



Bioinformatics against COVID-19

Research Group in Bioinformatics

MSc. Vicente Machaca Arceda

June 12, 2020

Overview



Introduction

- Objectives

- Presentation

- The purpose of Bioinformatics

- What is Bioinformatics?

Bioinformatics against COVID-19

- Protein structure prediction

- Drug discovery

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

Drug discovery

Objectives



- ▶ Understand what is Bioinformatics.

Objectives



- ▶ Understand what is Bioinformatics.
- ▶ Learn areas of research of Bioinformatics related to COVID-19.

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

Drug discovery



- ▶ MSc. Vicente Enrique Machaca Arceda.



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.
- ▶ Full time researcher at La Salle university.



- ▶ MSc. Vicente Enrique Machaca Arceda.
- ▶ Professor at UNSA university.
- ▶ Full time researcher at La Salle university.
- ▶ Leader of research group of Bioinformatics in Arequipa.



Year	Country	Title
2018	Brasil	Fast Car Crash Detection in Video
2016	Chile	Fast Face Detection in Violent Video Scenes
2016	Costa Rica	Real Time Violence Detection in Video with ViF and Horn-Schunck
2016	Costa Rica	Optimization model for face detection in video sequences
2015	Chile	Real Time Violence Detection in Video



Year	Country	Title
2020		DNA sequence similarity analysis using Chaos Game Representation
2020		Machine Learning and Chaos Game Representation for rapid classification of novel pathogens COVID-19 case study
2020	Canada	An analysis of k-mer frequency features with machine learning models for viral subtyping of Poliovirus and HIV-1 genomes
2020	Canada	Forecasting time series with Multiplicative Trend Exponential Smoothing and LSTM: COVID-19 case study
2020	USA	Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

Drug discovery

The purpose of Bioinformatics

Why a person has cancer?

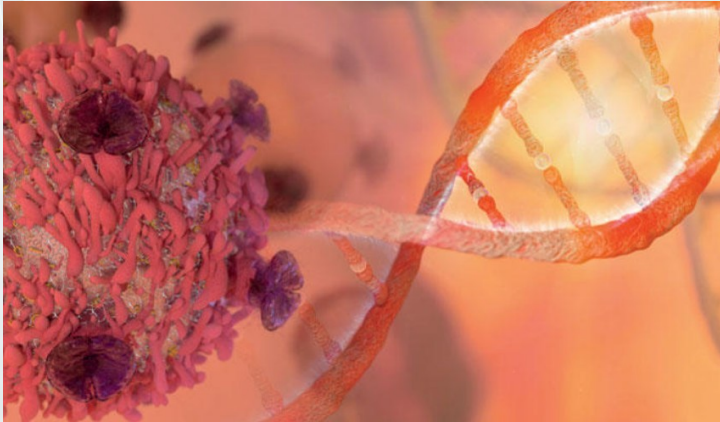


Figure: Why a person has cancer?

The purpose of Bioinformatics

Why some medicines no work in some persons?



Figure: Why some medicines no work in some persons?

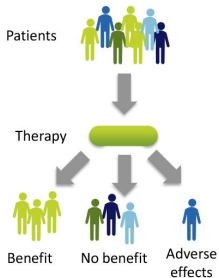
The purpose of Bioinformatics

Treatment Development



Without Personalized Medicine:

Some Benefit, Some Do Not



With Personalized Medicine:

