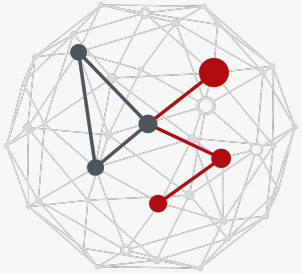
1222·2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO MATEMATICA



DATA SCIENCE
UNIVERSITY OF PADOVA

# STRUCTURAL BIOINFORMATICS

## CURRICULUM
## BIOLOGICAL DATA ANALYTICS

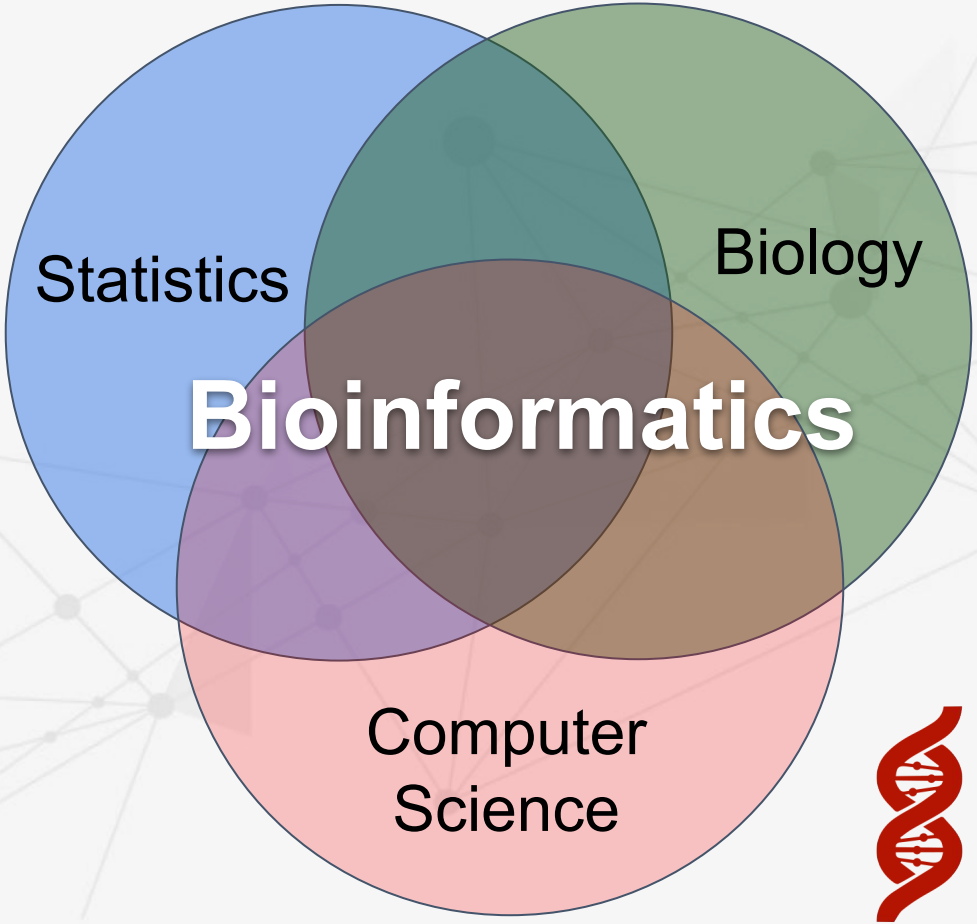
Master of Science in Data Science

Damiano Piovesan

# Bioinformatics

DIPARTIMENTO MATEMATICA

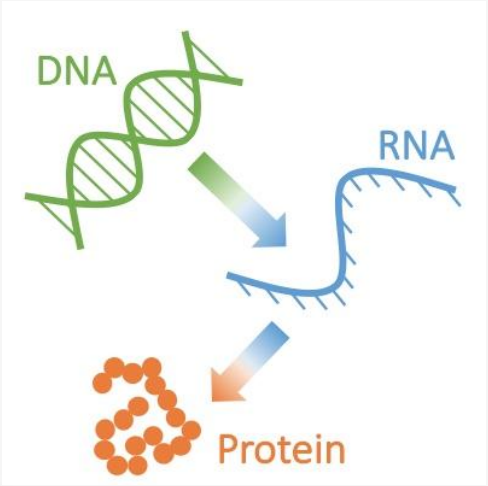
- Biological information can be studied with approaches typical of information theory
- Understanding biology requires statistical analyses
- Computer science as a tool for biological data management



