COX-1 is built in many different cells to create prostaglandins used for basic housekeeping messages throughout the body. The second enzyme is built only in special cells and is used for signaling pain and inflammation.

Unfortunately, aspirin attacks both. Since COX-1 is targeted, aspirin can lead to unpleasant complications, such as stomach bleeding.

Fortunately, specific compounds that block just COX-2, leaving COX-1 to perform its essential work, are now becoming available. These new drugs are selective pain-killers and fever reducers, without the unpleasant side-effects.

(Text from article Goodsell (2001), the whole blog can be read if further interest, text and image reused [CC-BY-4.0 license](#)).

Glossary

Arachidonic acid: an polyunsaturated fatty acid present in the phospholipids of the lipid membranes of the body's cells, and is abundant in the brain, muscles and liver.

Phospholipids: a key component of all cell membranes. They can form lipid bilayers because of having both hydrophilic and hydrophobic components.

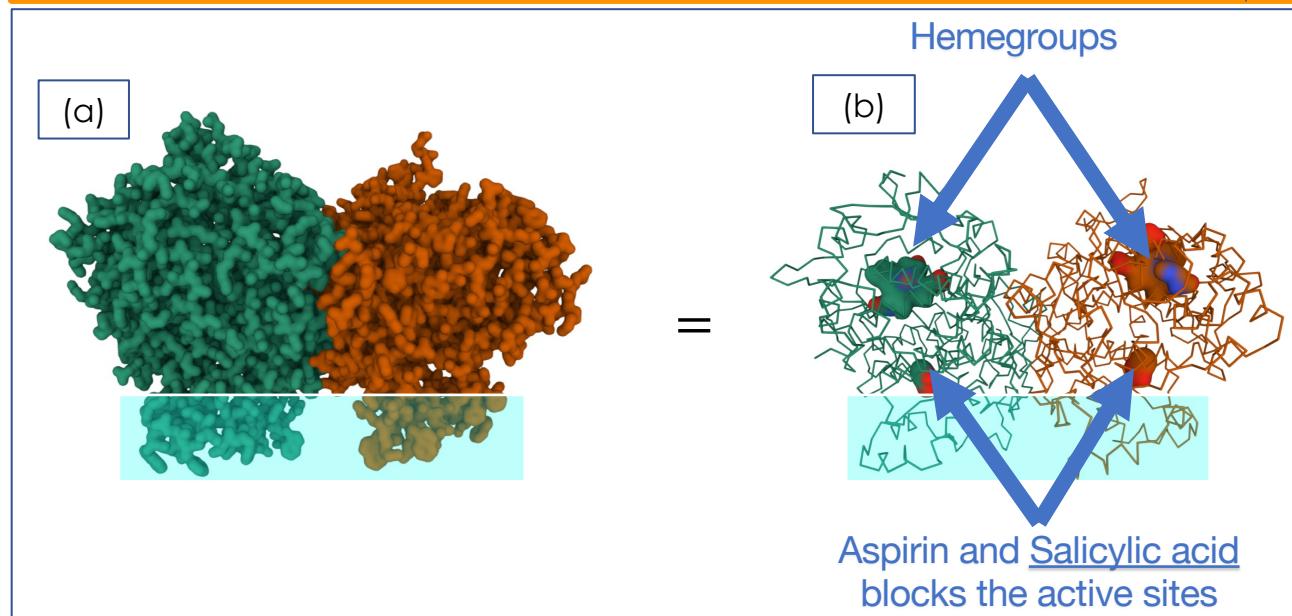


FIGURE: Protein Data Bank 3D View structure rendering of Cyclooxygenase (PDB entry 1PTH), of ovine (sheep) origin. In the 3D Viewer the Carbohydrates attached to protein are not shown, neither Non-standard molecules and Water from the crystal structure, these can be set to “not view” in graphical user interface of the 3D viewer.

(a) All-atom surface view of the protein where the green and orange monomer are respectively 604 amino-acids, The active enzyme is a dimer. a bit less since the N-,C-terminus were not included in reported PDB structure.

(b)

The same structure shown in an ribbon-representation of the monomer chains. The ligand salicylic acid and heme groups are shown in atom sphere (Gaussian sphere rendering).