Each Patient Receives the Right Medicine For Them

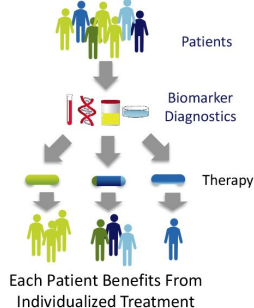


Figure: Personalized Medicine: New Approach to Treatment of Disease

The purpose of Bioinformatics

Protein structure prediction

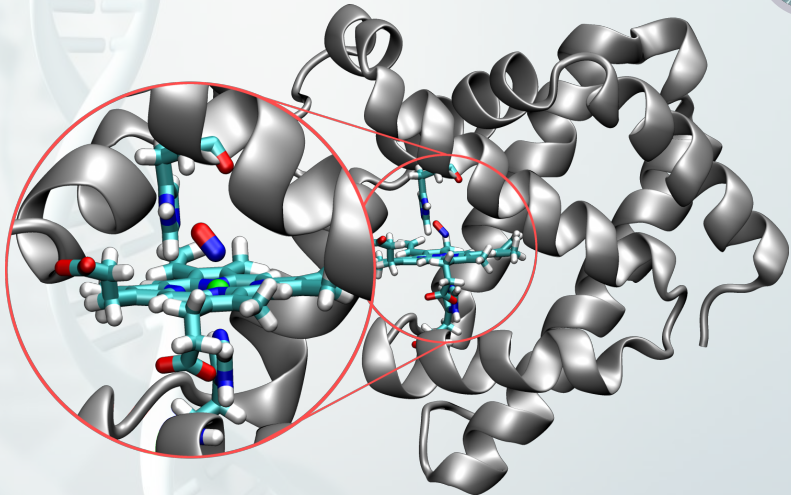


Figure: Computer simulation of protein-ligand.

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

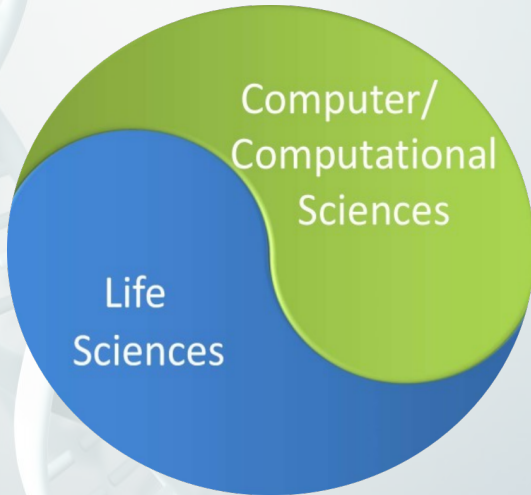
Drug discovery

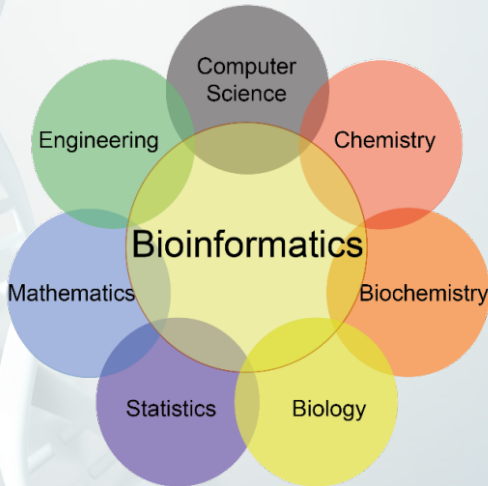
Introduction

What is Bioinformatics?



According to Luscombe et al.: **Bioinformatics** involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins [1].





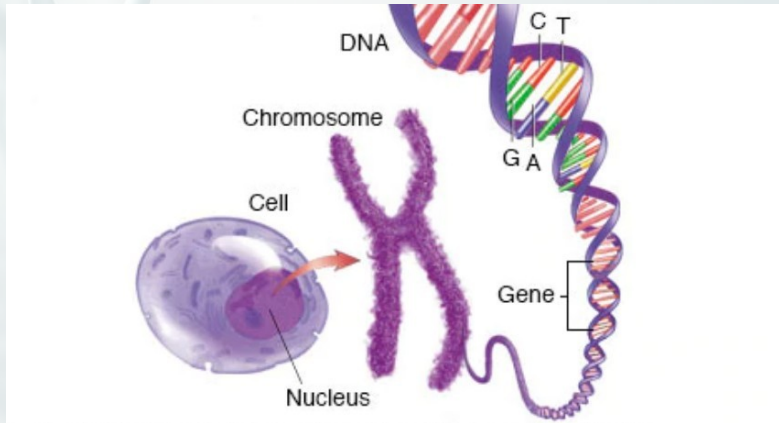


Figure: Where DNA is located [2].

