Science is fun but not only: career perspectives for young talented researchers in a modern society

Igor V. Tetko

Helmholtz Munich and BIGCHEM GmbH

July 11, 2023, On-line Zoom

HELMHOLTZ MUNICH

big chem

Agenda

- Undergraduate
- PhD
  - Marie Skłodowska-Curie program
- Translational studies

Undergraduate

Initial career opportunities

- HiWi (Research Assistant)
- Short research projects
- Erasmus

HiWi (Research Assistant)

Roland Arnold · 1st
Principal Investigator/Group Leader. Institute of Cancer and Genomic Sciences, University of Birmingham

University of Birmingham

United Kingdom · Contact info

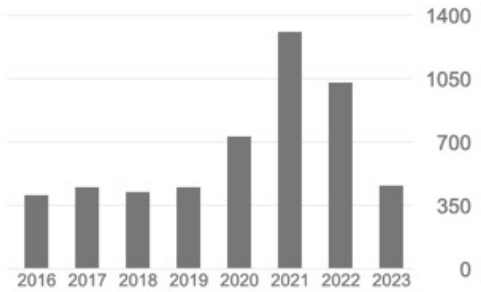
419 connections

Dr. Olivia Prazeres da Costa, PMP, Philipp Pagel, and 26 other mutual connections

Message More

HiWi 2001 – 2006
PhD 2007 – 2010

	All	Since 2018
Cited by		VIEW ALL
Citations	7531	4426
h-index	35	27
i10-index	47	41



SIMAP—structuring the network of protein similarities T Rattei, P Tischler, R Arnold, F Hamberger, J Krebs, J Krumsiek, ... Nucleic acids research 36 (suppl_1), D289-D292	39	2007
Gepard: a rapid and sensitive tool for creating dotplots on genome scale J Krumsiek, R Arnold, T Rattei Bioinformatics 23 (8), 1026-1028	707	2007
Using public resource computing and systematic pre-calculation for large scale sequence analysis T Rattei, M Walter, R Arnold, DP Anderson, W Mewes International Workshop on Grid Computing in Computational Biology, 11-18	3	2007
SIMAP: the similarity matrix of proteins T Rattei, R Arnold, P Tischler, D Lindner, V Stümpflen, HW Mewes Nucleic Acids Research 34 (suppl_1), D252-D256	59	2006
SIMAP—the similarity matrix of proteins R Arnold, T Rattei, P Tischler, MD Truong, V Stümpflen, W Mewes Bioinformatics 21 (suppl_2), ii42-ii46	52	2005
MIPS: analysis and annotation of proteins from whole genomes HW Mewes, C Amid, R Arnold, D Frishman, U Güldener, G Mannhaupt, ... Nucleic acids research 32 (suppl_1), D41-D44	749	2004
MIPS Arabidopsis thaliana Database (MAtDB): an integrated biological knowledge resource based on the first complete plant genome H Schoof, P Zaccaria, H Gundlach, K Lemcke, S Rudd, G Kolesov, ... Nucleic Acids Research 30 (1), 91-93	211	2002

HiWi (Research Assistant)

Susann Bader (She/Her) · 1st

Scientific Software Engineer @ Genedata

Bavaria, Germany · Contact info

214 connections

Yurii Sushko, Sergii Novotarsk

Message More

Genedata

Ludwig-Maximilians Universität München

Susann Bader (geb Vorberg)

FOLLOW

Postdoc Bioinformatics

Verified email at genzentrum.lmu.de

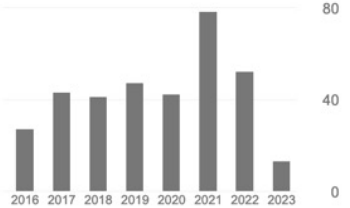
structural bioinformatics machine learning statistical modelling protein structure prediction

TITLE	CITED BY	YEAR
LocTree3 prediction of localization T Goldberg, M Hecht, T Hamp, T Karl, G Yachdav, N Ahmed, U Altermann, ... Nucleic acids research 42 (W1), W350-W355	285	2014
Modeling the biodegradability of chemical compounds using the online CHEmical modeling environment (OCHEM) S Vorberg, IV Tetko Molecular Informatics 33 (1), 73-85	67	2014
Synthetic protein alignments by CCMgen quantify noise in residue-residue contact prediction S Vorberg, S Seemayer, J Söding PLoS computational biology 14 (11), e1006526	20	2018

GET MY OWN PROFILE

Cited by VIEW ALL

	All	Since 2018
Citations	376	273
h-index	4	4
i10-index	3	3



Short Research Projects

Caroline Friedel · 1st
Associate Professor at Ludwig-Maximilians-Universität (LMU) München
Greater Munich Metropolitan Area · **Contact info**

Ludwig-Maximilians-Universität (LMU) München

Prof. Dr. Katharina Jahn

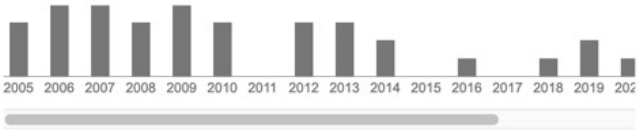
Freie Universität Berlin

Institut für Informatik
Biomedical Data Science
Professor

Address Takustr. 9
Room 009
14195 Berlin

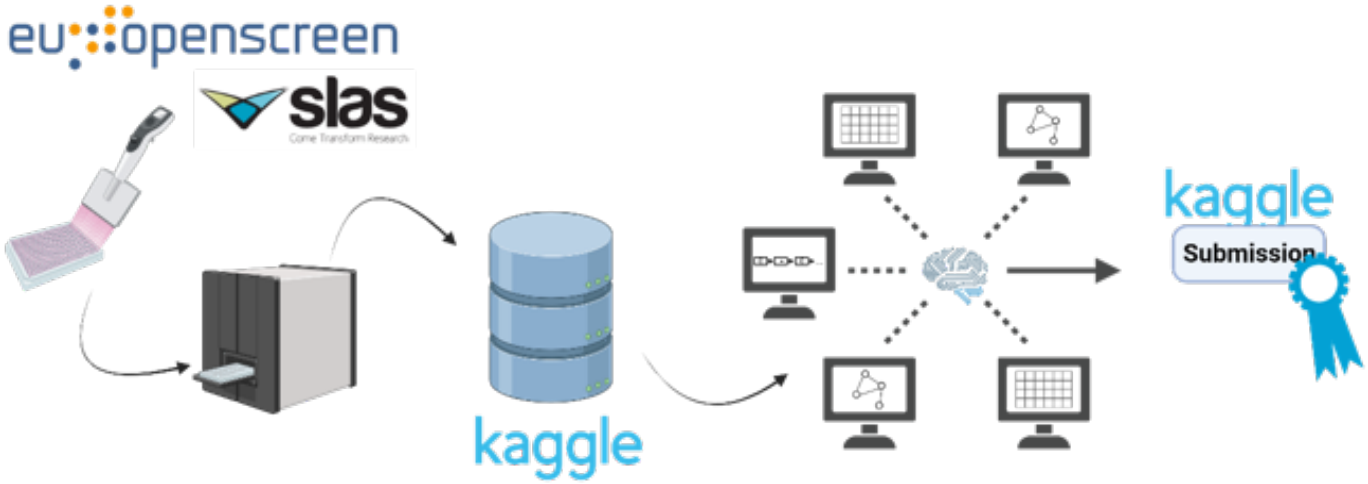
Support vector machines for separation of mixed plant–pathogen EST collections based on codon usage