Note the heme-groups, moreover where salicylic acid blocks the active sites. The heme groups are included as ligands which participate in the active site reaction. In each COX 604 amino-acid monomer, the inhibition site where salicylic acid binds is in the middle of three alpha-helices (coiled/folded amino-acid chains). When salicylic acid (and aspirin) binds it permanently inactivates the enzyme activity by forming a strong bond to SER 530, in a reaction called acetylation (Loll, 1995).

Moreover, this protein is attached to the membrane by the side of entrance to its two active sites, these crystal structures we see here is not incorporating the membrane; however is assumed that the structure is similar as in its biological surrounding and conditions in the cell.

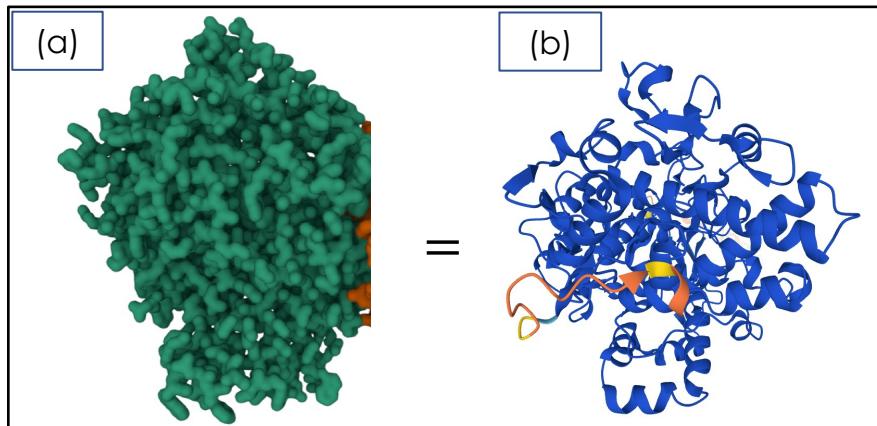


FIGURE: Structure rendering of a single 604 amino-acid monomer of cyclooxygenase.

(a) PDB 3D Viewer all-atom surface rendering (PDB entry 1PTH, of ovine or sheep origin), showing only one monomer of PDB 1PTH.

(b) AlphaFold structure prediction of a single 604 amino-acid monomer of Cyclooxygenase-1 (Human), to the right.

One can access the AlphaFold prediction of cyclooxygenase at their website <https://alphafold.ebi.ac.uk/>

The confidence measures can be retrieved for any residue at AlphaFold, moreover the predicted aligned error. Atleast currently the AlphaFold DataBase prediction predicts structures without bound ligands, hence no salicylic acid is present in the prediction only protein structure.

References (Harvard citation style)