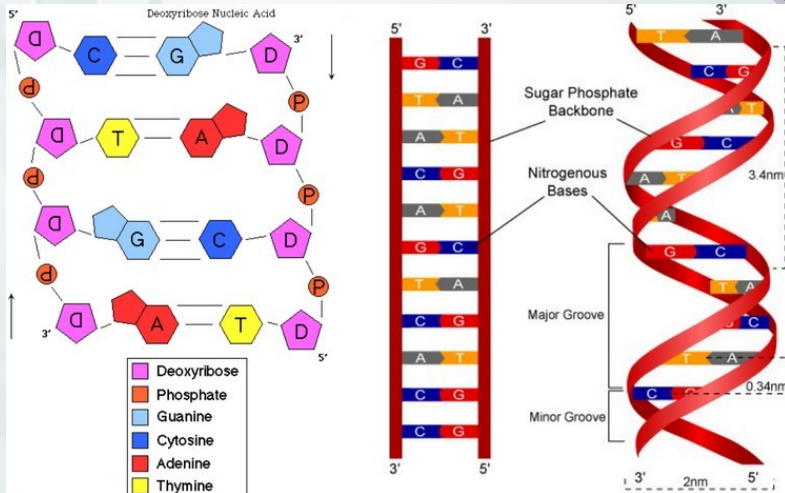


Figure: DNA structure [3].



The human genome is made of ~**3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].



The human genome is made of ~**3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].

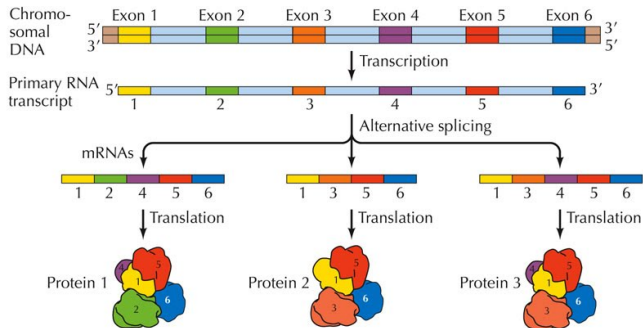
The HIV-1 genome is made of ~**20k bp** of DNA.
Meanwhile, the COVID-19 is made of ~**32k bp** [5].



The human genome is made of ~**3.2 billions bp** of DNA.
~6.4 billions of nucleotides [4].

The HIV-1 genome is made of ~**20k bp** of DNA.
Meanwhile, the COVID-19 is made of ~**32k bp** [5].

There are approximately **19000** to **25000** genes.
No one knows for sure [4].



THE CELL, Fourth Edition, Figure 5.5 © 2006 ASM Press and Sinauer Associates, Inc.

Figure: Alternative splicing [6].

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

Drug discovery

Protein structure prediction

Definition



Definition

The prediction of protein three-dimensional structure from amino acid sequence [7].

Protein structure prediction

Definition



Definition

The prediction of protein three-dimensional structure from amino acid sequence [7].

Methods

- ▶ X-ray crystallography.
- ▶ Nuclear magnetic resonance.
- ▶ Cryo-electron microscopy.

Protein structure prediction

Using computers



There are two approaches to predicting protein structures:

- ▶ Homology modeling.
- ▶ Physical modeling.

Protein structure prediction

Proteins in COVID-19



Figure: Graphical view of COVID-19 structure. Source: [8]

Protein structure prediction

Proteins in COVID-19

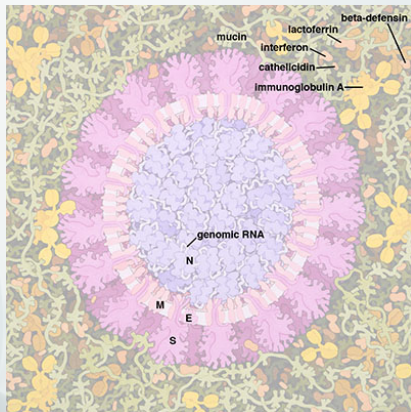


Figure: Membrane S (spike) protein, M (membrane) protein, membrane channel E (envelope) protein and the N (nucleocapsid) protein bound to the genomic RNA. Source: [8]