Authors	Caroline C Friedel, Katharina HV Jahn, Selina Sommer, Stephen Rudd, Hans W Mewes, Igor V Tetko
Publication date	2005/4/15
Journal	Bioinformatics
Volume	21
Issue	8
Pages	1383-1388
Publisher	Oxford University Press
Description	**Motivation:** Discovery of host and pathogen genes expressed at the plant–pathogen interface often requires the construction of mixed libraries that contain sequences from both genomes. Sequence identification requires high-throughput and reliable classification of genome origin. When using single-pass cDNA sequences difficulties arise from the short sequence length, the lack of sufficient taxonomically relevant sequence data in public databases and ambiguous sequence homology between plant and pathogen genes. **Results:** A novel method is described, which is independent of the availability of homologous genes and relies on subtle differences in codon usage between plant and fungal genes. We used support vector machines (SVMs) to identify the probable origin of sequences. SVMs were compared to several other machine learning techniques and to a probabilistic algorithm (PF …
Total citations	Cited by 36

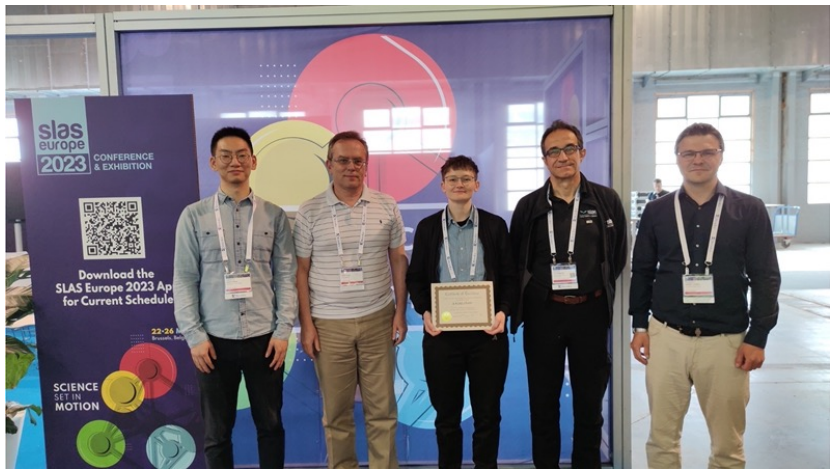


BSc and MSc projects

OCHEM consensus model wins Kaggle solubility challenge



Ms A. Kopp, BSc student at LMU



Erasmus (*European Community Action Scheme for the Mobility of University Students*)

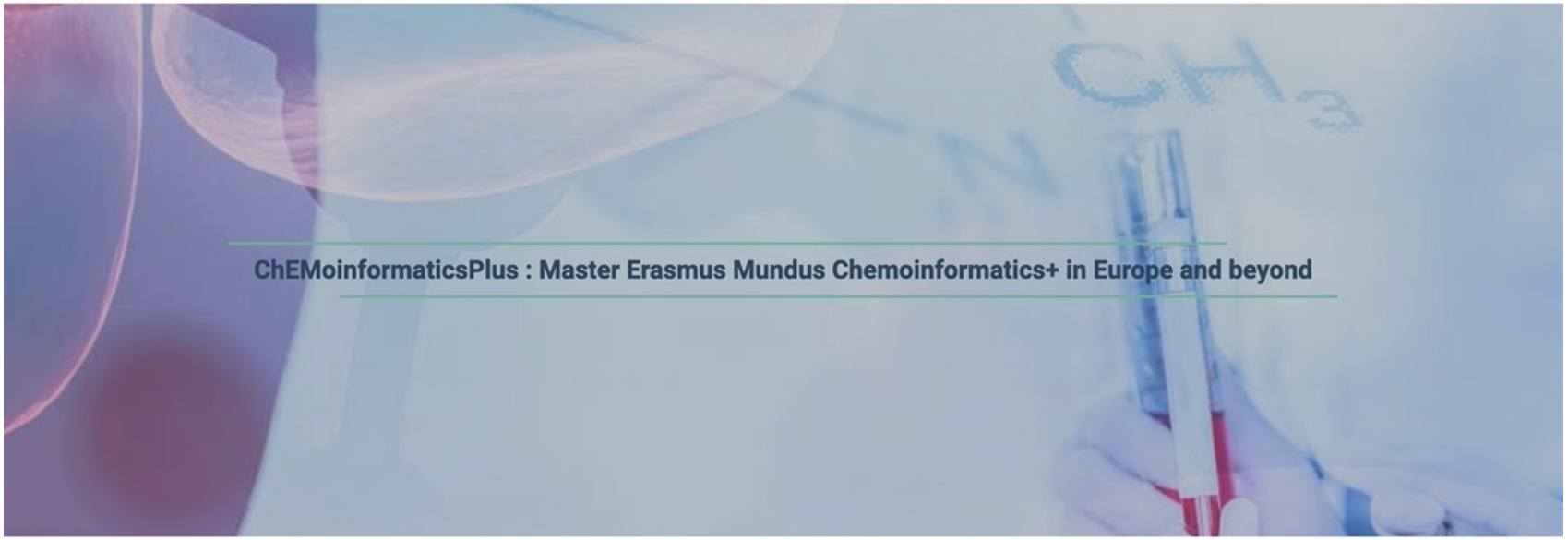
5600 organisation from world
38 countries

Erasmus Charter for Higher Education

https://erasmus-plus.ec.europa.eu

Erasmus of Rotterdam or **Erasmus (**1466 –1536) was a Dutch philosopher and Catholic theologian who is considered one of the greatest scholars of the Northern Renaissance.

Source of images: wiki and Erasmus web site



ChEMoinformaticsPlus: the Erasmus Mundus Master degree in Chemoinformatics across Europe and beyond

7 universities are involved in the two years *Erasmus Mundus Joint Master's Degree ChEMoinformaticsPlus* that will open to its first student's edition in **September 2022**. The Erasmus Mundus label is awarded by the European commission to transnational Master programs that contribute to excellence and internalization of Higher Education in Europe.

Chemoinformatics is a major discipline in theoretical chemistry, using artificial intelligence and data sciences to tackle current social and innovation challenges in Chemistry. It concerns the development, creation, organization, storage, dissemination, analysis, visualization and use of chemical information. Within the last 20 years, Chemoinformatics has become a corner stone for the chemical Industry:

"The use of artificial intelligence/ machine learning technologies has become a critical part in the drug discovery processes and chemoinformatics plays an important role in integrating AI/ML approaches into the field of medical chemistry." (Astellas Pharma Inc.),

PhD

PhD position announcements (external)

- https://euraxess.ec.europa.eu
- LinkedIn groups, e.g.
  - **Jobs in computational chemistry, computational drug design, cheminformatics & molecular modelling**
  - **Master Chemoinformatics (MSc)**
- http://ccl.net/chemistry/announcements/jobs/
- https://www.dfg.de/en/research_funding
- Nature, Science job announcements
- ResearchGate, Twitter
- Mailing lists, e.g., connectionists@cs.cmu.edu

Marie Skłodowska-Curie Actions 2021-2027

Developing talents, advancing research

Under Pillar I of Horizon Europe, the MSCA are the European Union's reference programme for doctoral education and postdoctoral training. They support researchers from all over the world, at all stages of their careers, with a focus on their training, skills and career development.

