

Looking for a good drug

A good drug molecule will not be created unless it is known which proteins it affects in our body. That is why, in drug design, it is important to utilise massive databases with all the discovered protein structures and protein families as well as knowledge about how proteins function in our cells.



The majority of the drugs in use are designed so that their target molecules are the body's biomolecules, i.e. proteins.

Most drugs take effect in the body by binding themselves to these targets like receptors of signal molecules. Receptors are natural targets for example for signal molecules such as neurotransmitters and hormones. They are specialised triggers of the cell associated with cellular signalling mechanisms.

The idea for drug design is to build small synthetic molecules that selectively affect the desired proteins. Most of the target proteins of drugs belong to only ten protein families, and up to half belong to only three families. Small molecules are able to absorb well into the bloodstream, allowing the drug to take effect. Depending on the location of the protein, the drug molecule has to penetrate the cell or transmit a signal outside the cell that affects the processes within the cell. The aim is to de-

sign the molecules, for example, in such a way that they slow down or accelerate the functioning of a particular protein.

In the past, little was known about in which part of the cell the drug takes effect. In 1980, 150 of these target areas of effect were known. However, that number has

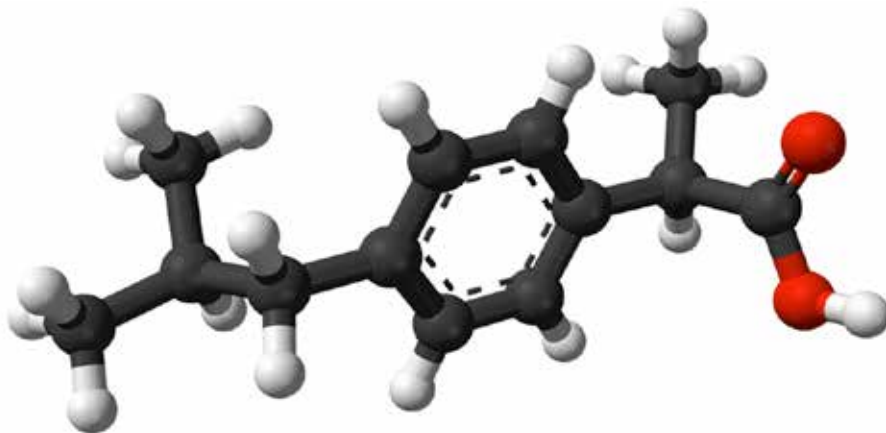
**“Approximately
2,500 drug molecules
are available
for medicine.”**

grown enormously with the determination of the genomes of organisms since, currently, already more than 5,000 possible target areas of effect are known. Approximately 2,500 drug molecules are available for medicine. The function of the human genome is being investigated more and more closely and, in the next few years, the number of known possible target are-

as of effect for drug ingredients may rise to 10,000. According to current estimates, our body has 2,000–3,000 proteins that are possible target proteins for a drug i. Existing drugs have been shown to work through only about 450 drug targets on a limited number of diseases. Thus, drug designers have two major goals – to build new safe molecules that can be used to safely affect known targets and, on the other hand, to study the use of known, safe drugs for new illnesses for which there is currently no drug approved by the authorities.

The goal of researchers is, among other things, to understand which structural and chemical characteristics of a drug molecule play a key role as they modify the function of proteins at the cellular level.

An effective drug can be developed once a three-dimensional structure of the target protein, which allows interaction with the drug molecule, is found. Chemical counterparts that recognise the amino



Ibuprofen, used in, for example, many painkillers, inhibits the function of the cyclooxygenase enzyme, reducing the production of chemicals and hormones called prostaglandins which are needed in the communication of pain receptors. This reduces the sensation of pain.

acids at the protein's binding site are built into the drug molecule. When this kind of molecule encounters the target protein in the body, it automatically finds its way to the binding site of the protein because attaching itself there is energetically advantageous for it.

The binding of a well-designed drug molecule to the target protein could be compared with putting on a wool glove. It fits firmly on the hand with precisely five fingers: it would be very uncomfortable for one with six or seven fingers. In addition, a left-hand glove fits poorly on the right hand.

The shape of proteins tells more about the function of the molecule than the amino acid sequence. Proteins with the same shape can function similarly biochemically even if their amino acid sequences differ from each other by more than 80%.

Once the structure of one member of a protein family has been determined, the structure of other proteins belonging to the same family can be predicted by modelling. Modelling which is carried out using a computer speeds up research because hundreds of times more protein amino acid sequences are known than protein structures that have already been determined by testing. It can roughly be said that the task of genomics is to determine the sequence of nucleotides. This sequence is translated into an amino acid polymer in the cell, but it starts to function only after the protein folds up into its

three-dimensional shape. This function is investigated through proteomics. Thus, cooperation between experts in genomics, proteomics and drug molecule modelling supports one another.