- Anderson, A.C., 2003. The process of structure-based drug design. *Chemistry & biology*, 10(9), pp.787-797.
- Arachidonic acid , 2022. Wikipedia. Available at: https://en.wikipedia.org/wiki/Arachidonic_acid (Accessed: 28 Mars 2022)
- Artificial Neural Network, 2022. Wikipedia. Available at: https://en.wikipedia.org/wiki/Artificial_neural_network (Accessed: 24 Mars 2022)
- AlphaFold Team, 2020. Using AI for scientific discovery. DeepMind. Available at: <https://www.deepmind.com/blog/alphafold-using-ai-for-scientific-discovery-2020> (Accessed: 28 Mars 2022)
- AlphaFold Team, 2021a. AlphaFold: A Solution to a 50-year-old Grand Challenge in Biology. DeepMind. Available at: <https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology> (Accessed: 23 Mars 2022)
- AlphaFold Team, 2021b. Enabling high-accuracy protein structure prediction at the proteome scale. Available at: <https://deepmind.com/research/publications/2021/enabling-high-accuracy-protein-structure-prediction-at-the-proteome-scale> (Accessed: 23 Mars 2022)
- Berg, J.M., 2005. Tymoczko, JL and Stryer, L. *Biochemistry*.
- Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. and Blaschke, T., 2018. The rise of deep learning in drug discovery. *Drug discovery today*, 23(6), pp.1241-1250. <https://doi.org/10.1016/j.drudis.2018.01.039>
- Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S.G., Thian, F.S., Kobilka, T.S., Choi, H.J., Kuhn, P., Weis, W.I., Kobilka, B.K. and Stevens, R.C., 2007. High Resolution Crystal Structure of an Engineered Human β 2-Adrenergic G protein-Coupled Receptor. *Science (New York, NY)*, 318(5854), p.1258.
- Cramer, P. AlphaFold2 and the future of structural biology. *Nat Struct Mol Biol* **28**, 704–705 (2021). <https://doi.org/10.1038/s41594-021-00650-1>
- DeepMind, 2020a. AlphaFold the making of a scientific breakthrough. 30 November. Available at: <https://youtu.be/gg7WjuFs8F4> (Accessed: 23 Mars 2022).
- DeepMind, 2020b. Protein folding explained. 30 November. Available at: <https://youtu.be/gg7WjuFs8F4> (Accessed: 23 Mars 2022).
- Farhad, M.. 2019. What Are Hidden Layers?, Medium, 20 May. Available at: <https://medium.com/fintechexplained/what-are-hidden-layers-4f54f7328263> (Accessed: 24 Mars 2022).
- Goodsell, D.S., 2001. Cyclooxygenase, RCSB PDB Molecule of the Month by [David S. Goodsell](#). Available at: <https://pdb101.rcsb.org/motm/17> (Accessed: 28 Mars 2022).
- Goodsell, D.S., Voigt, M., 2008, How do Drugs Work, RCSB PDB-101. Available at: <https://pdb101.rcsb.org/learn/flyers-posters-and-other-resources/flyer/how-do-drugs-work> (Accessed: 23 Mars 2022).
- Goodsell, D.S., 2016. Insulin Action 2016, Molecular Landscapes by David S. Goodsell. Available at: <https://pdb101.rcsb.org/sci-art/goodsell-gallery/insulin-action> (Accessed: 22 Mars 2022).
- Goodsell, D.S., 2021. Escherichia coli Bacterium, 2021, Molecular Landscapes by David S. Goodsell. Available at: <https://pdb101.rcsb.org/sci-art/goodsell-gallery/escherichia-coli-bacterium> (Accessed: 22 Mars 2022).
- Goodsell, D.S., n.d.. About Molecule of the Month, RCSB PDB Molecule of the Month by [David S. Goodsell](#). Available at: <https://pdb101.rcsb.org/motm/motm-about> (Accessed: 22 Mars 2022).
- Institute for Glycomics, 2021. What is Glycomics. 7 April. Available at: <https://www.youtube.com/watch?v=NqEgrAYN2Bc> (Accessed: 23 Mars 2022).
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), pp.583-589.
- Loll, P.J., Picot, D. and Garavito, R.M., 1995. The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase. *Nature structural biology*, 2(8), pp.637-643.

References (Harvard citation style)

Methods for Determining Atomic Structures, no date, RCSB PDB-101. Available at: <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure> (Accessed: 23 Mars 2022).

Mitchell, J.B., 2014. Machine learning methods in chemoinformatics. Wiley Interdisciplinary Reviews: Computational Molecular Science, 4(5), pp.468-481.

Mullard, A., 2021. What does AlphaFold mean for drug discovery?. *Nature reviews. Drug discovery*.

Omics, 2022. Wikipedia. Available at: <https://en.wikipedia.org/wiki/Omics> (Accessed: 23 Mars 2022)

Patel, L., Shukla, T., Huang, X., Ussery, D.W. and Wang, S., 2020. Machine learning methods in drug discovery. *Molecules*, 25(22), p.5277.

Phospholipid, 2022. Wikipedia. Available at: <https://en.wikipedia.org/wiki/Phospholipid> (Accessed: 28 Mars 2022)

XVIVO Scientific Animation, 2011. The Inner Life of the Cell. 11 July. Available at: <https://youtu.be/wJyUtbn0O5Y> (Accessed: 22 Mars 2022)

Yung, A., Welsh, C., Minderer M., 2013. Talking Back to the Brain: Using Light to Uncover the Language of Neurons. Available at: <https://sitn.hms.harvard.edu/seminars/2013/talking-back-to-the-brain-using-light-to-uncover-the-language-of-neurons/> (Accessed: 25 Mars 2022)

Zardecki, C., Dutta, S., Goodsell, D.S., Lowe, R., Voigt, M. and Burley, S.K., 2022. PDB-101: Educational resources supporting molecular explorations through biology and medicine. *Protein Science*, 31(1), pp.129-140.