ODJEL ZA INFORMATIKU
SVEUČILIŠTE U RIJECI

Research of New Drugs and Social Aspects of a Pandemic Caused by Coronavirus Using ICT

Ana Meštrović, Martina Ašenbrener Katić,
Vedran Miletić, Patrizia Poščić

ICT-20 @ COVID-19 online conference, MIPRO, 27 May 2020

DEPARTMENT OF INFORMATICS

UNIVERSITY OF RIJEKA

Radmile Matejčić 2, Rijeka
http://www.inf.uniri.hr/
ured@inf.uniri.hr

- One of the newest member institutions of the UNIRI
- Founded in April 2008, by separating from the Faculty of Humanities and Social Sciences
- Different studies of informatics since 1985 (informatics-pedagogy, mathematics-informatics, physics-informatics, …)
- Informatics single major programme was started in 2005
- Department moved to the new building at the Campus in 2012

Students & Employees

- Around 500 students (BSc+MSc):
- 250-350: Informatics
- 100 fixed educational programs:
  - Informatics and Mathematics
  - Informatics and Physics
  - Informatics and Polytechnics
- 40-50 PhD students

- 6 Full professors
- 4 Associate professors
- 7 Assistant professors
- 1 postdoc
- 1 senior lecturer
- 7 Research assistants
- 1 Office Manager

Research fields at the Department

- Intelligent systems
- Man-machine communications
- Computer vision (Action recognition)
- Complex networks (Language networks)
- Language technologies (NLP)
- Semantic technologies
- Business informatics (IS)
- e-learning (data mining in e-learning)

„COVID” science at the Department

- two projects submitted to the HRZZ calls
  - *„SARS-CoV-2 high-throughput virtual screening campaign and target protein inhibitors drug design”*
    - assist. prof. dr. Martina Ašenbrener Katić and dr. Vedran Miletić, Senior Lecturer
  - *„Multilayer Framework for the Information Spreading Characterization in Social Media during the COVID-19 Crisis”*
    - assoc. prof. dr. Ana Meštrović

ODJEL ZA INFORMATIKU
SVEUČILIŠTE U RIJECI

SARS-CoV-2 high-throughput virtual screening campaign and target protein inhibitors drug design (COVIDOCK)

Assist. Prof. Dr. Martina Ašenbrener Katić,
Dr. Vedran Miletić, Senior Lecturer
masenbrener@uniri.hr, vmiletic@uniri.hr

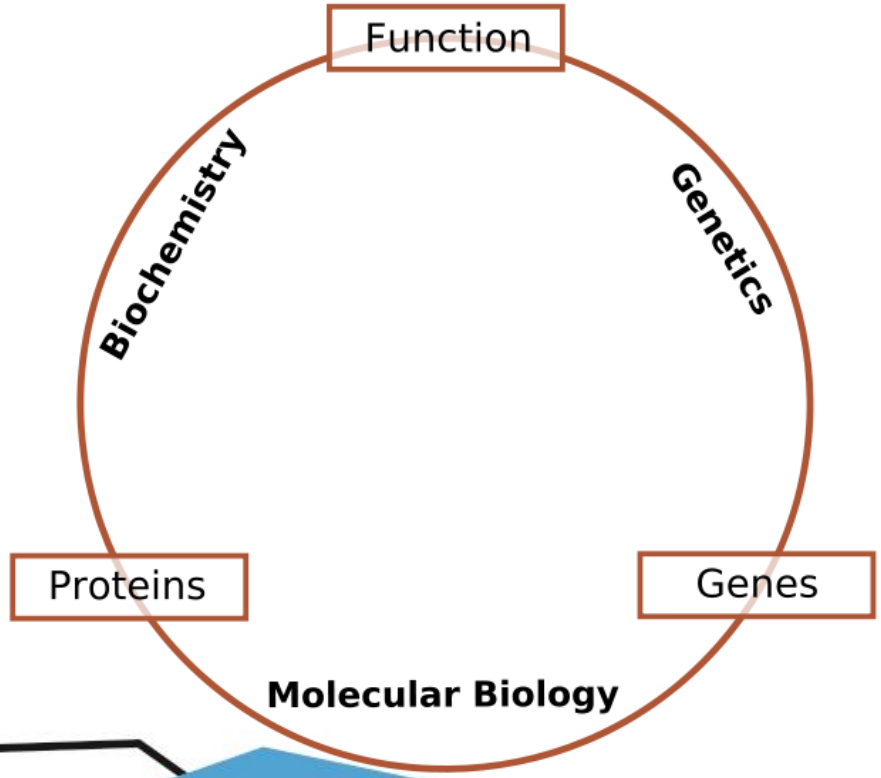
ICT-20 @ COVID-19 online conference, MIPRO, 27 May 2020

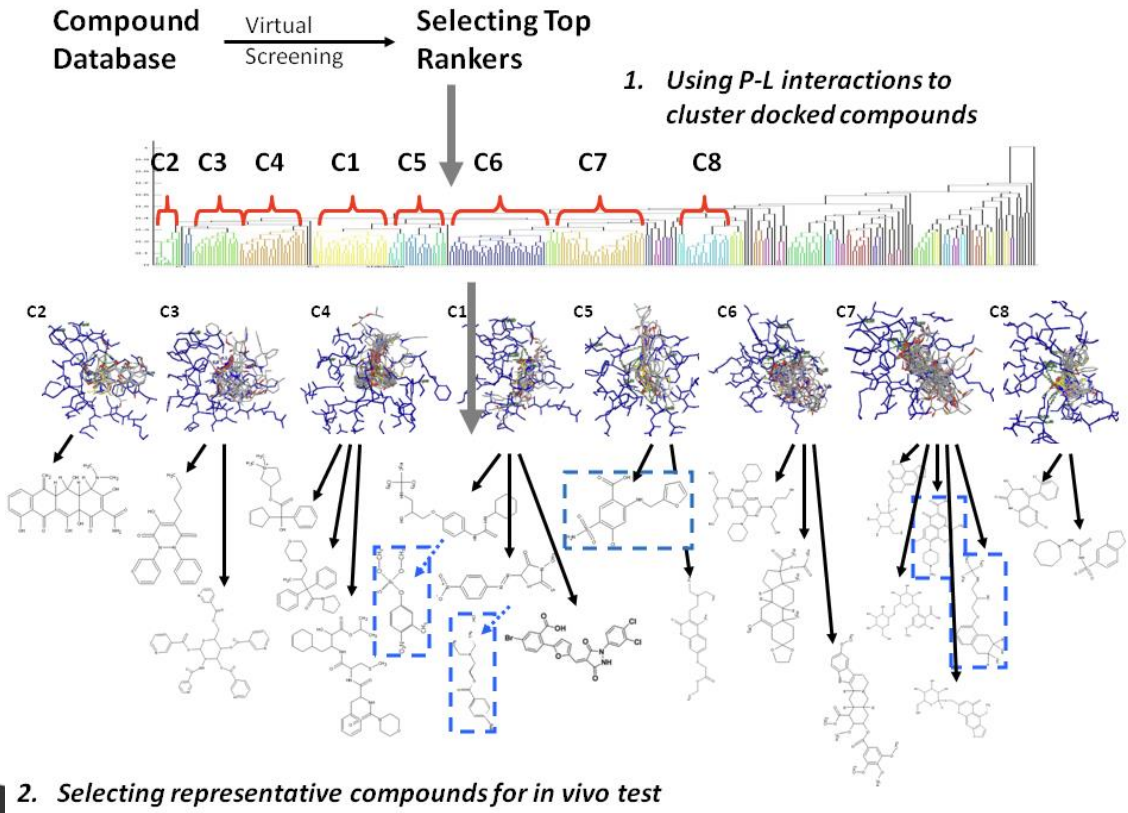
Structure-based drug design

"If you want to understand function, study structure," *I was supposed to have said in my molecular biology days.* ***I think that one should approach these problems at all levels, as was done in molecular biology.*** *Classical genetics is, after all, a black-box subject. The important thing was to combine it with biochemistry.*

Francis Crick, What Mad Pursuit (1989)