Under Horizon 2020 (2014-2020), the MSCA:

Funded **1080 doctoral programmes**, of which 156 industrial doctoral programmes and 76 joint doctorates

Involved **4 700 companies**, of which **2 200 SMEs**

Involved **37% of researchers** from non-EU countries and around **1300 organisations** from non-EU and non-associated countries

Since 1996

budget
14 billion €

researchers
140 000
(39 000 PhDs)

Horizon Europe (2021-2027)

budget
6.6 billion €

researchers
65 000
(25 000 PhDs)

Under Horizon Europe, the MSCA will:

Strengthen organisations
The MSCA support excellent doctoral and postdoctoral programmes and collaborative projects worldwide, promoting structuring impact on organisations

Foster research and innovation beyond academia
The MSCA boost ties between academia and other non-academic organisations with various incentives, increasing fellows' exposure to other sectors

Build international links
The MSCA are key in attracting talent to Europe, building international, strategic partnerships, and promoting global research mobility and science cooperation

The MSCA have 5 main actions

Postdoctoral Fellowships

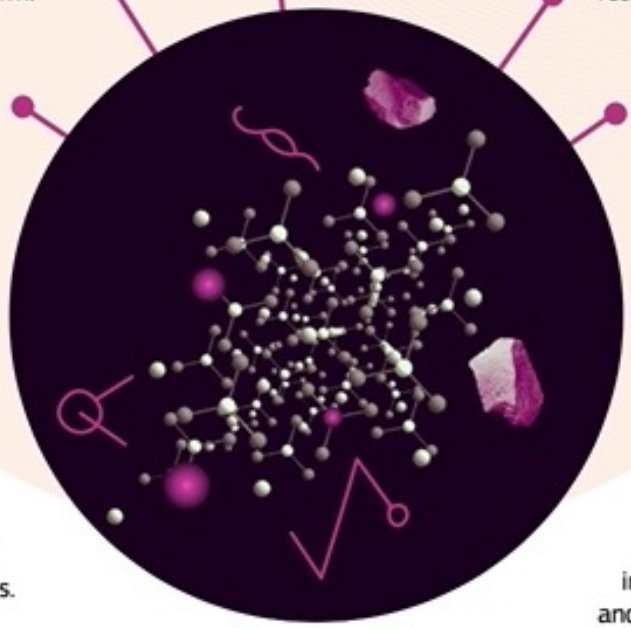
support researchers' careers and foster excellence in research and innovation. Researchers holding a PhD can carry out their research activities, acquire new skills and develop their careers abroad, whilst developing competences in non-academic sectors and working within interdisciplinary teams.

Doctoral Networks

implement doctoral programmes (including joint doctorates and industrial doctorates) **by international partnerships** of organisations from different sectors. They train highly-skilled doctoral candidates, stimulate their creativity, enhance their innovation capacities and boost their employability in the long-term.

Staff Exchanges

encourage short-term international and inter-sectoral exchanges of research and innovation staff through sustainable, collaborative projects in Europe and beyond. By doing so, they enhance knowledge and skills transfer and increase organisations' research and innovation capacities.



MSCA and Citizens

brings research and researchers closer to children, families and the public at large through the European Researchers' Night - the annual research communication and promotion event taking place at the end of September across EU Member States and Horizon Europe Associated Countries.

COFUND

co-finances regional, national and international doctoral and postdoctoral programmes for researchers' training and career development. The COFUND action spreads MSCA's best practices by setting high standards and excellent working conditions, and boosts training and international, interdisciplinary and inter-sectoral mobility.

Benefits of MSCA:

for organisations

- High-quality research training and supervision offered
- Build up strong research and innovation partnerships
- Strengthened research capacity
- Improved human resources and working conditions to attract the best researchers
- Sustainable knowledge transfer and new international and inter-sectoral collaborations
- Enhanced global visibility and attractiveness

for researchers

- New knowledge and skills in and outside academia
- Increased career prospects and employability
- Innovation-oriented mind-set, to convert ideas into products and services
- Networking and increased visibility in the European R&I community
- International, interdisciplinary and inter-sectoral experience and exposure
- Access to leading organisations and their teams

for European Research and Innovation

- Scientific excellence promoted in all countries
- Attract and retain talents in Europe
- High quality R&I for Europe's sustainable growth
- Increased strategic cooperation and brain circulation between countries, disciplines and sectors
- New links between research, industry and society
- Stronger European Research Area (ERA)

Nobel price to former MSCA fellow

On 7 October 2020, the Royal Swedish Academy of Science awarded the Nobel Prize in Chemistry to Emmanuelle Charpentier and Jennifer A. Doudna ‘for the development of a method for genome editing’.

Dr Charpentier is a former MSCA fellow and principal investigator involved in the training of young researchers in the field of genomics in the MSCA project ENLIGHT-TEN ITN.

MSCA ITN applications

- Strong scientific competitive program, success rate ca 10%
- Evaluation is done of the whole consortium not of the individual partner
- Complementarity of partners
- Strong training program (“Curriculum”)
- Strong training in complementarity skills

ECO http://ecoitn.eu

Environmental ChemOinformatics

ECO | NEWS | REACH | FELLOWS | SCHOOLS | PUBLICATIONS | CONTACT

News

SITE STRUCTURE

- ECO
  - Description
    - General information
    - Synergism
    - Methodology
    - Dissemination
    - Abbreviations
  - Training
  - Partners
  - Projects involved
- News
  - Archive
- REACH
- Fellows
- Schools
  - Final Conference 2013
  - Winter School 2013
  - Summer School 2012
  - Winter School 2012
  - Autumn School 2011
  - Winter School 2011
  - Autumn School 2010
- Publications
- Contact

ECO News and Events

31 PhDs were awarded to the former ECO fellows. See the impact of ECO ITN publications at Google Scholar.

Saturday, 7 March 2020

10 doctoral (PhD) positions in Big Data Analysis in Chemistry, Marie Skodowska-Curie ITN BIGCHEM (http://bigchem.eu)

BIGCHEM (BIG data in CHEMistry) is a Marie Skodowska-Curie Innovative Training Network (ITN) for Early Stage Researchers (ESR) funded by the European Commission under the H2020 Programme. The BIGCHEM ITN will provide a comprehensive and cross-disciplinary structured curriculum for doctoral students in large chemical data analysis using machine-learning, computational chemistry and chemoinformatics methods. The innovative research program will be implemented with the target users, large pharma companies and SMEs, which generate and analyse large chemical data.

Tuesday, 12 February 2016

ECO publication received 2016 SLAS ReadersChoice Awards

On January 26 2016, SLAS announced that article of Schorpp, K. et al Identification of Small-Molecule Frequent Hitters from AlphaScreen High-Throughput Screens *J. Biomol. Screen. 2014, 19 715-726* received 2016 JBS Readers Choice Award. ECO fellow Mrs. Elena Salmina contributed to the chemoinformatics analysis of this study during her short-term fellowship in HMGU, group of Dr. Tetko.

Tuesday, 26 January 2016

19th successful PhD in ECO network

On December 9 2015, Dr. Alessandra Pirovano successfully defended her PhD thesis at the Radboud University. Dr. Prirovano was ECO fellow at the same University. The topic of her thesis is "Quantifying biotransformation of xenobiotics in mammals" under supervisor of Prof. dr. ir. A.J. Hendriks.