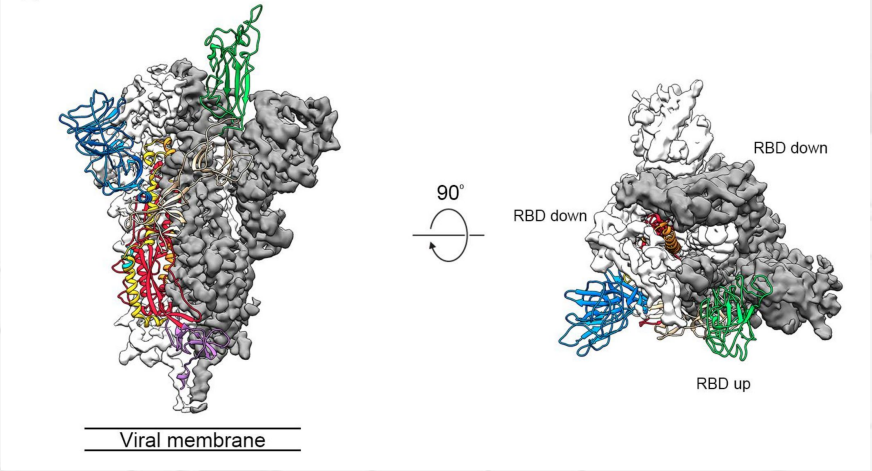
# Structural Bioinformatics

DIPARTIMENTO MATEMATICA

## The Central Dogma of life




## An example... **Coronavirus**



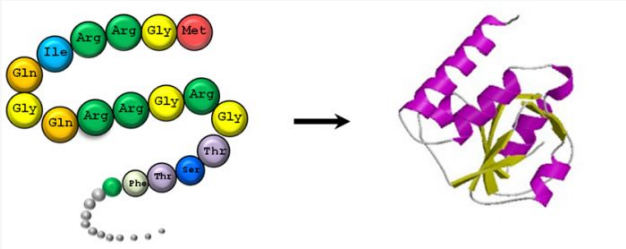

Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.

*Wrapp D et al.* ***Science.*** ***26 Feb 2020***

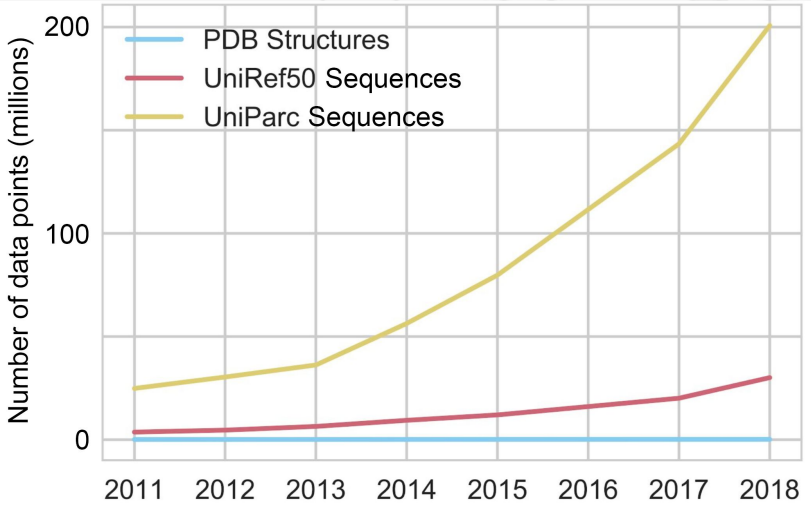
# Why structure prediction?

DIPARTIMENTO MATEMATICA

## From 1D to 3D




## The sequence-structure gap

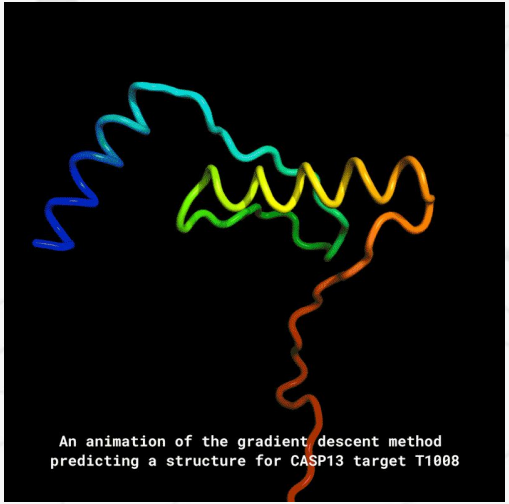


DIPARTIMENTO MATEMATICA

# The protein folding problem

An example, the protein folding problem:

- **Molecular Dynamics** simulations take **years** of calculation (on single CPU)
- **Deep learning** algorithms take **hours** / **minutes**




**AlphaFold**
Google DeepMind lab

DIPARTIMENTO MATEMATICA

# Course information

DIPARTIMENTO MATEMATICA

# Prof. Damiano Piovesan

- **Office hours**: Tuesday 15:00-16:00
- **Office**: Vallisneri building, ground floor (near the library)
- **Email:** damiano.piovesan@unipd.it
- **LinkedIn:** @DamianoPiovesan
- **Moodle**: STRUCTURAL BIOINFORMATICS 2025-2026 - **SCP7079278**
  - https://stem.elearning.unipd.it/course/view.php?id=**15125**
  - Enrollment key: **SB2026**

DIPARTIMENTO MATEMATICA

# Learning Objectives

**Knowledge & Understanding**

- Understand the fundamentals of protein structure (amino acids, chemical bonds, folding)
- Learn the principles of structural data acquisition (crystallography, databases)
- Grasp concepts of protein evolution, disorder, and aggregation
- Understand modern structure prediction methods (comparative modeling, ab initio, AlphaFold2)