Structure-based drug design



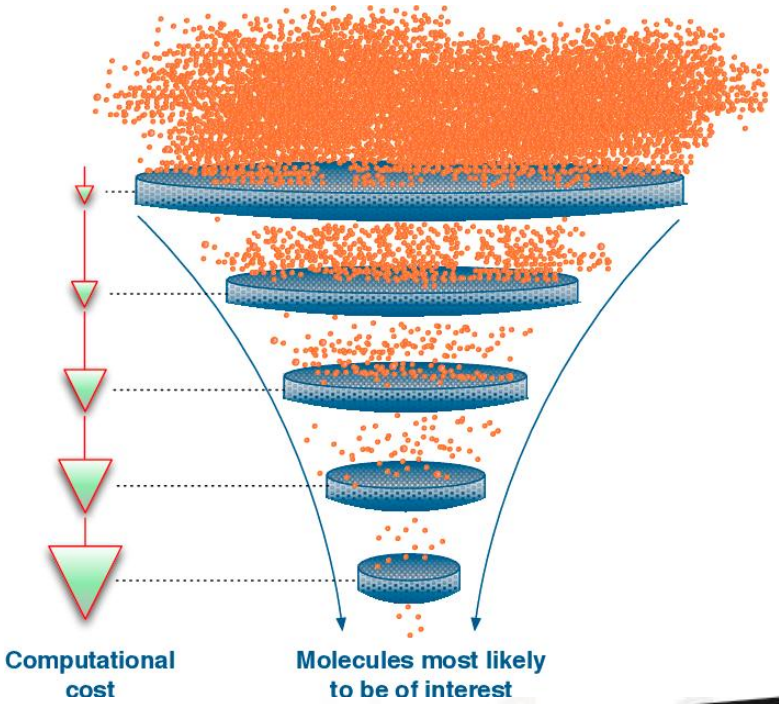


Compound screening

- Complete analysis of 1 ligand – 1 target:
  - Complete MD analysis of ligand – target system with solvation, cofactors and membranes (if applicable)
  - ADMET analysis (toxicity, solubility, Lipinsky rules, general pharmakokinetics considerations), synthesis pathways
- At least one week, provided one has the adequate structure and lab setup, trained team, and full time devotion to the problem at hand

Compound screening

- Average screening library has 100 000 compounds
  - 100 000 weeks => almost 2000 years
  - **Do you have 2000 years for initial screening?**



Pyzer-Knapp, Edward O., Changwon Suh, Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre and Alán Aspuru-Guzik. “What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery.” (2015).

And screening happened!

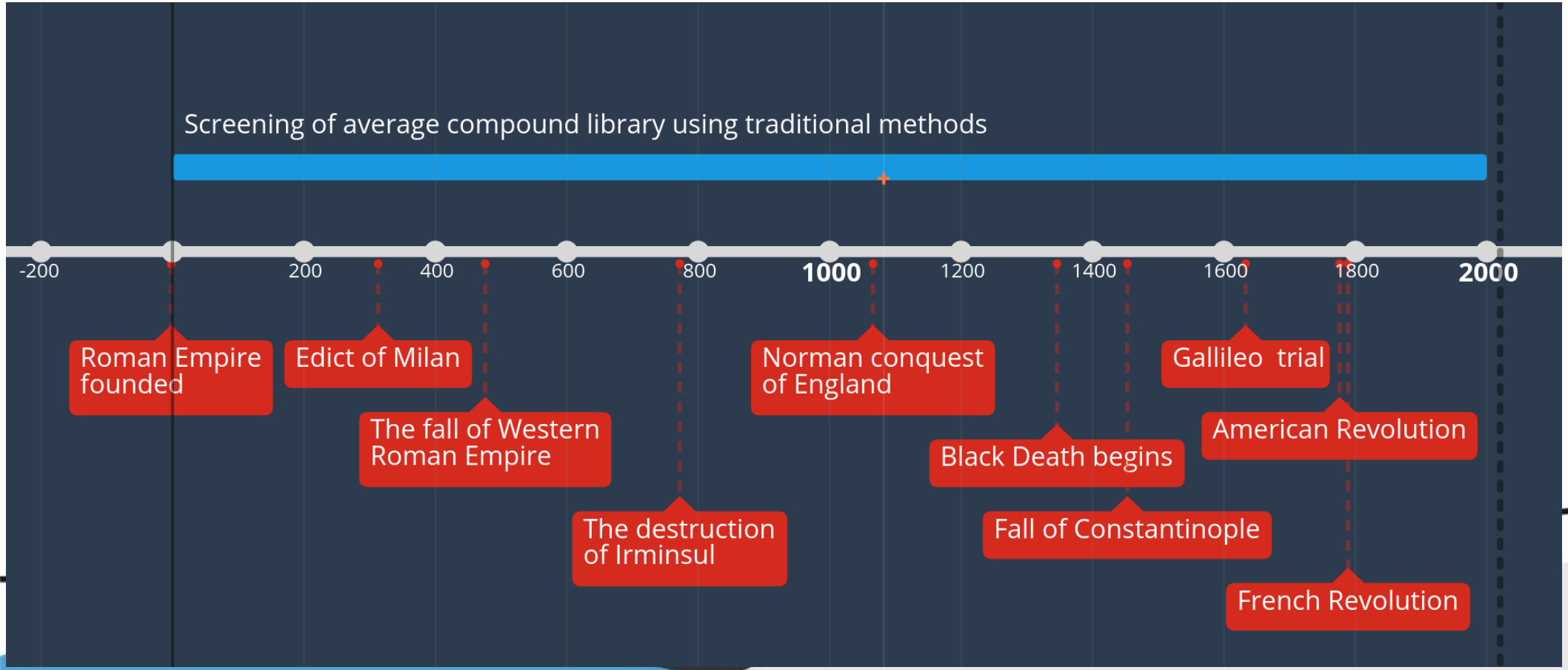
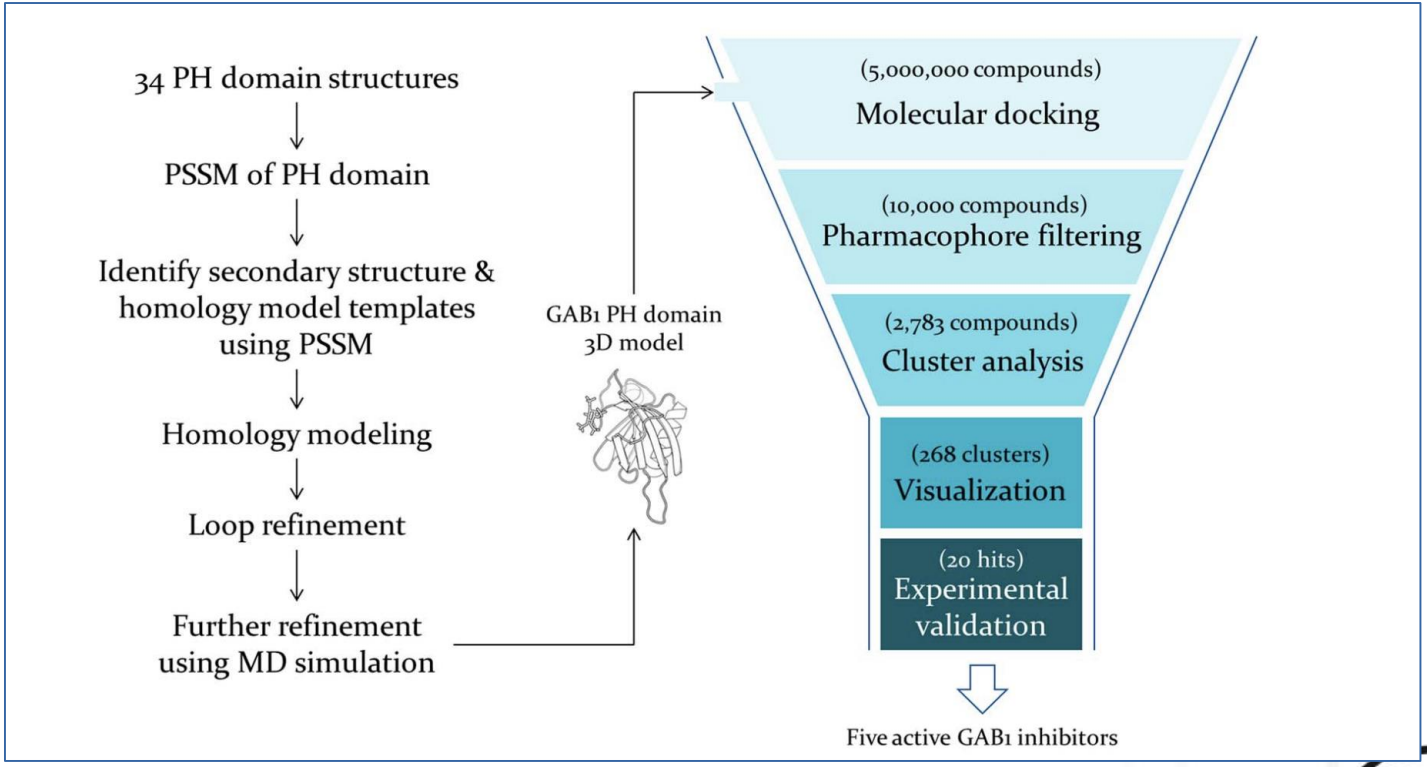


Table 1. Typical costs of experiments [14].