Protein structure prediction

AlphaFold method

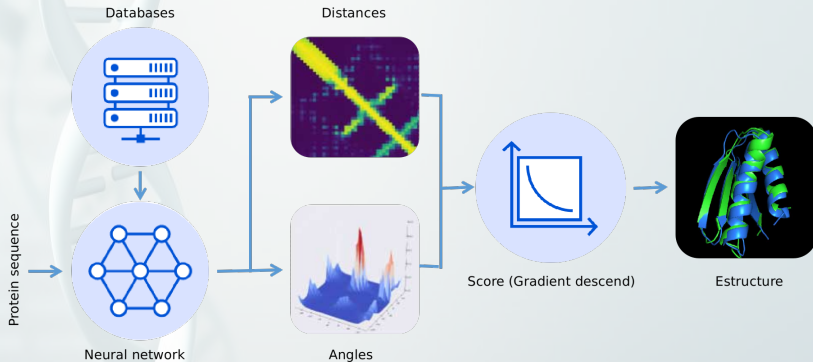


Figure: Protein structure prediction method proposed by AlphaFold. Source: [9]

Protein structure prediction

COVID-19 membrane protein

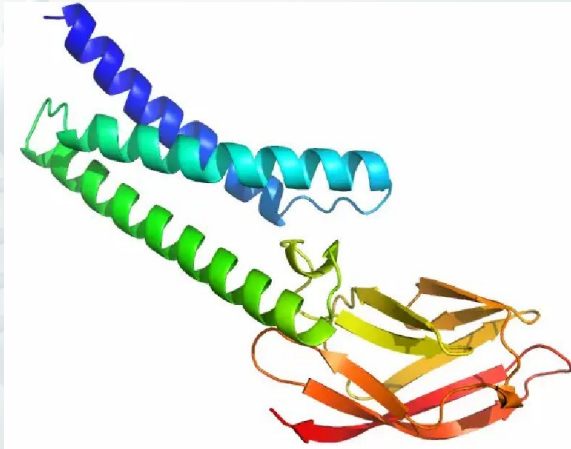


Figure: COVID-19 membrane protein. Source: [9]

Protein structure prediction

Tertiary representations of the S1 and S2 subunits of the spike protein

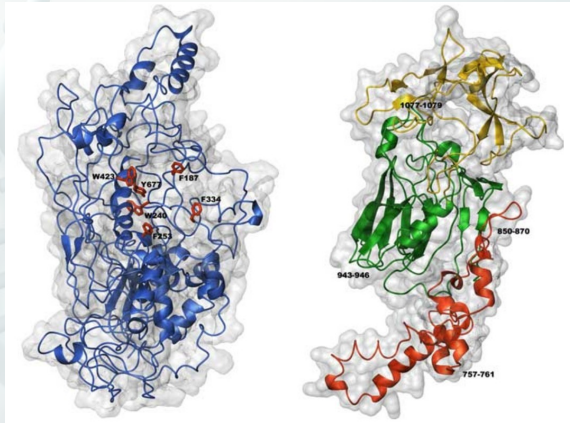


Figure: Tertiary representations of the S1 and S2 subunits of the spike protein using PsiPred. Source: [10]

Table of Contents



Introduction

Objectives

Presentation

The purpose of Bioinformatics

What is Bioinformatics?

Bioinformatics against COVID-19

Protein structure prediction

Drug discovery



Definition

Drug discovery is the process by new candidate medications are discovered [11].

Drug discovery

From a million to one

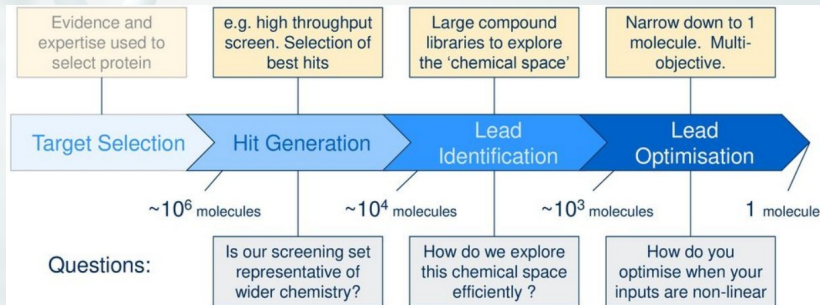


Figure: Process in drug discovery. Source: [12]