**Technical Skills**

- Master PyMOL for molecular visualization and scripting
- Develop proficiency in BioPython for structural analysis
- Apply statistical potentials and embeddings to structural problems
- Understand basics of molecular dynamics simulations

DIPARTIMENTO MATEMATICA

# Learning Objectives

**Analytical Competencies**

- Analyze and compare protein structures computationally
- Predict and evaluate protein disorder
- Work with structural databases and tools (RING mentioned)
- Develop a final project applying learned techniques

**Computational Focus**

- Strong emphasis on programming and scripting for structural analysis
- Integration of classical bioinformatics with modern AI/ML approaches (embeddings, AlphaFold2)

# Prospective student's background

DIPARTIMENTO MATEMATICA

We welcome students with a **good background** in

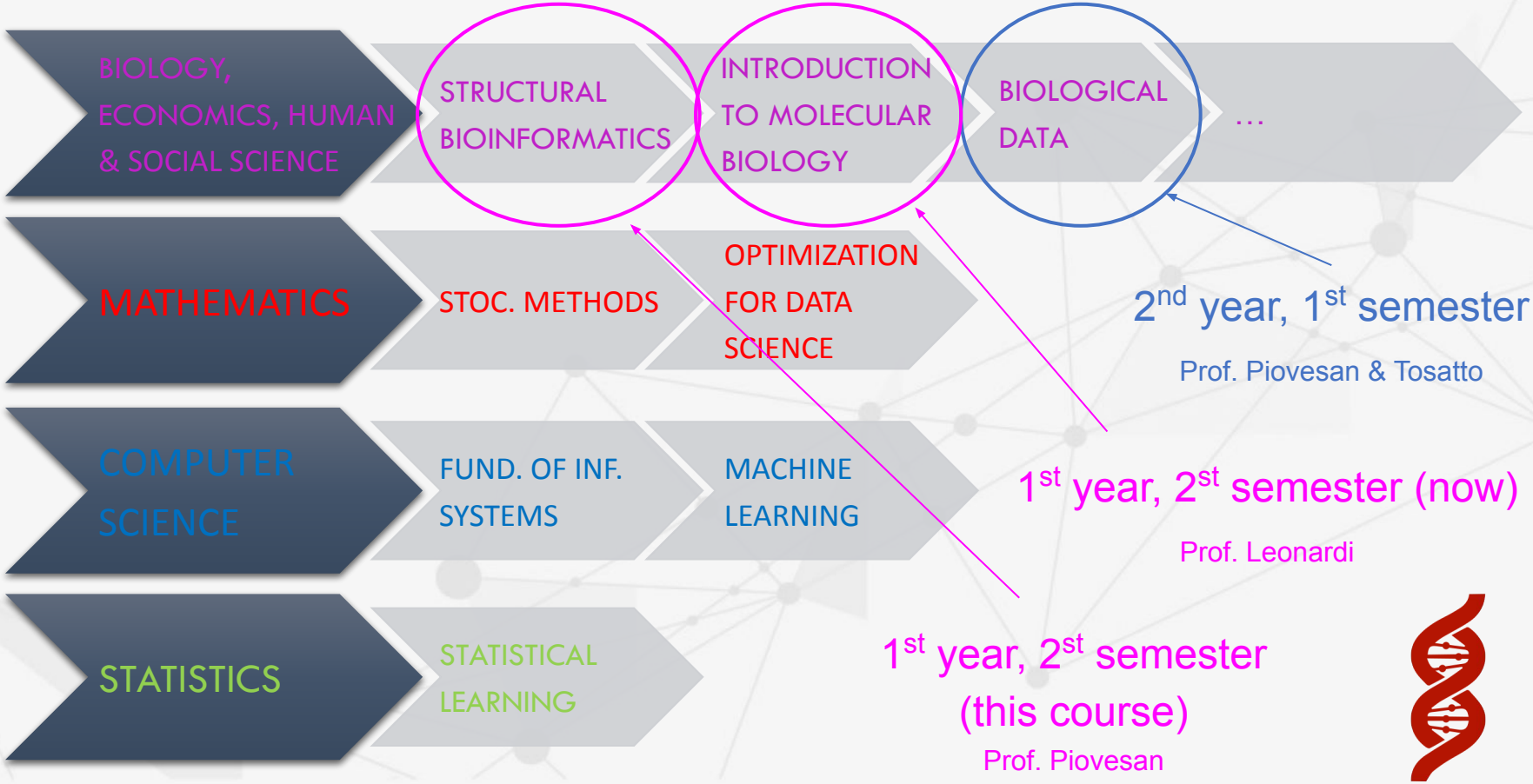
- Computer science, Computer Engineering, ICT
- Mathematics, Statistics, Physics
- Biology, Medicine and related (with math/statistics fundamentals)

Prospective students should

- Have an attitude on both basic science (biology, biochemistry, physics) and computer science (basic programming)