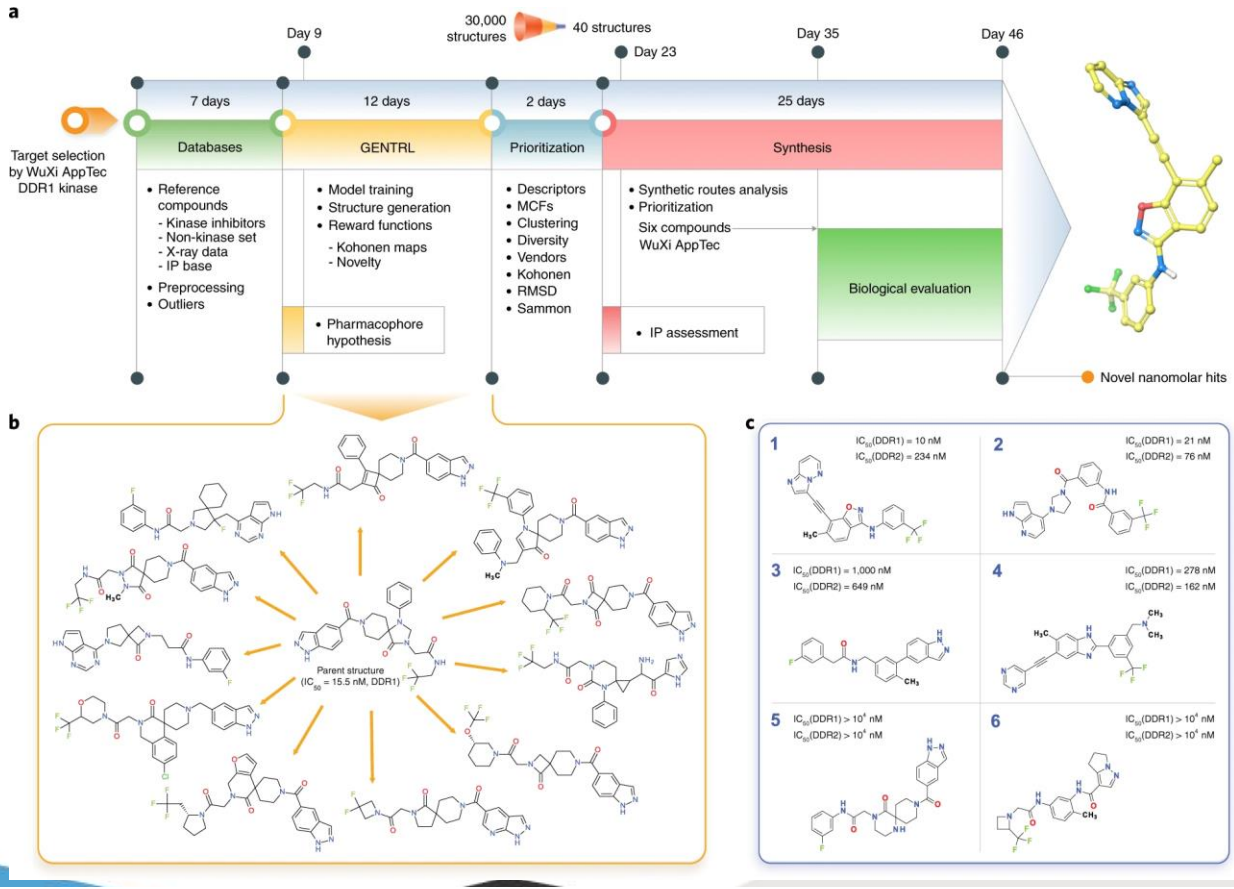
Experiment	Typical Cost per Compound ($)
Computer modeling	10
Biochemical assay	400
Cell culture assay	4,000
Rat acute toxicity	12,000
Protein crystal structure	100,000
Animal efficacy trial	300,000
Rat 2-years chronic oral toxicity	800,000
Human clinical trial	500,000,000

GAB1 inhibitors



Chen, L. et al. Novel Inhibitors Induce Large Conformational Changes of GAB1 Pleckstrin Homology Domain and Kill Breast Cancer Cells. PLoS Comput Biol 11, e1004021 (2015). DOI: 10.1371/journal.pcbi.1004021

DDR1 kinase inhibitors



Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol 37, 1038–1040 (2019). DOI: 10.1038/s41587-019-0224-x

Existing COVID-19 research

- Small molecule ligands: a number of public and private databases of screened compounds
- Peptide binders
  - 23-mer ACE2 peptide binds SARS-CoV-2 spike protein in vitro with 47nM affinity [1]
  - 31-mer peptidic scaffold linking two fragments grafted from hACE2 with glycine linker to bind to SARS-CoV-2 RBD [2]

[1] Zhang, G., Pomplun, S., Loftis, A. R., Loas, A. & Pentelute, B. L. The first-in-class peptide binder to the SARS-CoV-2 spike protein. bioRxiv 2020.03.19.999318 (2020) doi:10.1101/2020.03.19.999318.
[2] Huang, X., Pearce, R. & Zhang, Y. Computational Design of Peptides to Block Binding of the SARS-CoV-2 Spike Protein to Human ACE2. bioRxiv 2020.03.28.013607 (2020) doi:10.1101/2020.03.28.013607.

The first-in-class peptide binder to the SARS-CoV-2 spike protein

G. Zhang, S. Pomplun, A. R. Loftis, A. Loas, B. L. Pentelute

doi: https://doi.org/10.1101/2020.03.19.999318

Abstract

Coronavirus disease 19 (COVID-19) is an emerging global health crisis. With over 200,000 confirmed cases to date, this pandemic continues to expand, spurring research to discover vaccines and therapies. SARS-CoV-2 is the novel coronavirus responsible for this disease. It initiates entry into human cells by binding to angiotensin-converting enzyme 2 (ACE2) via the receptor binding domain (RBD) of its spike protein (S). Disrupting the SARS-CoV-2-RBD binding to ACE2 with designer drugs has the potential to inhibit the virus from entering human cells, presenting a new modality for therapeutic intervention. Peptide-based binders are an attractive solution to inhibit the RBD-ACE2 interaction by adequately covering the extended protein contact interface. Using molecular dynamics simulations based on the recently solved ACE2 and SARS-CoV-2-RBD co-crystal structure, we observed that the ACE2 peptidase domain (PD) α1 helix is important for binding SARS-CoV-2-RBD. Using automated fast-flow peptide synthesis, we chemically synthesized a 23-mer peptide fragment of the ACE2 PD α1 helix composed entirely of proteinogenic amino acids. Chemical synthesis of this human derived sequence was complete in 1.5 hours and after work up and isolation >20 milligrams of pure material was obtained. Bio-layer interferometry revealed that this peptide specifically associates with the SARS-CoV-2-RBD with low nanomolar affinity. This peptide binder to

Computational Design of Peptides to Block Binding of the SARS-CoV-2 Spike Protein to Human ACE2

Xiaoqiang Huang, Robin Pearce, Yang Zhang

doi: https://doi.org/10.1101/2020.03.28.013607

ABSTRACT

The outbreak of COVID-19 has now become a global pandemic and it continues to spread rapidly worldwide, severely threatening lives and economic stability. Making the problem worse, there is no specific antiviral drug that can be used to treat COVID-19 to date. SARS-CoV-2 initiates its entry into human cells by binding to angiotensin-converting enzyme 2 (hACE2) via the receptor binding domain (RBD) of its spike protein. Therefore, molecules that can block SARS-CoV-2 from binding to hACE2 may potentially prevent the virus from entering human cells and serve as an effective antiviral drug. Based on this idea, we designed a series of peptides that can strongly bind to SARS-CoV-2 RBD in computational experiments. Specifically, we first constructed a 31-mer peptidic scaffold by linking two fragments grafted from hACE2 (a.a. 22-44 and 351-357) with a linker glycine, and then redesigned the peptide sequence to enhance its binding affinity to SARS-CoV-2 RBD. Compared with several computational studies

Inhibition of SARS-CoV-2 infection (previously 2019-nCoV) by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion

Shuai Xia, Meiqin Liu, Chao Wang, Wei Xu, Qiaoshuai Lan, Siliang Feng, Feifei Qi, Linlin Bao, Lanying Du, Shuwen Liu, Chuan Qin, Fei Sun, Zhengli Shi, Yun Zhu, Shibo Jiang, Lu Lu

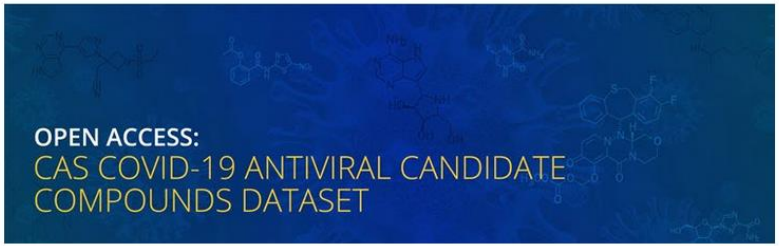
doi: https://doi.org/10.1101/2020.03.09.983247