REACH

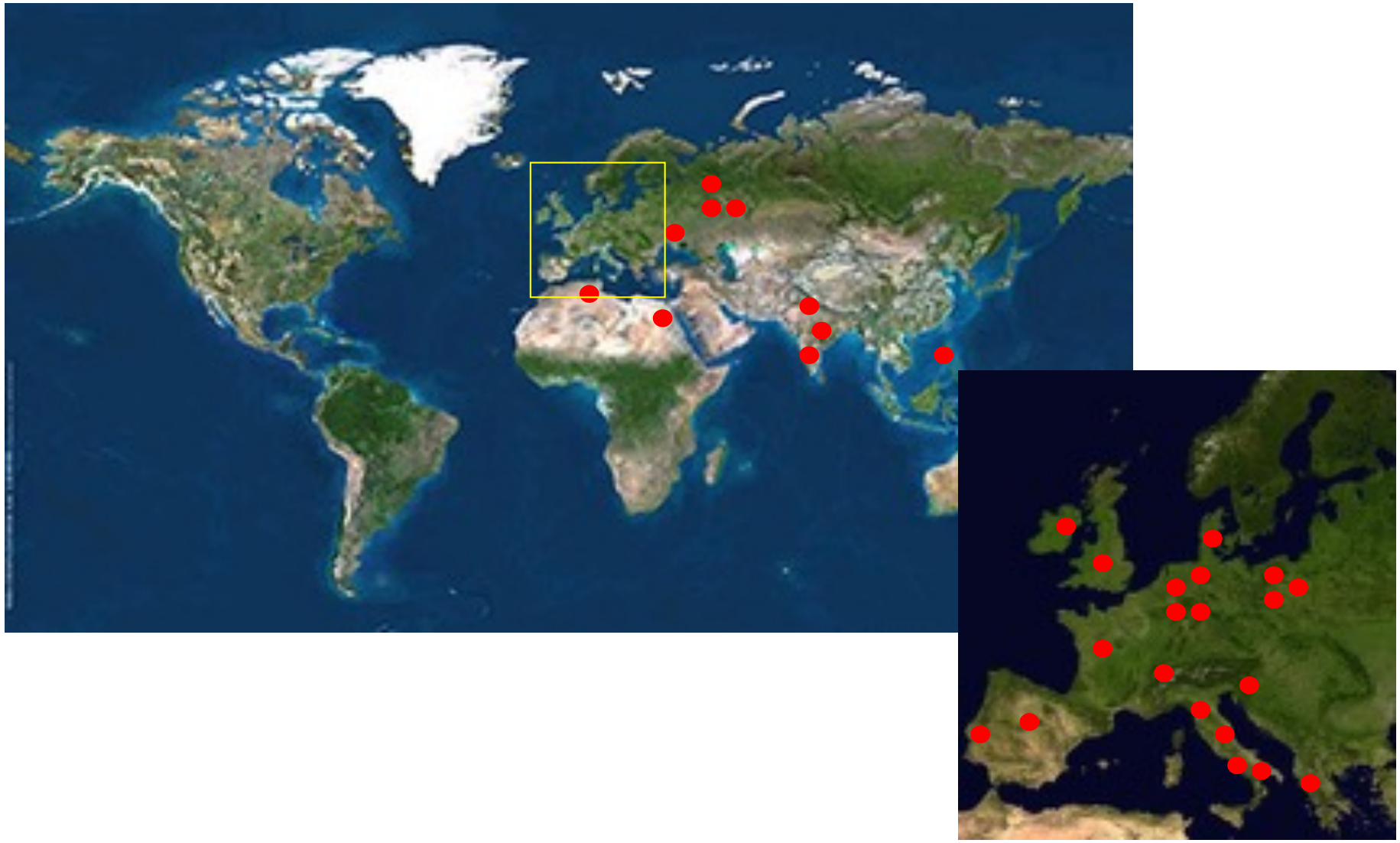
- **R**egistration, **E**valuation, **A**uthorisation and Restriction of **Ch**emical substances

European **Ch**emicals **A**gency (ECHA) in Helsinki

What is the REACH Timetable?

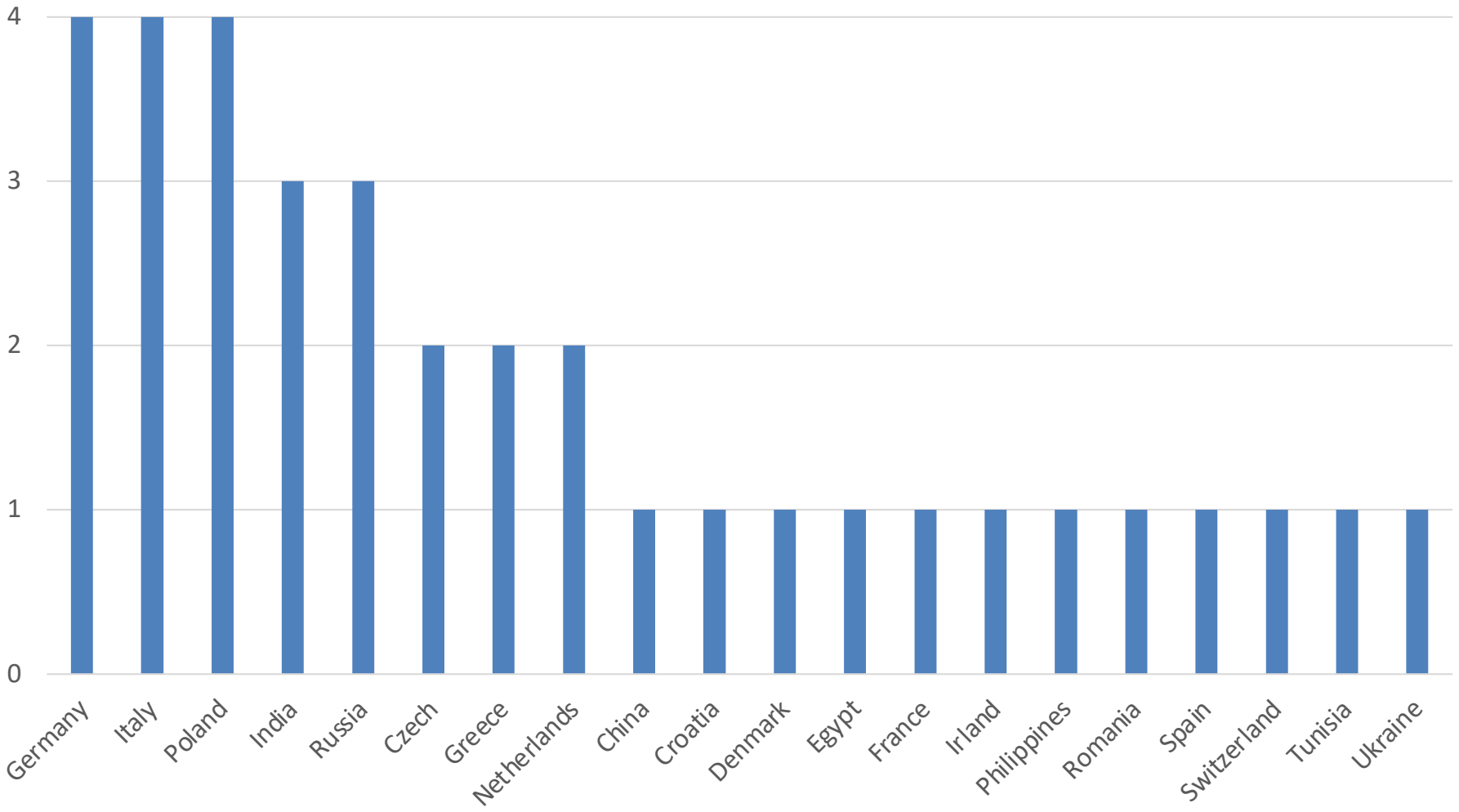
http://www.basf.com/group/corporate/en/sustainability/management-and-instruments/success-added-value