# BDA curriculum (for DS students)

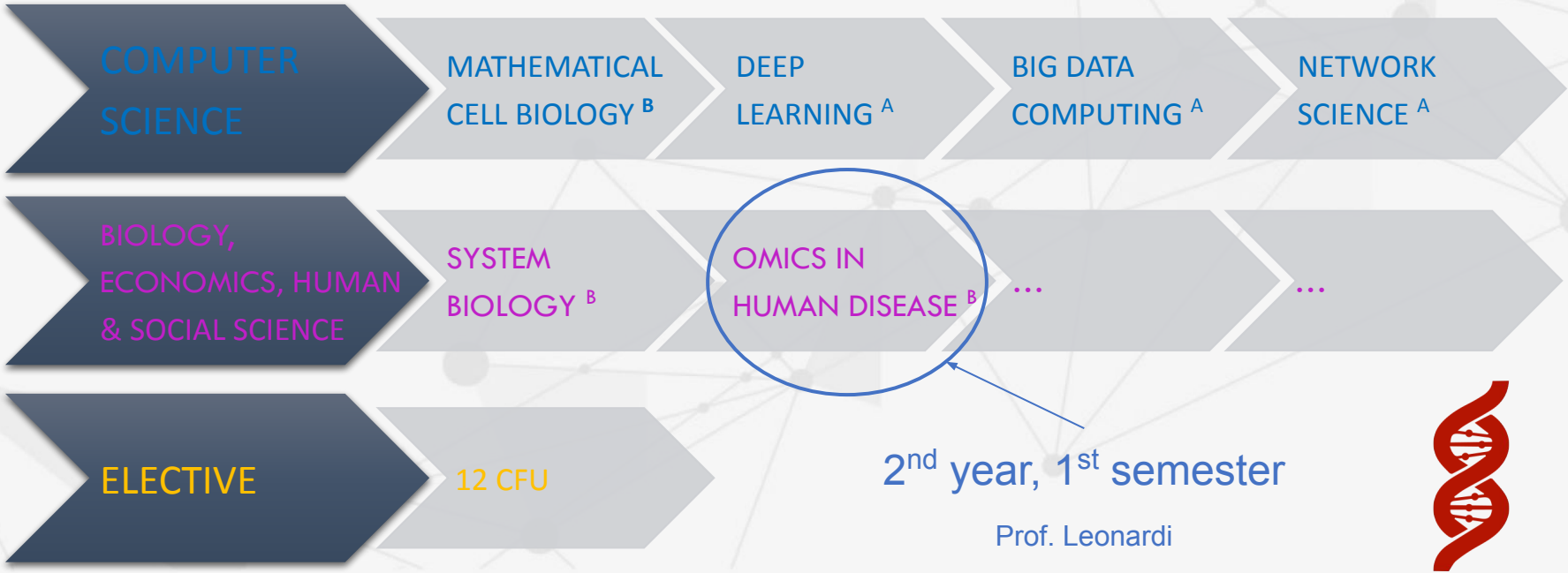
DIPARTIMENTO MATEMATICA



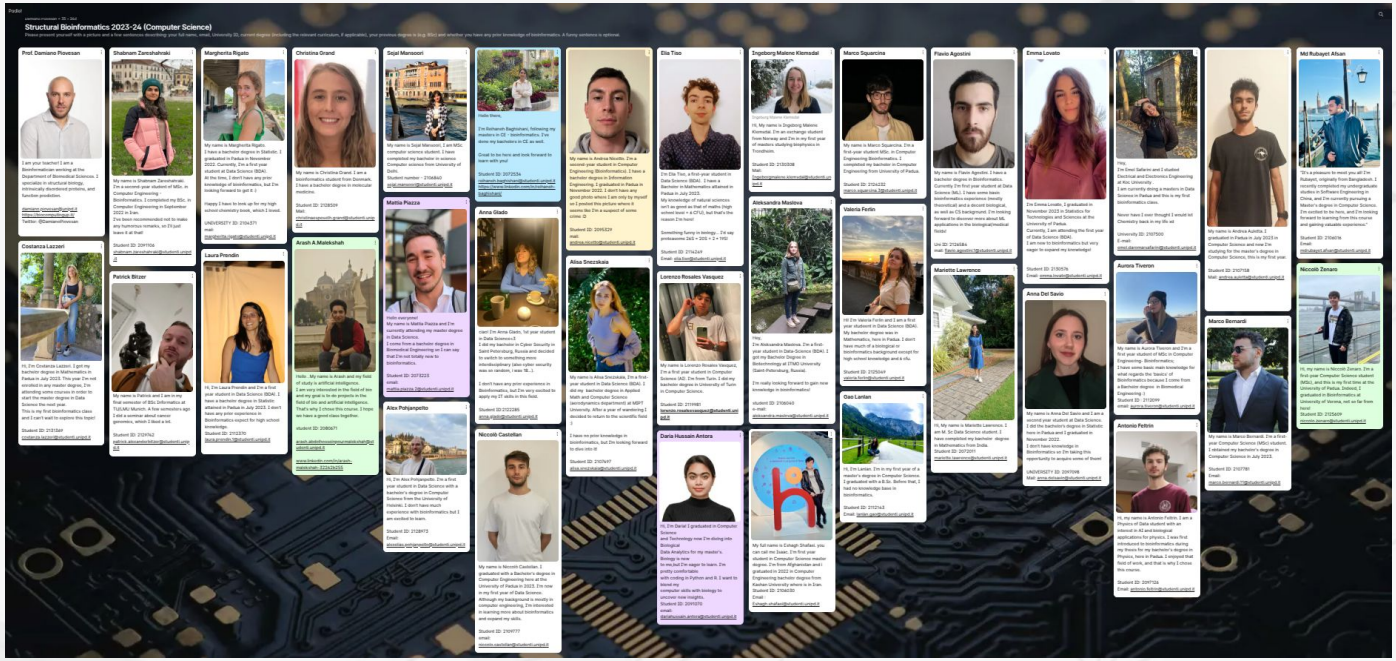
# BDA curriculum (for DS students)