Now published in *Cell Research* doi: 10.1038/s41422-020-0305-x

Abstract

The recent outbreak of coronavirus disease (COVID-19) caused by SARS-CoV-2 infection in Wuhan, China has posed a serious threat to global public health. To develop specific anti-coronavirus therapeutics and prophylactics, the molecular mechanism that underlies viral infection must first be confirmed. Therefore, we herein used a SARS-CoV-2 spike (S) protein-mediated cell-cell fusion assay and found that SARS-CoV-2 showed plasma membrane fusion capacity superior to that of SARS-CoV. We solved the X-ray crystal structure of six-helical bundle (6-HB) core of the HR1 and HR2 domains in SARS-CoV-2 S protein S2 subunit, revealing that several mutated amino acid residues in the HR1 domain may be associated with enhanced interactions with HR2 domain. We previously developed a pan-coronavirus fusion inhibitor, EK1, which targeted HR1 domain and could inhibit infection by divergent human coronaviruses tested, including SARS-CoV and MERS-CoV. We then generated a series of lipopeptides and found that the EK1C4 was the most potent fusion inhibitor against SARS-CoV-2 S protein-mediated membrane fusion and pseudovirus infection with IC_{50}s of 1.3 and 15.8 nM, about 241- and 149-fold more potent than that of EK1 peptide, respectively. EK1C4

Download CAS COVID-19 Antiviral Candidate Compounds Dataset



As a specialist in scientific information solutions, CAS is partnering with research organizations around the globe to tackle the complex and rapidly evolving challenge of COVID-19. Aligned with our mission as a division of the American Chemical Society, CAS is making a wide range of assets, expertise, and resources available to support this fight.

As part of this effort, CAS has built an open source dataset assembled from CAS REGISTRY® including known anti-viral drugs and related chemical compounds that are structurally similar to known antivirals. The dataset license terms support use for applications including research, data mining, machine learning, and analytics at no charge.

This dataset may be downloaded by filling in the form below

About the Dataset

The dataset is in SD file format (.sdf) and contains connection tables for nearly 50,000 chemical substances, along with related metadata such as CAS Registry Number® and physical properties for each substance (details).

LIFE DIGITAL CHEMISTRY

ARTICLE

COVID-19: 1.5 Billion Compounds to Undergo Virtual Screening

05.18.2020, by Jonathan Rangapanaiken

© Science Photo Library / Tek Image

Sections

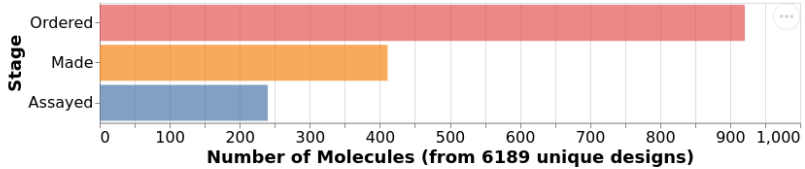
French scientists have set up a multidisciplinary, large-scale virtual screening project that should enable them, during the next 18 months,

Compound Tracker

Similarity/Substructure Search | Export CSV

Data Filters

Ordered:	No	Any	Yes
Made:	No	Any	Yes
Assayed:	No	Any	Yes



Show/Hide Columns

Molecular Properties

- [x] MW
- [x] cLogP
- [x] TPSA
- [x] Rotatable Bonds
- [x] Fraction sp3
- [x] HBA
- [x] HBD
- [x] Covalent Fragment
- [x] Covalent Warhead
- [ ] Criterion Violations

Activity

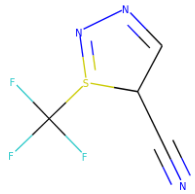
- [ ] IC50 (μM)
- [ ] Trypsin IC50 (μM)
- [ ] Average Inhibition @ 20 μM
- [ ] Average Inhibition @ 50 μM

ADMET

- [ ] Relative Solubility @ 20 μM
- [ ] Relative Solubility @ 100 μM

Showing 1 to 10 of 6,466 entries

Search:

Image	Molecule	MW	cLogP	TPSA	Rotatable Bonds	Fraction sp3	HBA	HBD	Covalent Fragment
	ABI-SAT-2adc218e-1 N#CC1C=NN=S1C(F)(F)F	180.99	1.20	48.51	0	0.50	3	0	false
	ABI-SAT-2adc218e-2 NC1CC1C(F)(F)F	125.05	0.90	26.02	0	1.00	1	1	false

Drug repositioning

International Journal of *Molecular Sciences*

MDPI

Article

Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds

André Fischer, Manuel Sellner, Santhosh Neranjan and Martin Smieško * and Markus A. Lill *

Computational Pharmacy, Department of Pharmaceutical Sciences, University of Basel, 4056 Basel, Switzerland; and.fischer@unibas.ch (A.F.); manuel.sellner@unibas.ch (M.S.); santhosh.neranjan@stud.unibas.ch (S.N.)

* Correspondence: martin.smiesko@unibas.ch (M.S.); markus.lill@unibas.ch (M.A.L.)

Received: 22 April 2020; Accepted: 16 May 2020; Published: 21 May 2020

check for updates

of the main protease (M^{pro}) of SARS-CoV-2. A screening of such a vast chemical space for SARS-CoV-2 M^{pro} inhibitors has not been reported before. After shape screening, two docking protocols were applied followed by the determination of molecular descriptors relevant for pharmacokinetics to narrow down the number of initial hits. Next, molecular dynamics simulations were conducted to validate the stability of docked binding modes and comprehensively quantify ligand binding energies. After evaluation of potential off-target binding, we report a list of 12 purchasable compounds, with binding affinity to the target protease that is predicted to be more favorable than that of the cocrystallized peptidomimetic compound. In order to quickly advise ongoing therapeutic intervention for patients, we evaluated approved antiviral drugs and other protease inhibitors to provide a list of nine compounds for drug repurposing. Furthermore, we identified the natural compounds (−)-taxifolin and rhamnetin as potential inhibitors of M^{pro}. Rhamnetin is already commercially available in pharmacies.

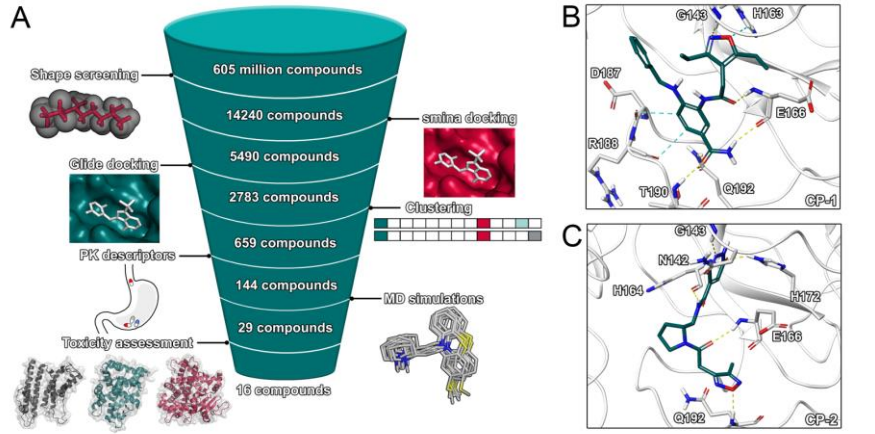
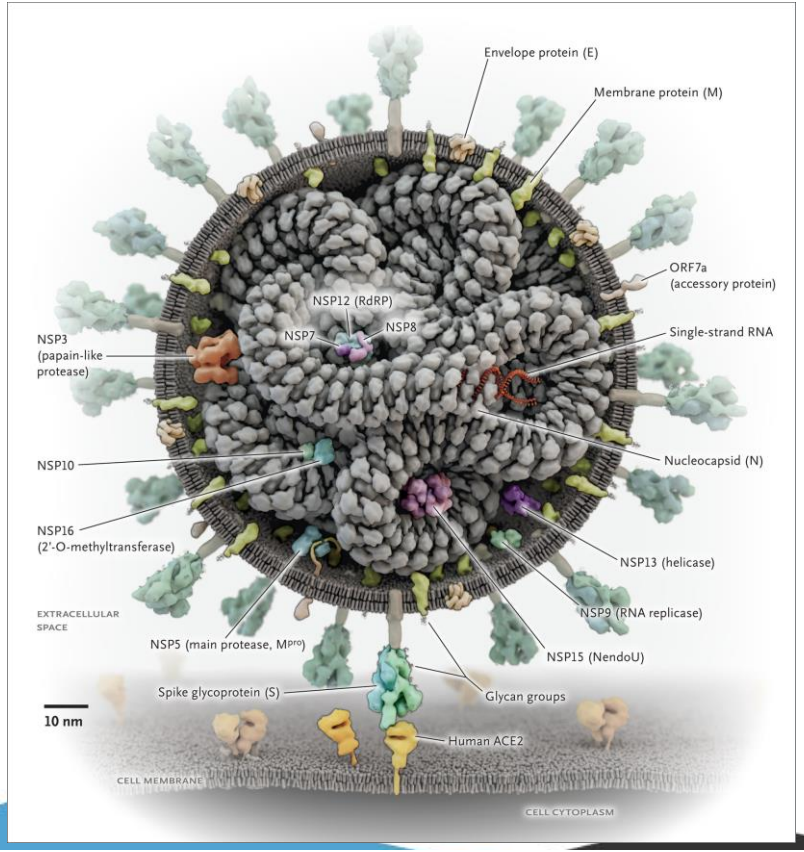
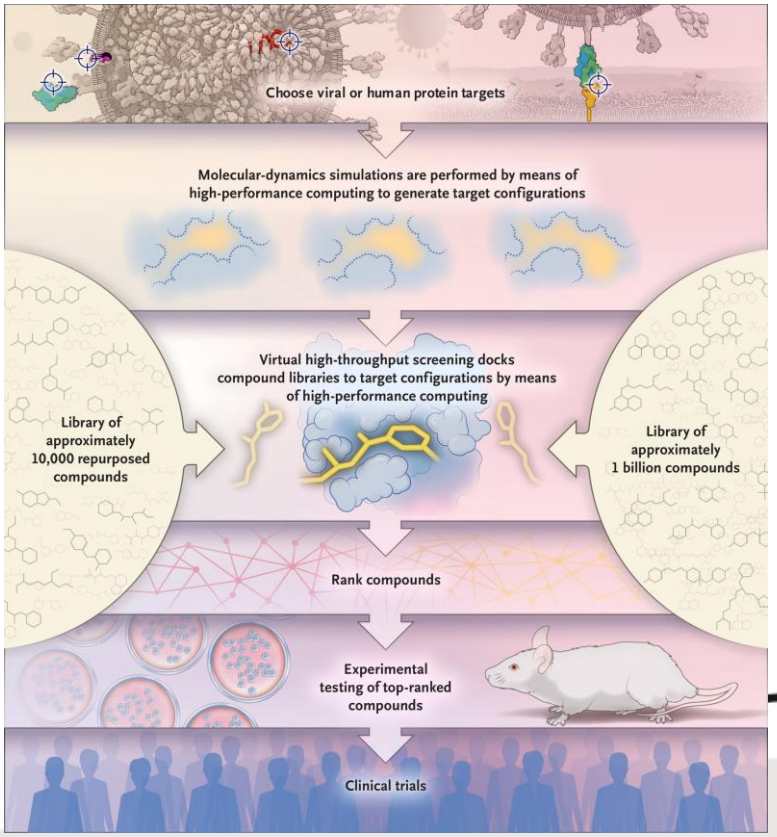