Origin of the fellows



Overall network

Origin of fellows per country



ECO http://ecoitn.eu

Environmental ChemOinformatics

ECO | NEWS | REACH | FELLOWS | SCHOOLS | PUBLICATIONS | CONTACT

News

SITE STRUCTURE

- ECO
  - Description
    - General information
    - Synergism
    - Methodology
    - Dissemination
    - Abbreviations
  - Training
  - Partners
  - Projects involved
- News
  - Archive
- REACH
- Fellows
- Schools
  - Final Conference 2013
  - Winter School 2013
  - Summer School 2012
  - Winter School 2012
  - Autumn School 2011
  - Winter School 2011
  - Autumn School 2010
- Publications
- Contact

ECO News and Events

31 PhDs were awarded to the former ECO fellows. See the impact of ECO ITN publications at Google Scholar.

Saturday, 7 March 2020

10 doctoral (PhD) positions in Big Data Analysis in Chemistry, Marie Skodowska-Curie ITN BIGCHEM (http://bigchem.eu)

BIGCHEM (BIG data in CHEMistry) is a Marie Skodowska-Curie Innovative Training Network (ITN) for Early Stage Researchers (ESR) funded by the European Commission under the H2020 Programme. The BIGCHEM ITN will provide a comprehensive and cross-disciplinary structured curriculum for doctoral students in large chemical data analysis using machine-learning, computational chemistry and chemoinformatics methods. The innovative research program will be implemented with the target users, large pharma companies and SMEs, which generate and analyse large chemical data.

Tuesday, 12 February 2016

ECO publication received 2016 SLAS ReadersChoice Awards

On January 26 2016, SLAS announced that article of Schorpp, K. et al Identification of Small-Molecule Frequent Hitters from AlphaScreen High-Throughput Screens *J. Biomol. Screen. 2014, 19 715-726* received 2016 JBS Readers Choice Award. ECO fellow Mrs. Elena Salmina contributed to the chemoinformatics analysis of this study during her short-term fellowship in HMGU, group of Dr. Tetko.

Tuesday, 26 January 2016

19th successful PhD in ECO network

On December 9 2015, Dr. Alessandra Pirovano successfully defended her PhD thesis at the Radboud University. Dr. Pirovano was ECO fellow at the same University. The topic of her thesis is "Quantifying biotransformation of xenobiotics in mammals" under supervisor of Prof. dr. ir. A.J. Hendriks.

The final School of the ECO project (September 2013)

ECO ITN publications

FOLLOW

GET MY OWN PROFILE

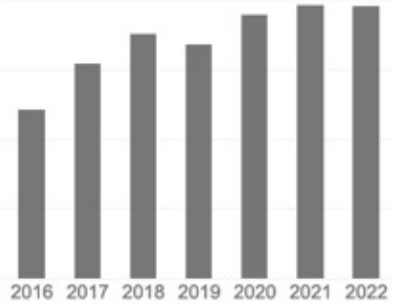
Environmental ChemOinformatic - FP7 Marie Curie Innovative Training Network
Verified email at ecoitn.eu

REACH chemoinformatics Ecotoxicology Nanotoxicology QSAR

TITLE	CITED BY	YEAR
Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information I Sushko, S Novotarskyi, R Körner, AK Pandey, M Rupp, W Teetz, ... Journal of computer-aided molecular design 25, 533-554	527	2011
Comparison of different approaches to define the applicability domain of QSAR models F Sahigara, K Mansouri, D Ballabio, A Mauri, V Consonni, R Todeschini Molecules 17 (5), 4791-4810	418	2012
Relative importance of microplastics as a pathway for the transfer of hydrophobic organic chemicals to marine life A Bakir, IA O'Connor, SJ Rowland, AJ Hendriks, RC Thompson Environmental pollution 219, 56-65	402	2016
Internalization and cytotoxicity of graphene oxide and carboxyl graphene nanoplatelets in the human hepatocellular carcinoma cell line Hep G2 T Lammel, P Boisseaux, ML Fernández-Cruz, JM Navas Particle and fibre toxicology 10, 1-21	377	2013
Quantitative structure–activity relationship models for ready biodegradability of chemicals K Mansouri, T Ringsted, D Ballabio, R Todeschini, V Consonni Journal of chemical information and modeling 53 (4), 867-878	237	2013
ToxAlerts: a web server of structural alerts for toxic chemicals and compounds with potential adverse reactions I Sushko, E Salmina, VA Potemkin, G Poda, IV Tetko Journal of chemical information and modeling 52 (8), 2310-2316	217	2012
Degree of deacetylation of chitosan by infrared spectroscopy and partial least squares IKD Dimzon, TP Knepper International journal of biological macromolecules 72, 939-945	137	2015

Cited by

	All
Citations	4686
h-index	34
i10-index	58



Co-authors


Igor V. Tetko
BIGCHEM GmbH & Helmholt


Roberto Todeschini
University of Milano-Bicocca -


Peijnenburg Willie
RIVM - Center for Safety of S


A. Jan Hendriks
Professor of Environmental S


Faizan Sahigara
Business Development Mana

Examples of careers

Kamel Mansouri [orcid:0000-0002-6426-8036]

FOLLOW

Computational Chemist at NIH/NIEHS/DTT/NICEATM
Verified email at nih.gov - Homepage

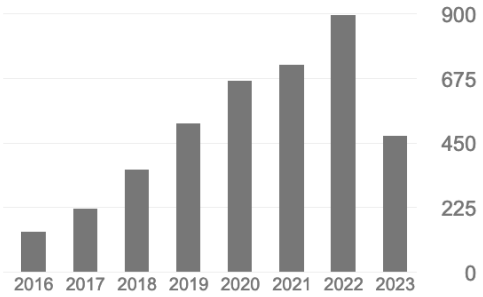
QSAR cheminformatics computational chemistry computational toxicology

GET MY OWN PROFILE

TITLE	CITED BY	YEAR
The CompTox Chemistry Dashboard: a community data resource for environmental chemistry AJ Williams, CM Grulke, J Edwards, AD McEachran, K Mansouri, ... Journal of cheminformatics 9, 1-27	624	2017
ToxCast chemical landscape: paving the road to 21st century toxicology AM Richard, RS Judson, KA Houck, CM Grulke, P Volarath, ... Chemical research in toxicology 29 (8), 1225-1251	495	2016
Comparison of different approaches to define the applicability domain of QSAR models F Sahigara, K Mansouri, D Ballabio, A Mauri, V Consonni, R Todeschini Molecules 17 (5), 4791-4810	418	2012
OPERA models for predicting physicochemical properties and environmental fate endpoints K Mansouri, CM Grulke, RS Judson, AJ Williams Journal of cheminformatics 10 (1), 1-19	294	2018
CERAPP: collaborative estrogen receptor activity prediction project K Mansouri, A Abdelaziz, A Rybacka, A Roncaglioni, A Tropsha, A Varnek, ... Environmental health perspectives 124 (7), 1023-1033	277	2016
Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 *In Vitro* High-Throughput Screening Assays for the Estrogen Receptor RS Judson, FM Magpantay, V Chickarmane, C Haskell, N Tania, J Taylor, ... Toxicological Sciences 148 (1), 137-154	262	2015