DIPARTIMENTO MATEMATICA

- Two **optional** + two **electives** courses
- Optional: **two** from **group A** and **one** from **group B**



# Padlet




https://padlet.com/unipd/structural-bioinformatics-2025-26-computer-science-fsznlf2ncnxeayfy

DIPARTIMENTO MATEMATICA

# Student's sheet

- Used to exchange information and data with the students
- Filled by me once you add your card to the Padlet

SB_2022/23_students

File Edit View Insert Format Data Tools Extensions Help

A1:F1 Structural Bioinformatics

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | **Structural Bioinformatics** | | | | | |
| 2 | ***Please add your contacts and group number (1, 2, 3, ...) by 17th of May 2023*** ***Groups should be of 3/4 people, possibly with complementary skills (Bio + Math/CS)*** | | | | | |
| 3 | **Surname** | **Name** | **ID** | **Email** | **Degree program** | **Project group** |
| 4 | Pertile | Andrea Valentina | 2089070 | andreavalentina.pertile@studenti.unipd.it | Data science | group 6 |
| 5 | Ballarini | Marco | 2096997 | marco.ballarini@studenti.unipd.it | Data science | group 6 |
| 6 | Hassan | Syed Fahad | 2044060 | syedfahad.hassan@studenti.unipd.it | Computer Science | group 8 |
| 7 | Bottarelli | Federico | 2086505 | federico.bottarelli@studenti.unipd.it | Data science | |
| 8 | Yaseen | Mutahir | 2052021 | mutahir.yaseen@studenti.unipd.it | Computer Science | group 8 |
| 9 | Tabish | Muhammad | 2041579 | muhammad.tabish@studenti.unipd.it | Computer Science | group 8 |
| 10 | Airom | Omid | 2071818 | omid.airom@studenti.unipd.it | Data science | group 5 |
| 11 | Perlini | Daniela | 2097338 | daniela.perlini@studenti.unipd.it | Data science | group 7 |
| 12 | Canel | Alessandro | 2097570 | alessandro.canel@studenti.unipd.it | Computer Science | group 4 |
| 13 | Graves | Tanner | 2073559 | tanneraaron.graves@studenti.unipd.it | Data science | group 3 |
| 14 | Andriolo | Matteo | 2060889 | matteo.andriolo.2@studenti.unipd.it | Computer engineering | group 1 |
| 15 | Uderzo | Marco | 2096998 | marco.uderzo@studenti.unipd.it | Data Science | group 3 |
| 16 | Bedin | Veronica | 2097013 | veronica.bedin.1@studenti.unipd.it | Physics of data | group 4 |
| 17 | Giustini | Piermarco | 2056755 | piermarco.giustini@studenti.unipd.it | Computer engineering | group 1 |
| 18 | ~~Pisacreta~~ | ~~Giulia~~ | ~~2055299~~ | ~~giulia.pisacreta@studenti.unipd.it~~ | ~~Computer engineering~~ | |
| 19 | Kalhor | Amirhosein | 2072056 | amirhosein.kalhor@studenti.unipd.it | Data Science | group 5 |
| 20 | Stefanovska | Elena | 2085310 | elena.stefanovska@studenti.unipd.it | Data Science | group 2 |
| 21 | Riccò | Lorenzo | 2091906 | lorenzo.ricco@studenti.unipd.it | Computer Engineering | group 4 |
| 22 | Kralevska | Angela | 2072071 | angela.kralevska@studenti.unipd.it | Data Science | group 2 |
| 23 | Alhousseini | Nour | 2081230 | nour.alhousseini@studenti.unipd.it | Data Science | group 3 |
| 24 | Lawrence | Mariette | 2072011 | mariette.lawrence@studenti.unipd.it | Data Science | |
| 25 | Adebayo | Hazeezat Adebimpe | 2090254 | hazeezatadebimpe.adebayo@studenti.unipd.it | Data Science | group 3 |
| 26 | Pase | Emanuele | 2097904 | emanuele.pase@studenti.unipd.it | Computer Engineering | group 4 |
| 27 | Tomaselli | Francesco | 2089207 | francesco.tomaselli@studenti.unipd.it | Data Science | group 6 |
| 28 | Bartocci | Giulia | 2076673 | giulia.bartocci@studenti.unipd.it | Data science | group 6 |
| 29 | Antora | Daria Hussain | 2091070 | dariahussain.antora@studenti.unipd.it | Data Science | |
| 30 | Syed Riaz | Raza | 2041583 | syedriaz.raza@studenti.unipd.it | Computer Science | group 8 |
| 31 | Ferrero | Elisa | 2095871 | elisa.ferrero@studenti.unipd.it | Data Science | group 7 |
| 32 | Mandal | Rana | 2041577 | rana.mandal@studenti.unipd.it | Computer Science | group 9 |