Figure 2. Main virtual screening workflow and binding poses of top two ligands. (**A**) Virtual screening workflow. (**B**) Binding pose of CP-1. (**C**) Binding pose of CP-2.

HTVS campaign of commercially available small-molecule ligands





Parks, J. M. & Smith, J. C. **How to Discover Antiviral Drugs Quickly.** *New England Journal of Medicine* (2020).

Peptide binder design

Substrate Channeling *via* a Transient Protein-Protein Complex: The case of D-Glyceraldehyde-3-Phosphate Dehydrogenase and L-Lactate Dehydrogenase

Željko M. Svedružić, Ivica Odorčić, Christopher H. Chang, Draženka Svedružić

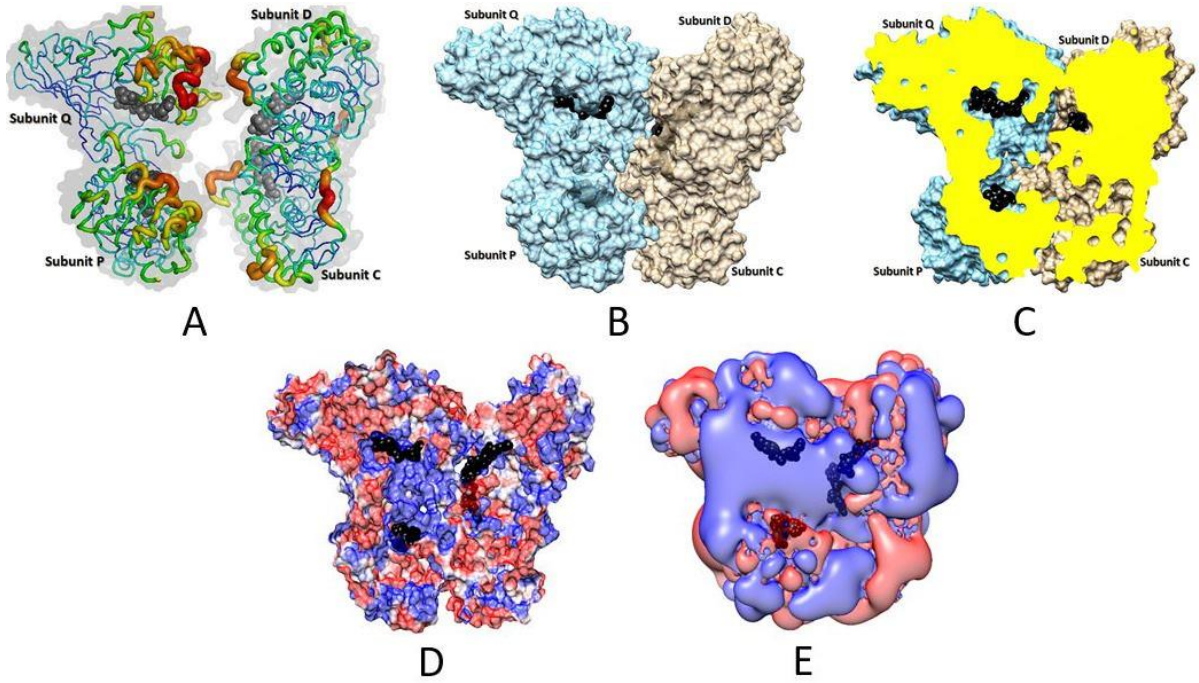
doi: https://doi.org/10.1101/2020.01.22.916023

Abstract | Full Text | Info/History | Metrics | Preview PDF

Abstract

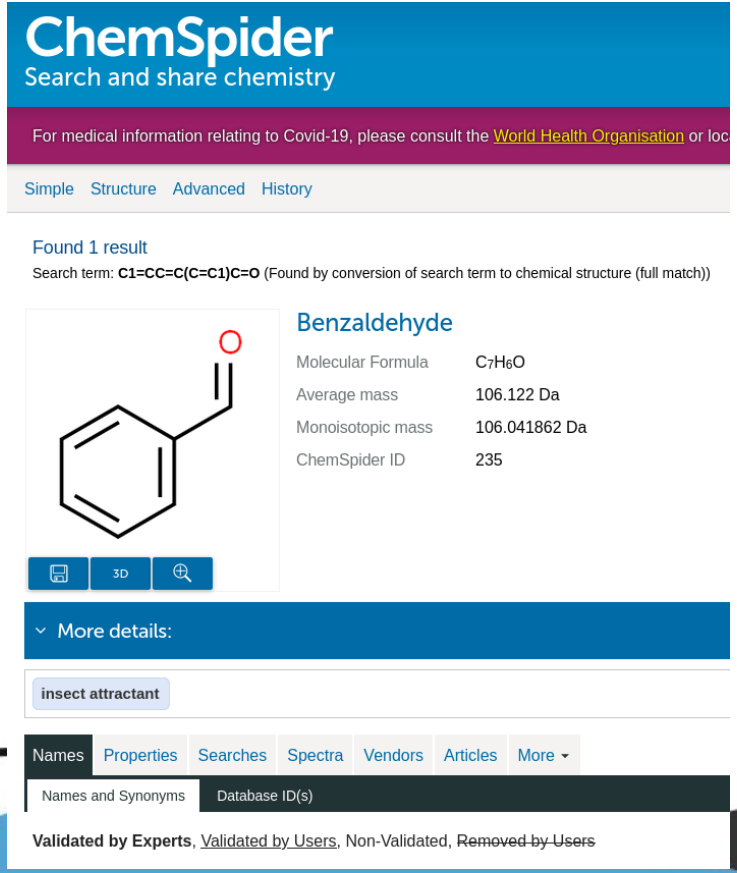
Background D-Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and L-lactate dehydrogenase (LDH) can form a complex that can regulate the major metabolic pathways, however, the exact mechanism remains unknown. We analyzed a possibility of NADH-channeling from GAPDH-NADH complex to LDH isozymes using enzymes from different cells.