Cited by VIEW ALL

	All	Since 2018
Citations	4163	3652
h-index	26	25
i10-index	42	42



Public access VIEW ALL

1 article 25 articles

not available available

Based on funding mandates

Christian Eschauzier · 1st

Global Director Environment at AkzoNobel

Netherlands · **Contact info**

500+ connections

Ian Ken Dimzon, Inge van Driezum, and 4 other mutual connections

AkzoNobel

University of Amsterdam

Message More

Faizan Sahigara · 1st

Business Development & Brand Manager | QSAR Modeling Expert

Paris, Île-de-France, France · **Contact info**

1,418 followers · **500+** connections

Joachim Vogt, Dr. Ahmed Sayed, MBA, and 149 other mutual connections

Message More

CEHTRA

King's College London, U. of London

Ian Ken Dimzon (He/Him) · 1st

Chairperson of the Department of Chemistry at Ateneo de Manila University

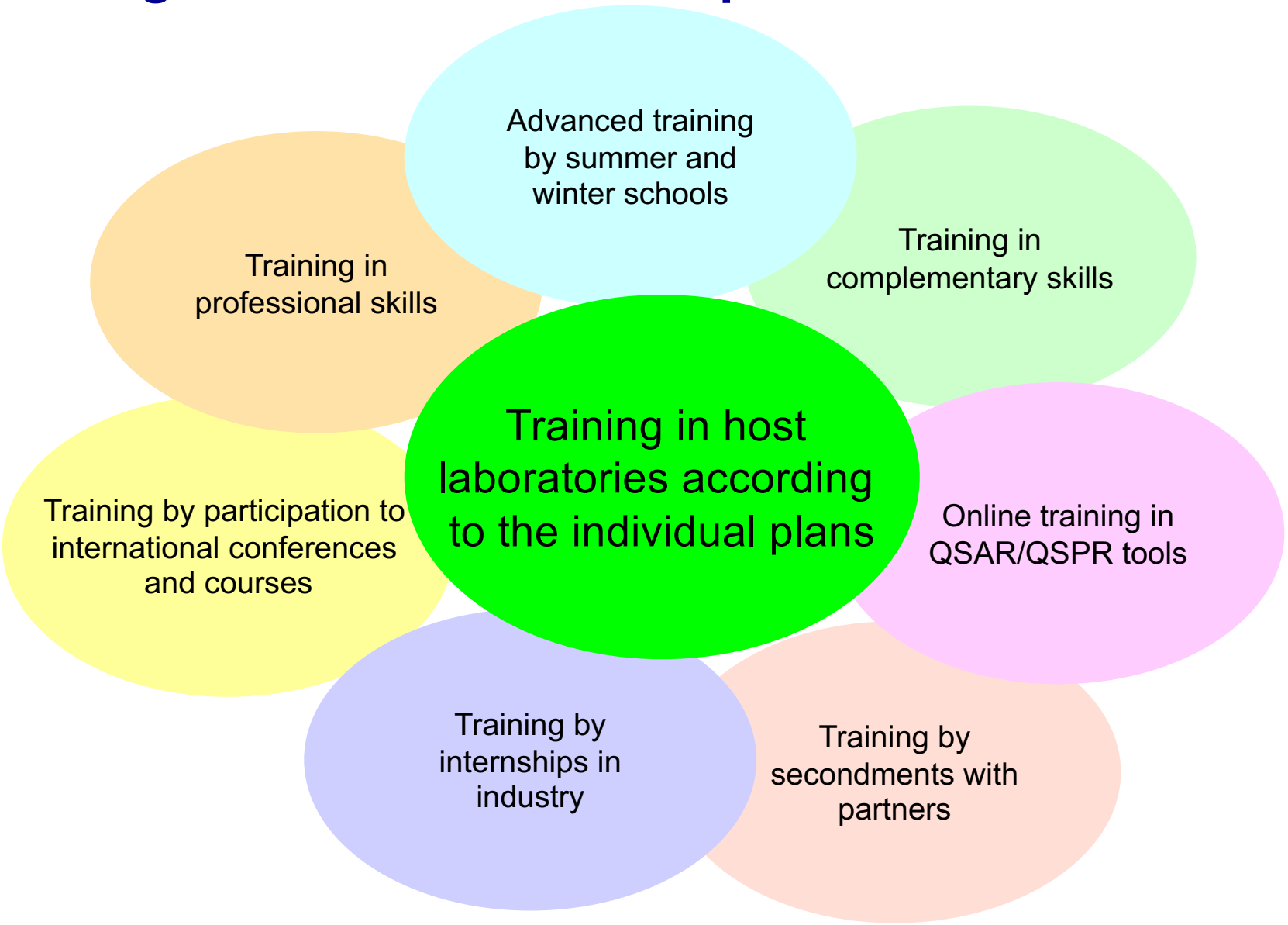
Metro Manila · **Contact info**

500+ connections

Dr. Ahmed Sayed, MBA, Yurii Sushko, and 40 other mutual connections

Message More

Training activities of ECO: steps to success in career



Accelerated Early staGe drug dIScovery

AEGIS

Accelerated Early staGe drug dIScovery

EC HORIZON 2020 Marie Skłodowska-Curie
Innovative Training Network (ITN)

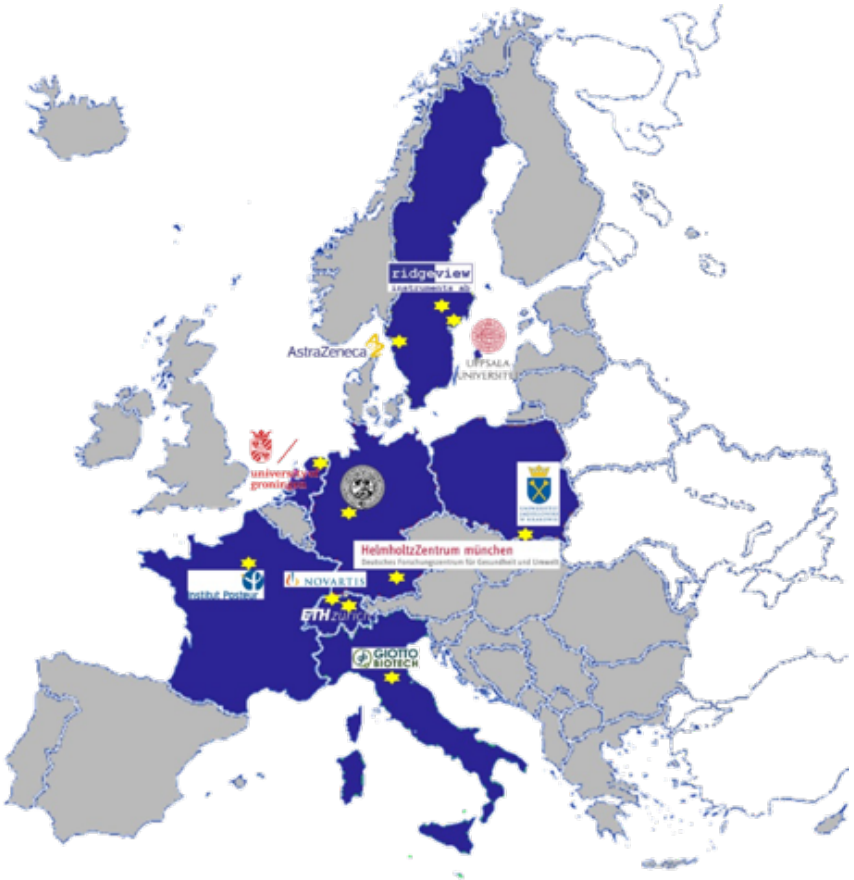
Coordination:
Michael Sattler, Helmholtz Zentrum München