# Evaluation criteria

DIPARTIMENTO MATEMATICA

**Practicals (2 x 15 points; 20% of total)**

- Work assignments during practical sessions, to be returned via email
- Tentatively end-March and end-April
- Mandatory to access the final exam

**Group project (60 points; 40% of total)**

- Code, data and written report to hand in (at least) ten days before exam
- Each group will have a different project
- Groups of 2-3 students

**Final oral examination (60 points; 40% of total)**

- Discussion of the project and written report
- Questions from the course material

**Optional bonuses**

- Interesting contribution to lecture

Scores are converted to thirty points maximum. Threshold for 18 will be ca. 50% of the maximum score, in each part

DIPARTIMENTO MATEMATICA

# Success stories

jmb Journal of Molecular Biology

Go to Journal of Molecular Biology on ScienceDirect

Available online 27 February 2021, 166900

In Press, Journal Pre-proof

Research Article

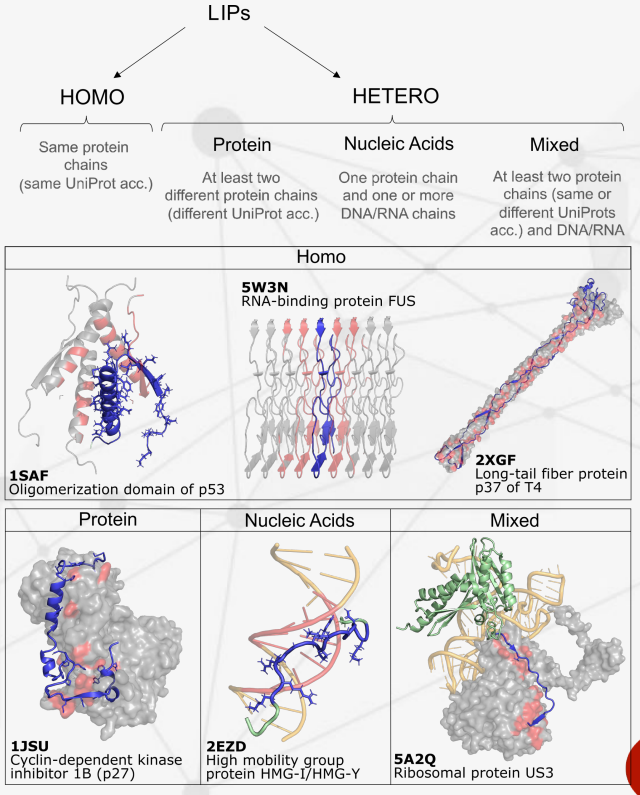
# FLIPPER: predicting and characterizing linear interacting peptides in the Protein Data Bank

Alexander Miguel Monzon, Paolo Bonato, Marco Necci, Silvio C.E. Tosatto, Damiano Piovesan

Dept. of Biomedical Sciences, University of Padua, Via Ugo Bassi 58/B, Padua, 35121, Italy

Received 3 December 2020, Revised 22 February 2021, Accepted 22 February 2021, Available online 27 February 2021.

CS student 2018/19




JMB - Impact Factor 6.151, 1st quartile considering “Biochemistry and Molecular Biology” category

# Disorder prediction

*Bioinformatics*, 36(22-23), 2020, 5533–5534
doi: 10.1093/bioinformatics/btaa1045
Advance Access Publication Date: 16 December 2020
Applications Note

Structural bioinformatics

## MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavors in proteins

**Marco Necci[1], Damiano Piovesan[1], Damiano Clementel[1], Zsuzsanna Dosztányi[2] and Silvio C. E. Tosatto[1,*]**

[1]Department of Biomedical Sciences, University of Padua, 35121 Padova, Italy and [2]MTA-ELTE Lendület Bioinformatics Research Group, Department of Biochemistry, ELTE Eötvös Loránd University, Budapest, Hungary

*To whom correspondence should be addressed.
Associate Editor: Yann Ponty

Received on August 5, 2020; revised on November 3, 2020; editorial decision on December 5, 2020; accepted on December 7, 2020

DS student 2018/19

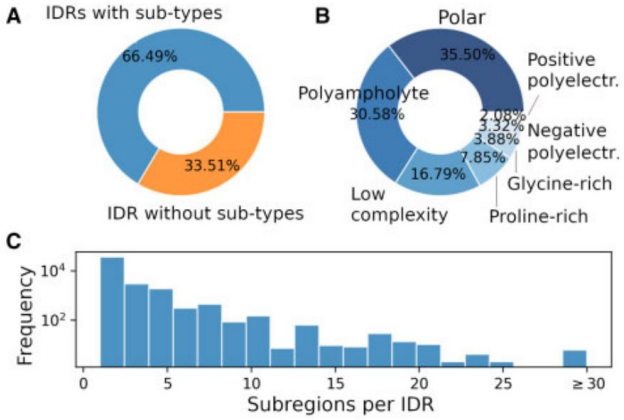



**Fig. 1.** Abundance of IDRs and subregions. MobiDB-lite results for the human proteome. (**A**) Fraction of IDRs with (blue) and without subtypes (orange). (**B**) Distribution of IDR subtypes. (**C**) Distribution of subregions detected per IDR, plotted on a logarithmic frequency scale (*y*-axis)