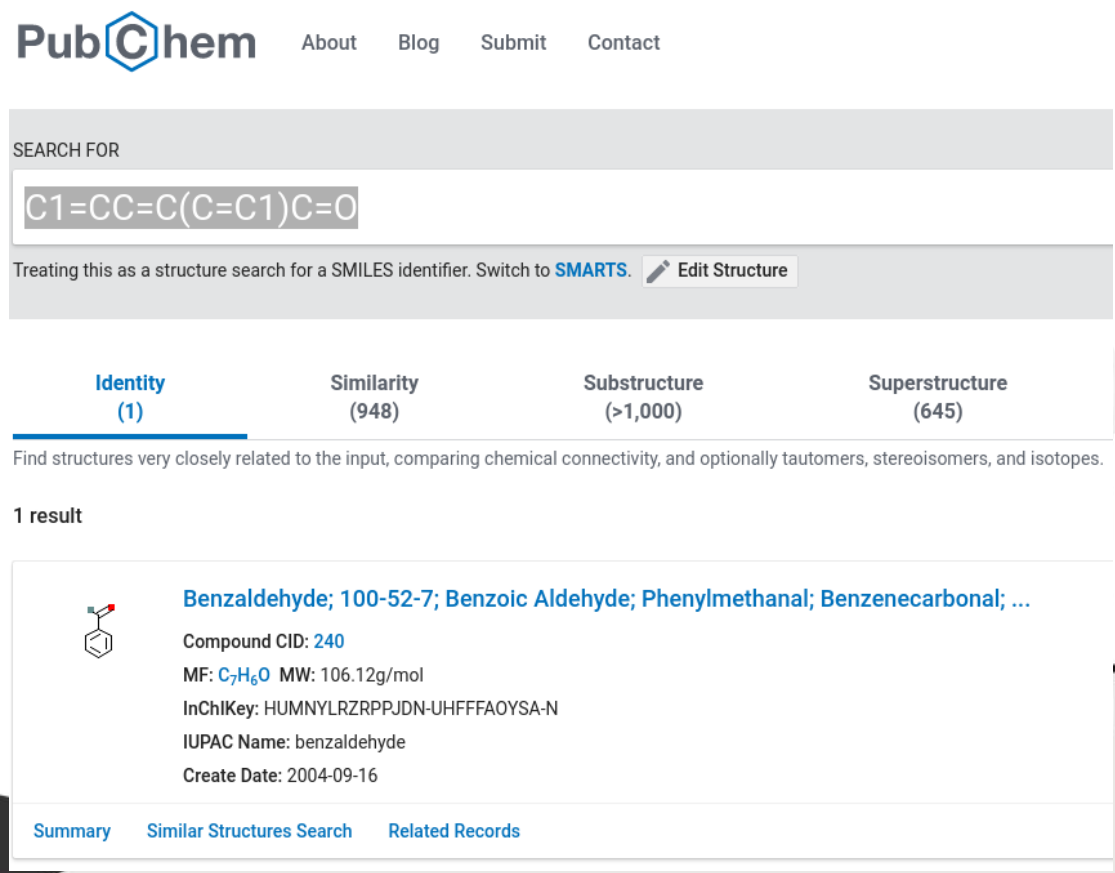
Results Enzyme-kinetics and NADH-binding studies showed that LDH can use GAPDH-NADH complex as a substrate. LDH activity with GAPDH-NADH complex was challenged with anti-LDH antibodies to show that the channeled and the diffusive reactions always take place in parallel. The channeling path is dominant only in assays with limiting free-NADH concertation that mimic cytosolic conditions. Analytical ultracentrifugation showed that the channeling does not require a high affinity complex. Molecular dynamics calculations and protein-protein interaction studies showed that LDH and GAPDH can form a leaky channeling complex only at



(Accepted for publication in *Nature Scientific Reports* on 21 May 2020)

Design and development of the compound database





Design and development of the docking server

Protein selection

Upload PDB or MOL2

Cavity mapping

Cavity mapping method

- ○ Two sphere method
- ◉ Reference ligand method

Reference ligand method

Upload SDF or MDL

Ligand docking

Ligand source

- ○ Enter molecules as text
- ◉ Upload a file containing molecules

Upload a file containing molecules

Upload SMI, SMILES, SDF, or MDL

Input format

- ○ SMILES
- ◉ SDF

Run docking

Deployment

- Bura supercomputer (existing equipment)
- 3 dedicated servers, rack, and switch (new equipment)
- 6 GPU compute nodes (new equipment)

Scientific infrastructure impact

- Reusable software, replicable results
  - The database and the web server will be maintained for at least 2 years after the project ends
  - All software developed in the project will be released under an open-source license
- Can be used in the future for any HTVS project by any research group on any (super)computer

General impact

- Viral pandemics regularly occurred during the 21st century
  - SARS epidemic (2002-2003)
  - Swine flu epidemic (2009-2010)
  - Middle East respiratory syndrome outbreaks (2012-ongoing)
  - Current COVID-19 pandemic (2019-ongoing)
- Viruses will continue to mutate and therefore the new epidemics in the future are **imminent** and **unavoidable**
  - Novel vaccines cannot be developed prior to specific viral mutation
  - Part of the population will inevitably be affected by whichever viral strain emerges – a large number of patients in dire need of an efficient and specific drug for their health condition
  - Viral drug discovery remains one of the most important fields of medicinal chemistry and molecular biology

Conclusion

- This call for funding aims to fund projects that can in a short time
  - Collect the maximal possible amount of information to understand COVID-19 physiology
  - **Design drugs specific for SARS-CoV-2 targets**
- Our group can address the drug design challenge in a fast and efficient manner using
  - The high-performance computing hardware
  - A large set of cutting-edge open-source software tools
  - Commercial compound databases
  - A unique mix of complementary expertise in computational and medicinal biochemistry and informatics

ODJEL ZA INFORMATIKU
SVEUČILIŠTE U RIJECI

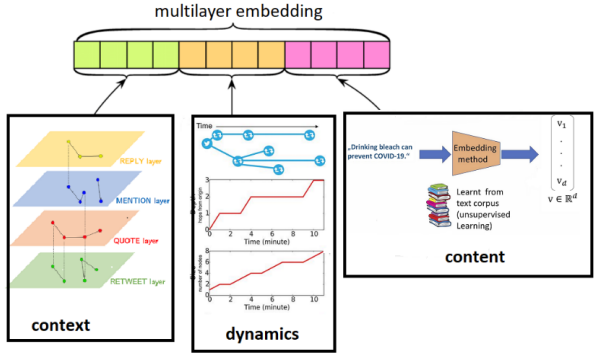
„Multilayer Framework for the Information Spreading Characterization in Social Media during the COVID-19 Crisis”

Assoc. Prof. Dr. Ana Meštrović
amestrovic@inf.uniri.hr

ICT-20 @ COVID-19 online conference, MIPRO, 27 May 2020

Outline

- Introduction
- Motivation and COVID-19 chalenges
- Problem description and related work
- Our approach: **multilayer framework**
- Conclusion
- Project team



Introduction

- **Fake news** refers to rumors, misinformation, disinformation or mal-information, half-truth information, conspiracy theories, etc.
- Fake news spreading in social media has been a huge problem during COVID-19 pandemic: dangerous, harmful for the society
- Many public authorities try to prevent and stop fake news spreading: government organizations, scientific community, etc. – by giving scientific explanations related to the COVID-19 rumors

WHO: Myth busters

While several drug trials are ongoing, there is currently no proof that hydroxychloroquine or any other drug can cure or prevent COVID-19.
The misuse of hydroxychloroquine can cause serious side effects and illness and even lead to death.
WHO is coordinating efforts to develop and evaluate medicines to treat COVID-19.

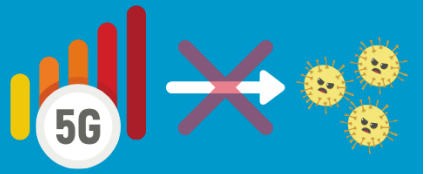
FACT:
There are currently no drugs licensed for the treatment or prevention of COVID-19

World Health Organization #Coronavirus #COVID19

27 April 2020

Viruses cannot travel on radio waves/mobile networks.
COVID-19 is spreading in many countries that do not have 5G mobile networks.
COVID-19 is spread through respiratory droplets when an infected person coughs, sneezes or speaks.
People can also be infected by touching a contaminated surface and then their eyes, mouth or nose.

FACT:
5G mobile networks DO NOT spread COVID-19



World Health Organization #Coronavirus #COVID19

8 April 2020

Hot peppers in your food, though very tasty, cannot prevent or cure COVID-19.
The best way to protect yourself against the new coronavirus is to keep at least 1 metre away from others and to wash your hands frequently and thoroughly.
It is also beneficial for your general health to maintain a balanced diet, stay well hydrated, exercise regularly and sleep well.

FACT:
Adding pepper to your soup or other meals DOES NOT prevent or cure COVID-19.

World Health Organization #Coronavirus #COVID19

27 April 2020

Methanol, ethanol, and bleach are poisons. Drinking them can lead to disability and death.
Methanol, ethanol and bleach are sometimes used in cleaning products to kill the virus on surfaces – however you should never drink them. They will not kill the virus in your body and they will harm your internal organs.