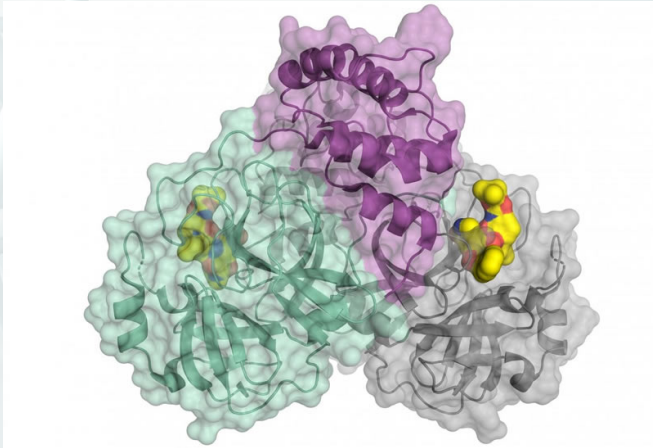


Figure: Schematic representation of the coronavirus protease. Source: [13]

Drug discovery

N3 inhibitor of main protease

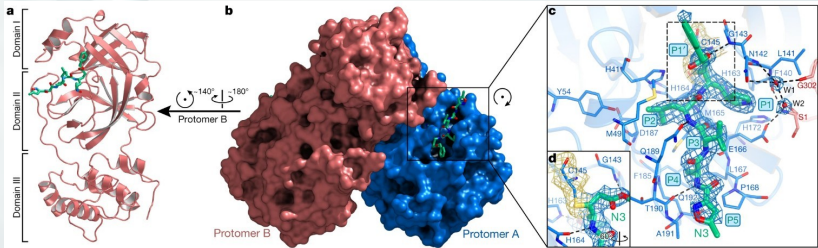
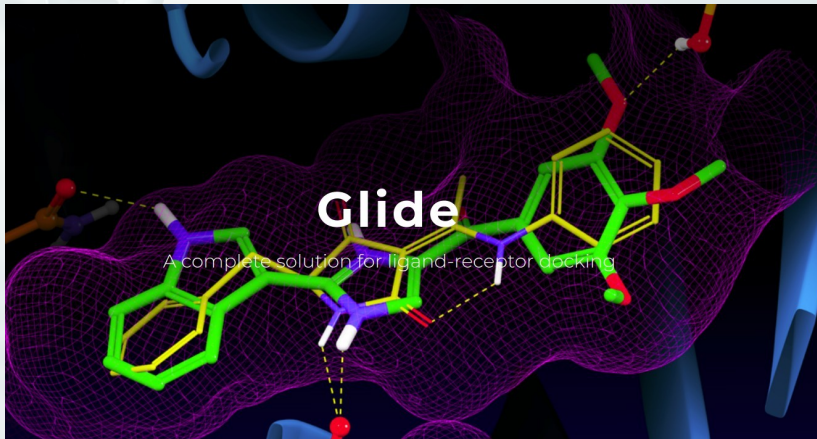


Figure: The crystal structure of SARS-CoV-2 main protease N3 inhibitor.
Source: [14]





Molecular docking

Molecular docking is a computer simulation procedure to predict the conformation of a receptor-ligand complex [15]

Algorithms used:

- ▶ Fast shape matching (take into account the geometric).
- ▶ Simulated Annealing.
- ▶ Genetic algorithms.
- ▶ Tabu search.