Bioinformatics - Impact Factor 6.931, 1st quartile

# Function prediction

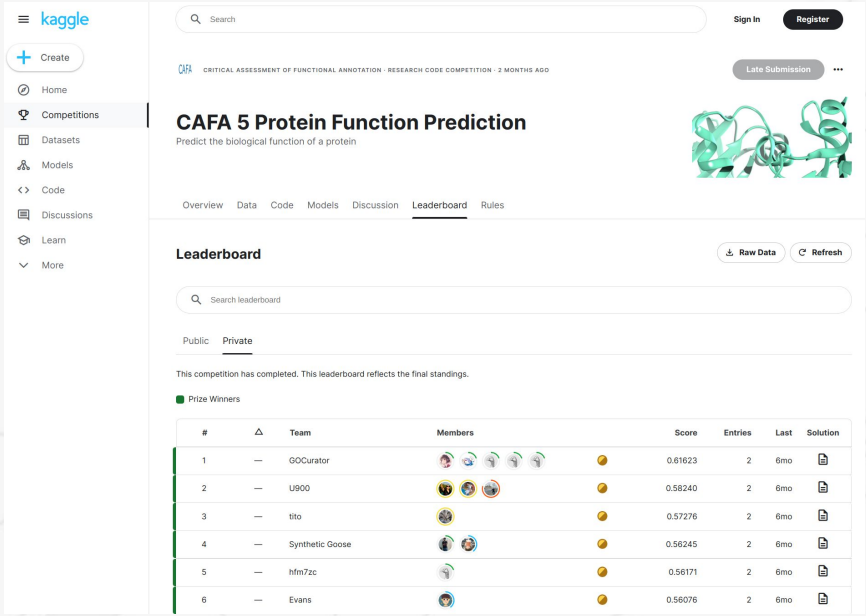
## CAFA-evaluator: A Python Tool for Benchmarking Ontological Classification Methods

Damiano Piovesan[1,*], Davide Zago[1], Parnal Joshi[2,3], M. Clara De Paolis Kaluza[4], , Mahta Mehdiabadi[1], Rashika Ramola[4], Alexander Miguel Monzon[5], Walter Reade[6], Iddo Friedberg[3], Predrag Radivojac[4], Silvio C. E. Tosatto[1]

[1]Department of Biomedical Sciences, University of Padova, Padova 35121, Italy.
[2]Program in Bioinformatics and Computational Biology, Iowa State University, Ames IA, 50011 USA.
[3]Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames IA, 50011 USA.
[4]Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA.
[5]Department of Information Engineering, University of Padova, Padova 35121, Italy.
[6]Kaggle, San Francisco, CA, USA

CS student 2021/22
PhD student 2023/now

CS student 2020/21




Bioinformatics Advances - Impact Factor 2.4, 2nd quartile

# Structural biology

*Bioinformatics*, 2025, **41(7)**, btaf395
https://doi.org/10.1093/bioinformatics/btaf395
Advance Access Publication Date: 10 July 2025
**Applications Note**

OXFORD

Structural bioinformatics

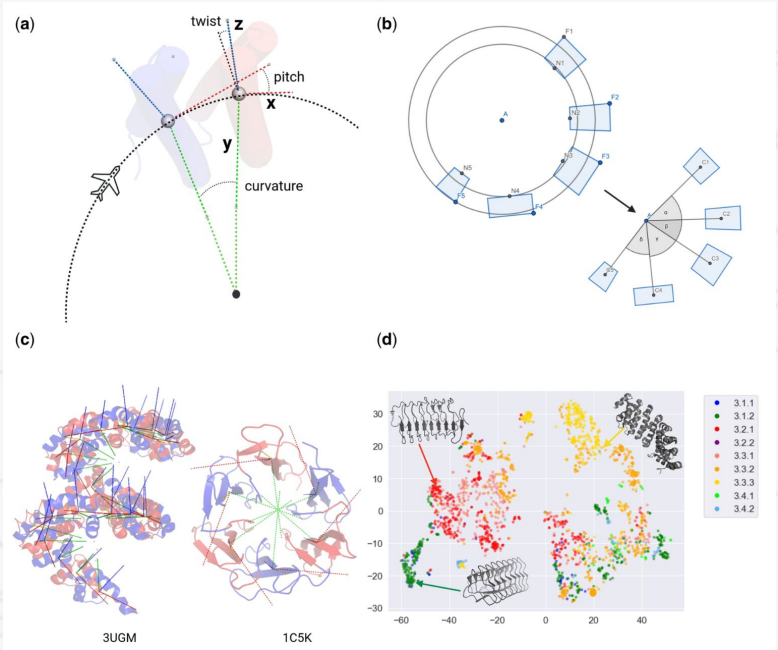
## GeomeTRe: accurate calculation of geometrical descriptors of tandem repeat proteins

**Zarifa Osmanli[1,†], Elisa Ferrero[1,2,†], Alexander Miguel Monzon[1], Silvio C.E. Tosatto[1,3], Damiano Piovesan[1,*]**

[1]Department of Biomedical Sciences, University of Padova, Padova, 35121, Italy
[2]Galilean School of Higher Education, University of Padova, Padova, 35132, Italy
[3]Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, 70126, Italy
*Corresponding author. Department of Biomedical Sciences, University of Padova, Via U. Bassi 58/b, Padova 35121, Italy. E-mail: damiano.piovesan@unipd.it.
[†]= equal contribution.