To protect yourself against COVID-19, disinfect objects and surfaces, especially the ones you touch regularly. You can use diluted bleach or alcohol for that. Make sure you clean your hands frequently and thoroughly and avoid touching your eyes, mouth and nose.

FACT:
Drinking methanol, ethanol or bleach DOES NOT prevent or cure COVID-19 and can be extremely dangerous

World Health Organization #COVID19 #coronavirus

5 April 2020

Project goals

- **AI approach**: **automatic detection of fake news** spreading in social media in the context of COVID-19 padnemic
- Project tasks and focus:
  - to perform quantitative and qualitative analysis of textual information related to COVID-19 in social media
  - to identify which characteristics of information spreading can differentiate between various types of information spreading
- There is a large number of scientific papers and studies of fake news:
  - more that 50.000 published papers
  - exponential growth

COVID-19 Chalenges

- COVID-19 crisis brings a whole new realm of challenges in terms of large communication volumes that results with:
  - massive datasets,
  - new terminology,
  - new aspects and new specific topics:
    - mortality rate, case fatality rate,
    - exponential curves, the pandemic spreading data,
    - healthcare issues,
    - government policies,
    - various restrictions, quaratntene, lock down,
    - and other socio-economic issues related to the pandemic, etc...

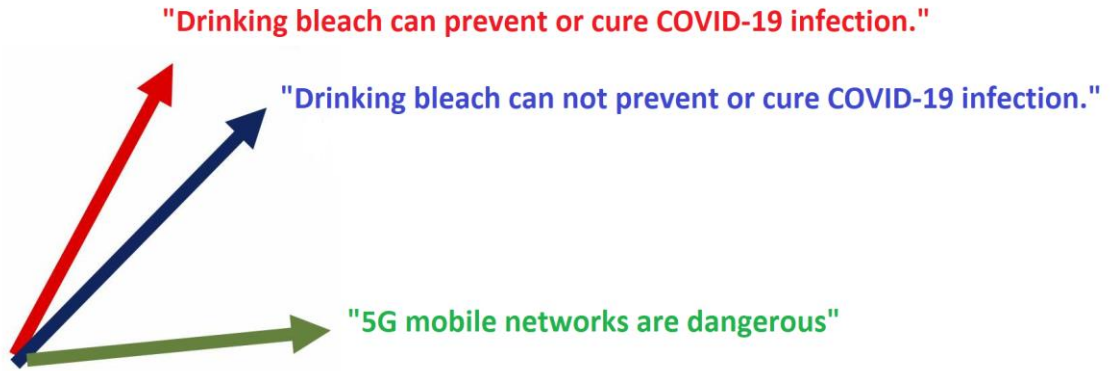
Task description

How to automatically detect spreading of fake news/fake information?

- **Information**: text message (post/**tweet**/comment, ...)
- Binary **classification task**: fake news, true news
- AI Approaches: machine learning algorithms, neural networks, ...
- **Manually annotated datasets** (fake/true): learning and evaluation
- **Features**: numerical representation as vectors in n-dimensional space
  - text (**content**),
  - user/social network based features (**context**),
  - spreading during the time (**dynamics**).

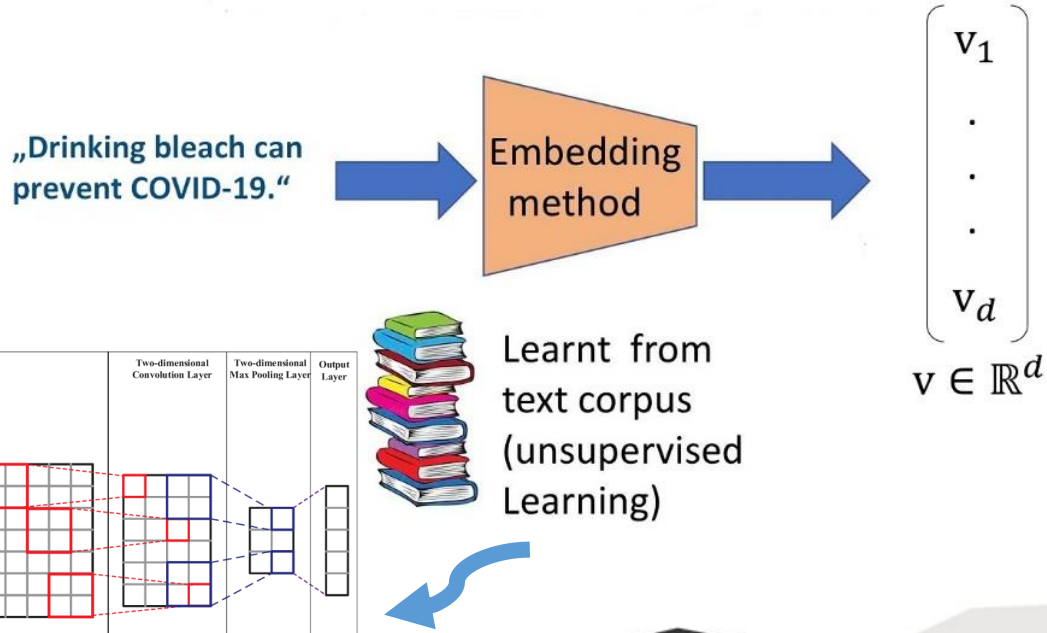
Text classification task

- Traditional approaches: vectors of sentences/textual documents represented as vectors in n-dimensional vector space
- Various weighting schemes: word frequencies, TF-IDF, ...
- Cosine similarity
- Problems:
  - High Dimensionality
  - Vocabulary problems
  - ...



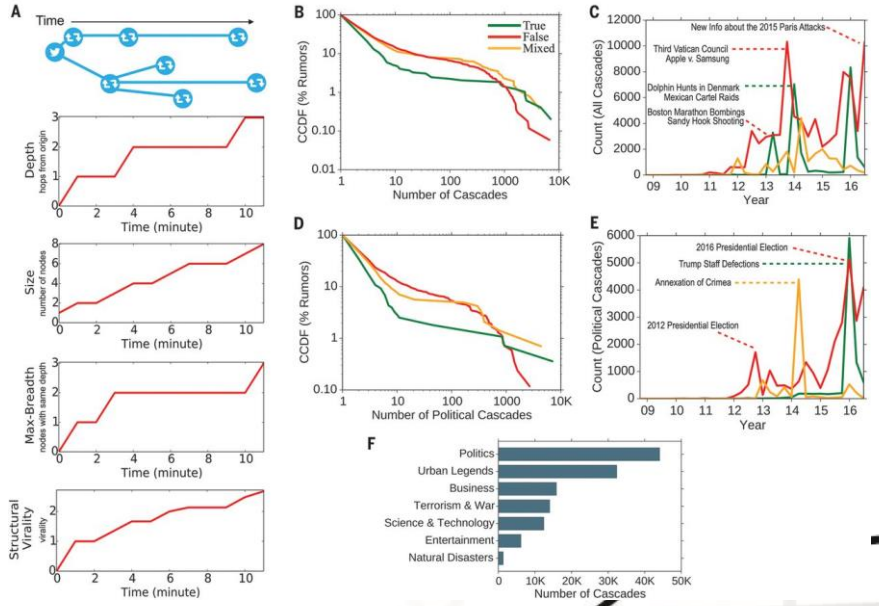
Text classification task

- Recent trends: deep learning (word/sentence embeddings)



Related work

- **Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The spread of true and false news online." *Science* 359, no. 6380 (2018): 1146-1151.**
- They investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017
- ~126,000 stories tweeted by ~3 million people
- False news spread significantly **farther, faster, deeper, and more broadly** than the truth in all categories of information



Related work

- **Ruchansky, Natali, Sungyong Seo, and Yan Liu. "Csi: A hybrid deep model for fake news detection." In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 797-806. 2017.**
- They combine two set of features of an article to detect fake news:
  - the temporal patern of user activity on a given article (temporal features)
  - the behavior of users (user features)
- They use neural networks and Adam algorithm
- They show that user features improve the classification

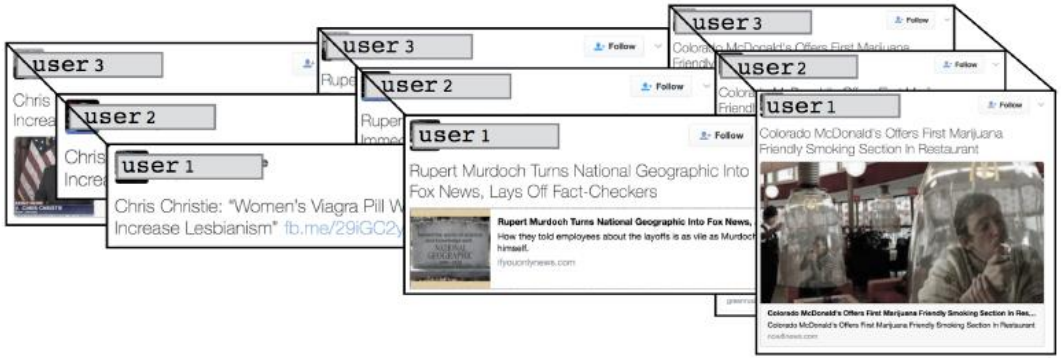
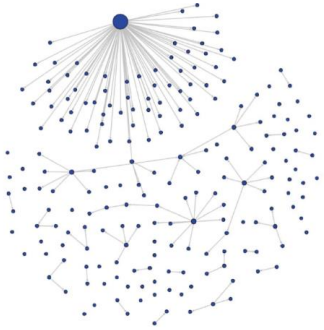


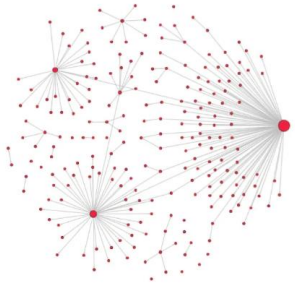
Figure 1: A group of Twitter accounts who shared the same set of fake articles.