Protein structures and locations in databases

Even though there is much information, the development of new drugs is quite challenging. Only 5% of drug ingredient candidates progress through laboratory testing even to treatment tests on animals. Of those, only a few per cent will ultimately be suitable as medications. It has been estimated that up to 75% of the price of drugs is due to the costs of failed pharmaceutical development projects.

One major challenge is minimising side effects. With the development of genomics, drug molecules have been found to have an individual effect. Historically, drugs have been developed assuming that people are similar in terms of their biochemistry but, in reality, we are unique at the cellular level in the same way as people are slightly different physically. When small drug molecules are used to try and influence the situation of a diseased body in a healing way, these individual differences at the molecular level may affect the performance of the drug.

By collecting and storing human biological data, it will be possible in the future to target drug molecules for treatment

“The shape of proteins tells more about the function of the molecule than the amino acid sequence.”

purposes that do exactly what they should in that exact situation, and tailored to the person who needs the medication. This is called personalised medicine.

A particular gene produces a specific protein affected by the drugs. When the DNA base sequence of a person's genome is known, it is also possible to deduce the basic structure of the corresponding protein in that person. Like DNA, a protein is also a string consisting of successive building blocks, and a specific block of a gene always corresponds to a specific block of a protein.

One person may have – inherited or caused by the environment – a change in one DNA nucleotide that is reflected in a protein through this chain. That change may be just where the protein should receive signals from elsewhere in the body or interact with a drug molecule. By storing protein structures and sharing them to be used by researchers, this phenomenon can be controlled and understood. The shapes of the drug molecule and the protein molecule can be matched to each other so that the drug is adapted to the situation, allowing the drug to adhere and take effect as effectively as possible. Many cancer treatments are based on this. The genome of a tumour changes over time. Tumours at different stages can be affected through drugs, but the shape of the drug molecules must take into account the changes in the shape of growth-stimulating proteins.

That is why especially proteins whose three-dimensional structure can be determined by tests or predicted through modelling are studied in drug design. The adherence of the drug molecule can be studied using modern computer modelling software in which the three-dimensional protein and drug models are matched to each other. This also enables the tailoring of the ideal drug shape.

Usually, a drug takes effect by adhering to a defective protein in the body and altering its function. An ideal drug does only this; it does not interfere with healthy



proteins or cause other side effects. Up to the present, we have been happy to find one protein affecting a disease and a drug molecule that is moderately effective against it.

Now, the entire arsenal of proteins and drug molecules can be screened and the best candidates selected. This is due to the advancement of molecular biology, computer computing power and databases. It is now possible to screen the entire protein range of the body.

The Protein Data Bank, i.e. the PDB protein database, includes more than 100,000 protein structures divided into protein families. The members of a protein family are usually similar in terms of their three-dimensional structure, which is why they also function in a similar manner.

The PDB database is maintained by the international consortium Worldwide Protein Data Bank (wwPDB). It is tasked with maintaining individual macromolecular structural data that is freely available to researchers.

The Human Protein Atlas is a Swedish-based programme started in 2003 with the aim to map all the human pro-

teins in cells, tissues and organs. Various omics technologies are used in the mapping, meaning technologies in which all genes or the proteins produced by them are studied simultaneously. These include antibody-based imaging, mass spectrometry-based proteomics, transcriptomics and systems biology. All the data collected is open to researchers.

In January 2015, the Human Protein Atlas published a map showing the locations of 17,000 different proteins in the human body, providing valuable information for drug design. The map included the locations of proteins that were the target proteins of approved drugs. Researchers can view proteins in 32 different tissues, representing all of the most significant tissues and organs in the body.

In December 2017, the Human Protein Atlas released version 18. At that time, the database contained 26,000 antibodies targeting proteins encoded by almost 17,000 genes. It corresponded to 87% of protein-encoding human genes.

Tommi Nyrönen
Ari Turunen

“The PDB database is maintained by the international consortium Worldwide Protein Data Bank.”

MORE INFORMATION:

CSC – IT Center for Science

is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.

<http://www.csc.fi>

<https://research.csc.fi/cloud-computing>

ELIXIR

builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.

<http://www.elixir-finland.org>

<http://www.elixir-europe.org>

ELIXIR FINLAND

Tel. +358 9 457 2821 • e-mail: servicedesk@csc.fi
www.elixir-europe.org/about-us/who-we-are/nodes/finland

www.elixir-finland.org

ELIXIR HEAD OFFICE

EMBL-European Bioinformatics Institute
www.elixir-europe.org