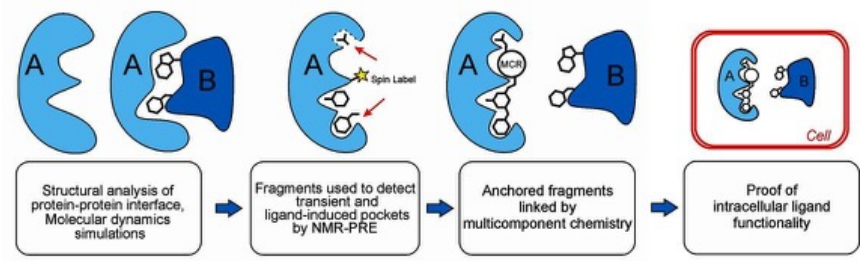
NOVARTIS

AstraZeneca

ridgeview instruments ab

GIOTTO BIOTECH

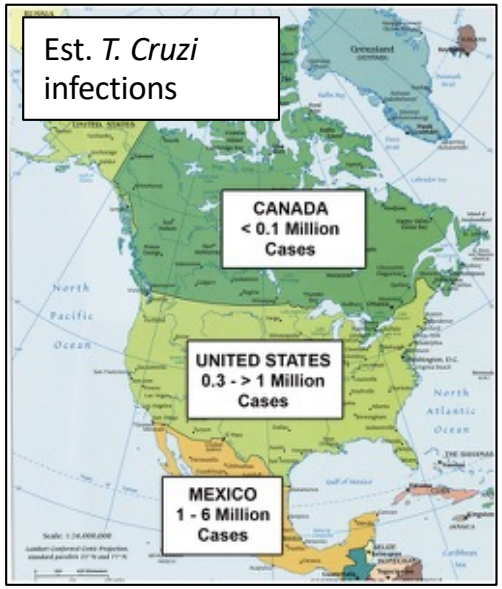




Target: Trypanosomatids caused by Kinetoplastida

Human Diseases

Disease	Parasite	Epidemilogy	Region
African sleeping sickness	*Trypanosoma brucei*	7216 Confirmed cases (2012) 20 000 estimated total 70 million in endangered area	Equatorial Africa. „Tse-tse zone“
Chagas disease	*Trypanosoma Cruzi*	41 000 New cases (2010) 10 000 Deaths	**USA, Canada** Latin America, Spain,
Leischmaniasis	*Leishmania Donovani*	1.5-2 million cases yearly, 52 000 deaths 350 million in endangered area	88 (sub)tropical countries, mainly India but also Southern America and **Texas**

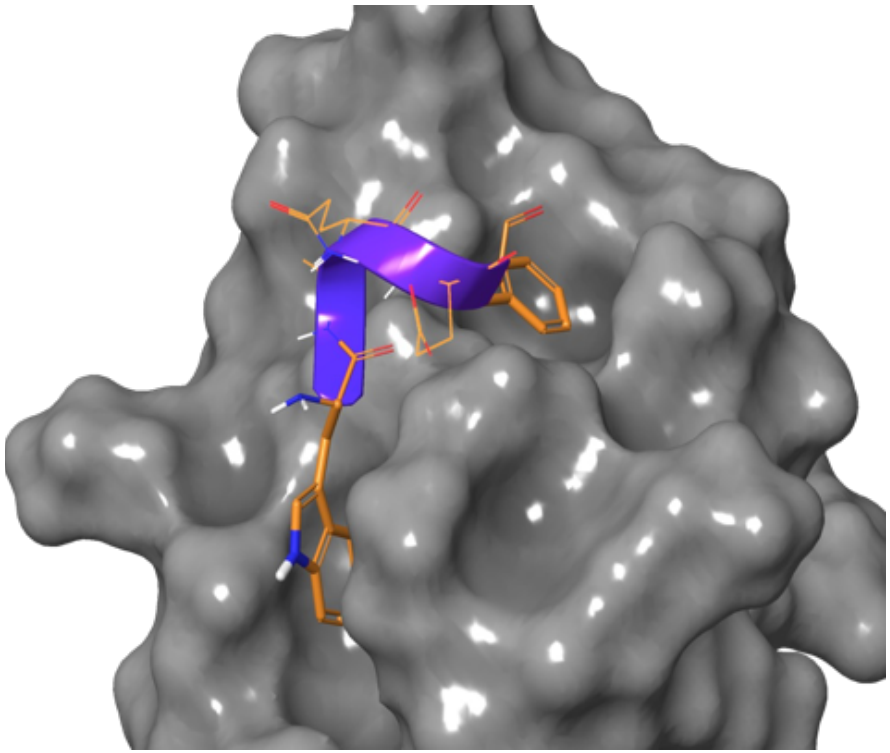


Animal Diseases

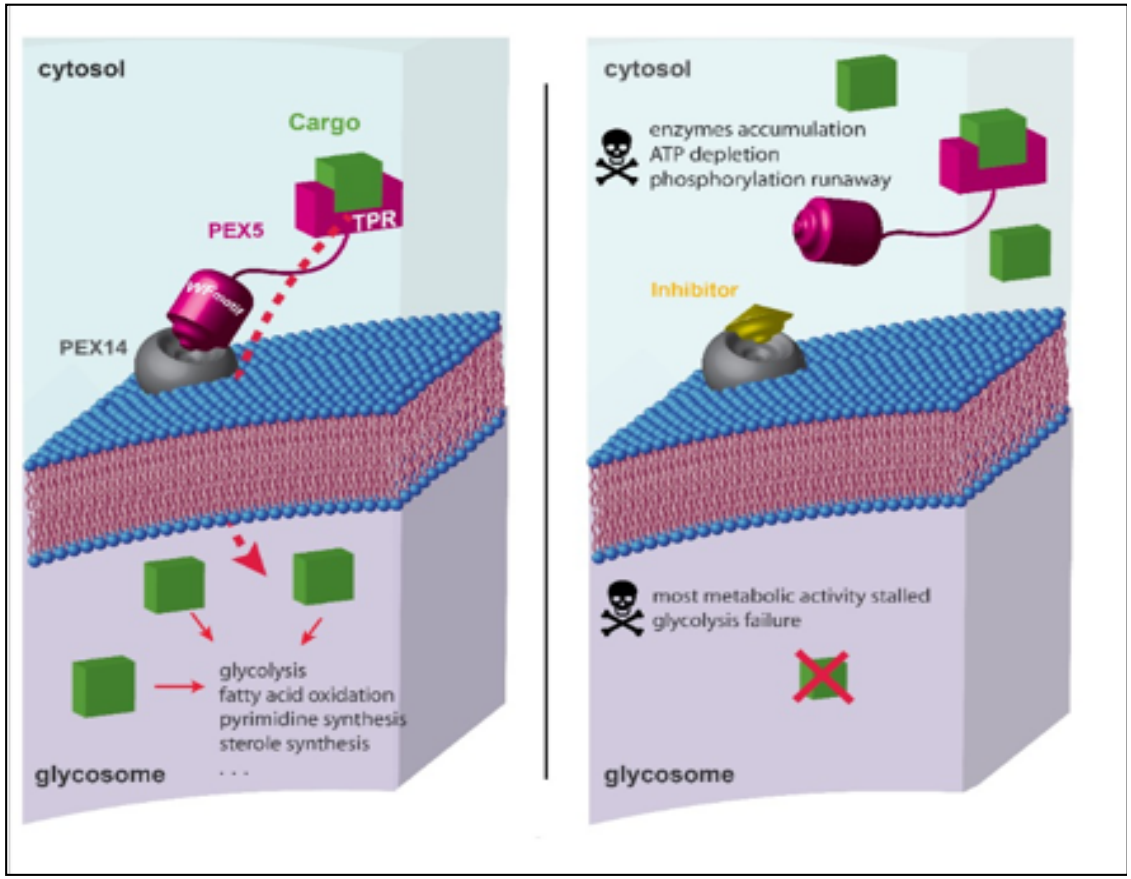
Disease	Parasite	Animals	Region
Nagana	*T. Congolese* *T. Vivax*	Cattle, sheep, pigs, horses, camels, monkeys	37 African countries
Dourine (Covering sickness)	*T. Equiperdum*	Horses, donkeys, mules **-untreatable-**	Africa, Asiatic Russia, Middle East **Europe (2011 Outbreak in Italy)**
Surra	*T. Evansi* *T. Suis*	Horses, donkeys, mules, cattle, camels,	South America, Africa, Middle East, Philippines
Sheep Ked	*T. Melophagium*	Sheep (benign)	**Northern UK**, Croatia (86%), Turkey (8%)
	T. Irwini	Koalas	Australia