Related work

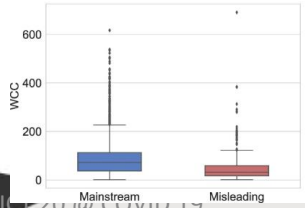
- **Pierri, Francesco, Carlo Piccardi, and Stefano Ceri. "Topology comparison of Twitter diffusion networks effectively reveals misleading information."** ***Scientific reports*** **10, no. 1 (2020): 1-9.**
- Network based approach
- Various network properties
- They show that there are differences between fake and true information spreading in social networks

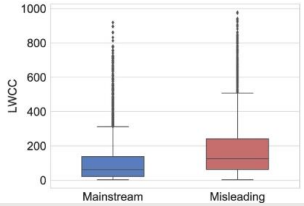


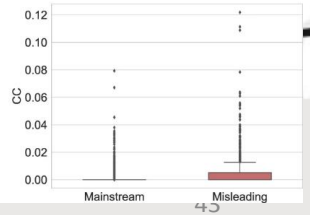
Features:
WCC = 70 **LWCC** = 72 **CC** = 0
DWCC = 3 **SCC** = 205 **LSCC** = 1 **KC** = 1



Features:
WCC = 21 **LWCC** = 178 **CC** = 0.03
DWCC = 5 **SCC** = 220 **LSCC** = 2 **KC** = 2





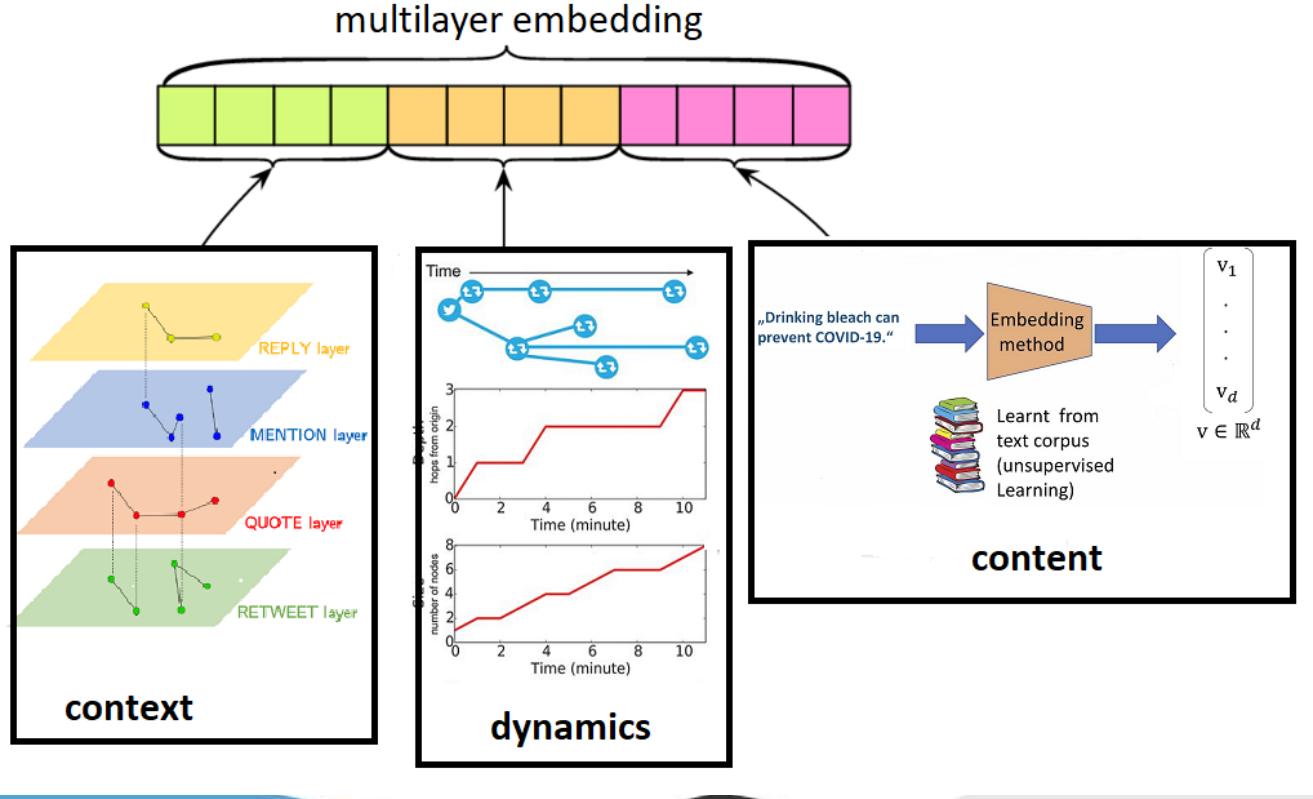


Our approach: **multilayer framework**

- Multilayer framework – integrate three sets of features

1. CONTENT
   - senetence embeddings
2. CONTEXT
   - user properties provided as multilayer network properties on the global, middle and local scale
3. DYNAMICS
   - cascade dynamics, information changes over time

Our approach: **multilayer framework**



Our approach: **multilayer framework**

- **Phase 1 – Data and Framework Setup**
  - **Multilayer framework definition**
  - **COVID-19 News Article Corpora**: crawling news related to COVID-19, comments and other metadata from
    - Croatian newspaper sites: index.hr, vecernji.hr, 24sata.hr, etc.
    - crawling data from the Croatian government and other official web sites related to COVID-19: koronavirus.hr, hzjz.hr
  - **COVID-19 Social Networks Messages dataset**: crawling publicly available data (e.g. messages, posts, comments, tweets, retweets, etc.) from the social networks (e.g. Twitter, Facebook) related to COVID-19 communication in the Croatian language.

Our approach: **multilayer framework**

- **Phase 2 – Content-based Analysis**
  - **Language model development and evaluation**: activities related to training word/sentence embeddings using deep learning models (BERT model)
  - **Content-based analysis**: activities related to various NLP tasks such as keyword extraction, articles/messages classification, etc.
  - **Statistical analysis of the COVID-19 corpus and datasets**
- **Phase 3 – Multilayer Network Analysis**
  - **Multilayer network construction**
  - **Context-based analysis**: calculating the network measures on the global, middle and local network scale for various networks
  - **Dynamic-based analysis**: analysis of the cascade effects and network dynamics

Our approach: **multilayer framework**

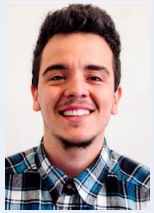
- **Phase 4 –fake news detection model developement**
  - Dataset pipeline building
  - **Model development**: initial experiments and performance tests, parameter tweaking, revised model development and experiments,
  - **Model evaluation** (accuracy, precision, recall, F-measures, ROC, AUC curves)

- **Phase 5 – Dissemination**

Conclusion

- Fake news detection during the COVID-19 pandemic – chalenging task
- Large datasets require powerful supercomputers
- We will integrate knowledge from two domanis:
  - social networks analysis
  - natural language processing
- More precisely, we will apply:
  - network based measures, multilayer networks
  - statistical analsisys of the spreading dynamic
  - deep learning, neural networks

Project team

	**Project leader:** Assoc. professor **Ana Meštrović**, PhD Expertise in: NLP, Complex networks		**Karlo Babić**, MSc Doctoral student NLP
	Professor **Sanda Martinčić-Ipšić**, PhD Expertise in: NLP, Complex networks		**Milan Petrović**, MSc Doctoral student Social network analysis, NLP
	**Slobodan Beliga**, PhD Postdoctoral student Expertise in: NLP, Complex networks		Assoc. Prof. **Mihaela Matešić,** PhD, Faculty of Humanities and Social Sciences in Rijeka, University of Rijeka, Croatia Expertise in: Linguistics

Experts outside University of Rijeka:
Zoran Levnajić, PhD, Faculty of Information Studies, Slovenia
Marko Pranjić, MSc, Trikoder d.o.o., Croatia

Questions