Drug discovery

Protease Inhibitors Designed Using Generative Deep Learning Approaches

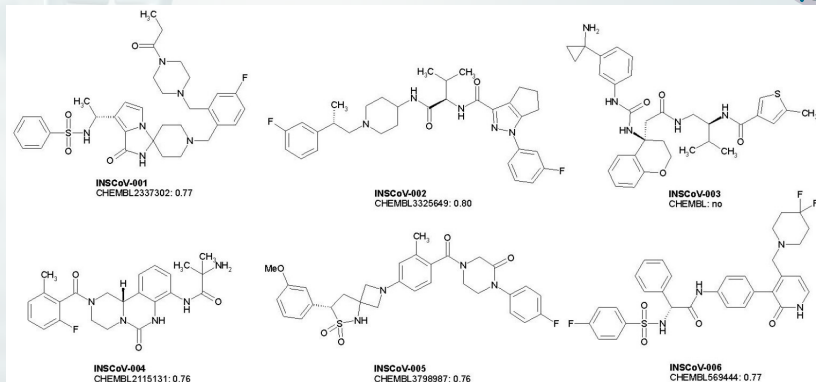


Figure: Representative examples of the structures generated to target the main protease of 2019-nCoV. Novelty was assessed using similarity search in ChEMBL Database. Source: [16]

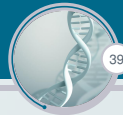


- [1] N. M. Luscombe, D. Greenbaum, and M. Gerstein, “What is bioinformatics? a proposed definition and overview of the field,” *Methods of information in medicine*, vol. 40, no. 04, pp. 346–358, 2001.
- [2] M. Clinics, “How genetic disorders are inherited,” <https://www.mayoclinic.org/tests-procedures/genetic-testing/multimedia/genetic-disorders/sls-20076216?s=2>, 2020, accessed: 2020-03-20.
- [3] NAU, “Dna structure,” http://www2.nau.edu/lrm22/lessons/dna_notes/dna_notes.html, 2020, accessed: 2020-03-20.
- [4] J. M. Archibald, *Genomics: A Very Short Introduction*. Oxford University Press, 2018, vol. 559.

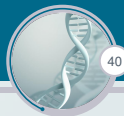
References II



- [5] G. S. Randhawa, M. P. Soltysiak, H. El Roz, C. P. de Souza, K. A. Hill, and L. Kari, “Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study,” *bioRxiv*, 2020.
- [6] G. BIO, “Gen. bio,” <https://sites.google.com/site/bio1040genbio2/home>, 2020, accessed: 2020-03-20.
- [7] B. Kuhlman and P. Bradley, “Advances in protein structure prediction and design,” *Nature Reviews Molecular Cell Biology*, vol. 20, no. 11, pp. 681–697, 2019.
- [8] D. Goodsell, “Coronavirus,” *RCSB Protein Data Bank*, Feb. 2020. [Online]. Available: https://doi.org/10.2210/rcsb_pdb/goodsell-gallery-019
- [9] A. Senior, J. Jumper, D. Hassabis, and P. Kohli, “Alphafold: Using ai for scientific discovery,” 2020.



- [10] O. Spiga, A. Bernini, A. Ciutti, S. Chiellini, N. Menciassi, F. Finetti, V. Causarono, F. Anselmi, F. Prischi, and N. Niccolai, “Molecular modelling of s1 and s2 subunits of sars coronavirus spike glycoprotein,” *Biochemical and biophysical research communications*, vol. 310, no. 1, pp. 78–83, 2003.
- [11] U. Food and Drug, “The drug development process,” 2020.
- [12] C. Nantasenamat, “Computational drug discovery: Machine learning for making sense of big data in drug discovery,” 2020.
- [13] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, “Crystal structure of sars-cov-2 main protease provides a basis for design of improved α -ketoamide inhibitors,” *Science*, vol. 368, no. 6489, pp. 409–412, 2020.



- [14] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng *et al.*, “Structure of m pro from sars-cov-2 and discovery of its inhibitors,” *Nature*, pp. 1–5, 2020.
- [15] R. Dias, J. de Azevedo, and F. Walter, “Molecular docking algorithms,” *Current drug targets*, vol. 9, no. 12, pp. 1040–1047, 2008.
- [16] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, D. Polykovskiy, R. Shayakhmetov, A. Filimonov, P. Orekhov, Y. Yan, O. Popova, Q. Vanhaelen, A. Aliper, and Y. Ivanenkov, “Potential COVID-2019 3c-like protease inhibitors designed using generative deep learning approaches,” Feb. 2020. [Online]. Available: <https://doi.org/10.26434/chemrxiv.11829102.v2>

A white 3D DNA double helix structure is positioned on the left side of the image, extending from the top to the bottom. The background is a light blue gradient. The text "Thank you" is centered in the middle of the image.

Thank you