Associate Editor: Jianlin Cheng

DS student 2022/23

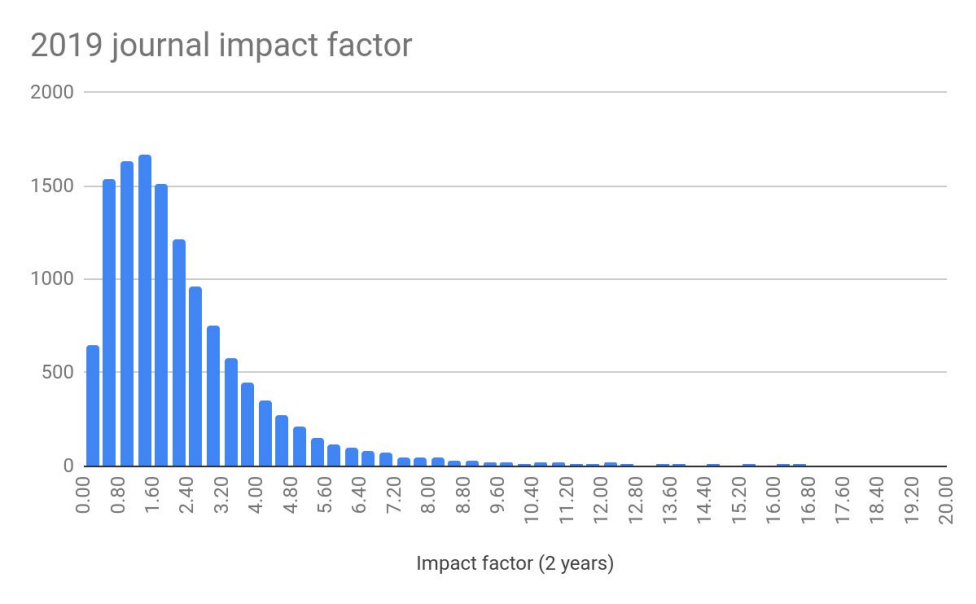



Bioinformatics - Impact Factor 6.931, 1st quartile

# Journal impact factor

DIPARTIMENTO MATEMATICA

$$\mathrm{IF}_{2017} = \frac{\mathrm{Citations}_{2017}}{\mathrm{Publications}_{2016} + \mathrm{Publications}_{2015}} = \frac{74090}{880 + 902} = 41.577$$




Clarivate InCite Journal Citation Report (https://jcr.clarivate.com)

# Journals by category

DIPARTIMENTO MATEMATICA

| Rank | Category | # Journals | Total Cites | Median Impact Factor |
|---|---|---|---|---|
| 1 | ECONOMICS | 373 | 1,116,324 | 1.411 |
| 2 | MATHEMATICS | 325 | 565,491 | 0.794 |
| 3 | MATERIALS SCIENCE, MULTIDISCIPLINARY | 314 | 5,001,499 | 2.528 |
| **4** | **BIOCHEMISTRY & MOLECULAR BIOLOGY** | **297** | **3,962,918** | **3.167** |
| 5 | **NEUROSCIENCES** | **272** | **2,548,868** | **3.052** |
| 6 | **PHARMACOLOGY & PHARMACY** | **271** | **1,769,982** | **2.681** |
| 7 | ENGINEERING, ELECTRICAL & ELECTRONIC | 266 | 2,097,197 | 2.204 |
| 8 | ENVIRONMENTAL SCIENCES | 265 | 2,622,685 | 2.400 |
| 9 | EDUCATION & EDUCATIONAL RESEARCH | 263 | 421,337 | 1.585 |
| 10 | MATHEMATICS, APPLIED | 261 | 631,053 | 1.162 |
| 11 | **ONCOLOGY** | **244** | **2,187,273** | **3.297** |
| 12 | **PLANT SCIENCES** | **234** | **1,292,836** | **1.576** |
| 13 | MANAGEMENT | 226 | 892,977 | 2.465 |
| 14 | **SURGERY** | **210** | **1,327,558** | **1.901** |
| 15 | **CLINICAL NEUROLOGY** | **204** | **1,438,476** | **2.611** |
| 16 | GEOSCIENCES, MULTIDISCIPLINARY | 200 | 1,166,162 | 2.207 |
| 17 | **CELL BIOLOGY** | **195** | **2,359,746** | **3.683** |
| 18 | PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH | 193 | 1,050,617 | 2.104 |
| 19 | LINGUISTICS | 187 | 177,076 | 0.895 |
| 20 | POLITICAL SCIENCE | 181 | 270,661 | 1.366 |