Source: WHO Fact sheet N°259 FAO- PAAT
Hotez. PJ. Et al PLoS Negl Trop Dis. 7(10): e2300.

Hypothesized Target



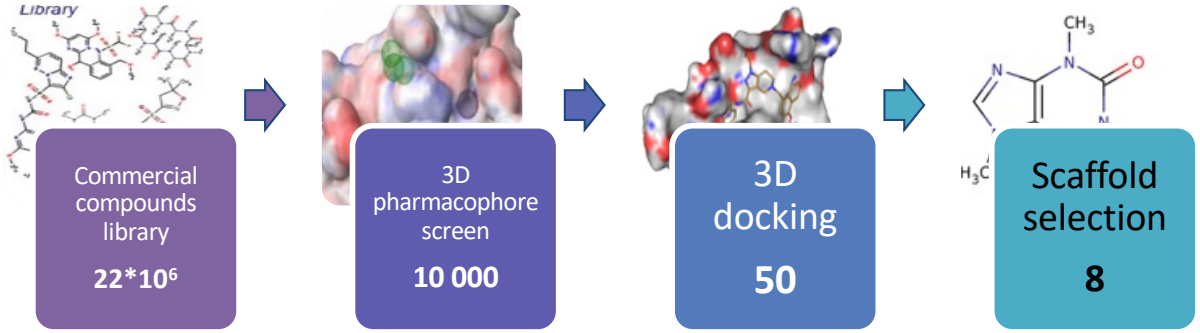
PPI interface

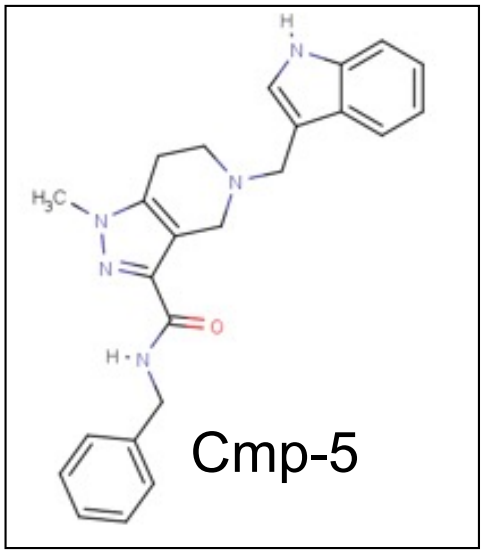


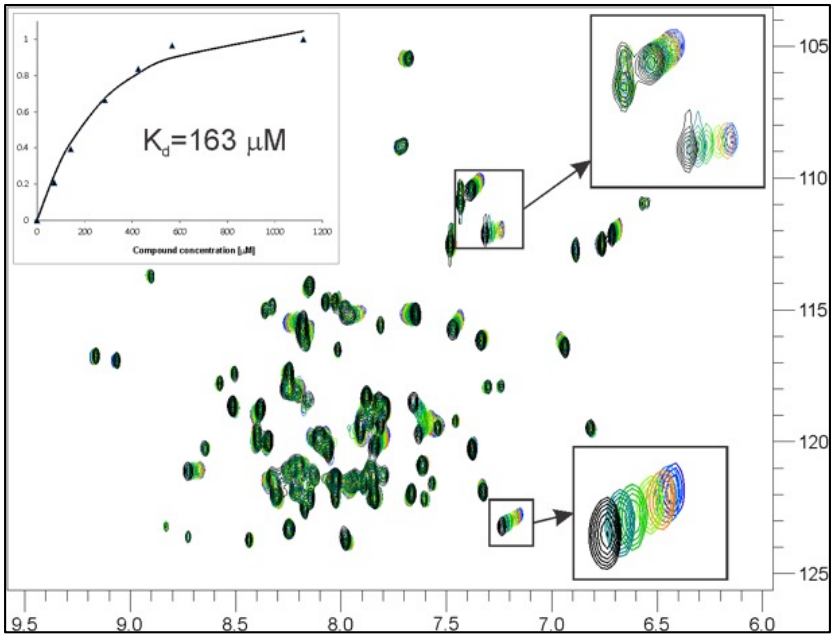
Neufeld, C. … Sattler, M. (2009) EMBO J. 28: 745-754

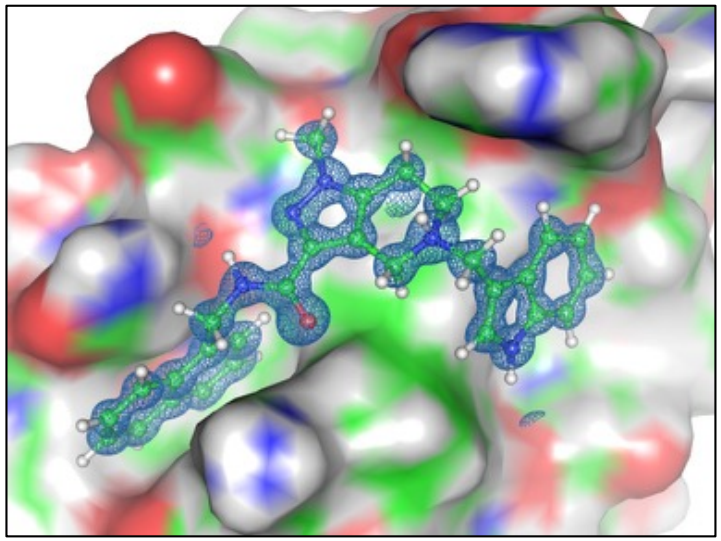
Peroxins Pex14/Pex5 are responsible for transport of glycosomal enzymes from cytoplasm to glycosomes for glucose metabolism

Hit identification









Dawidowski, M., … Popowicz, G. M. Science 2017, 355, 1416-1420.

BigChem http://bigchem.eu

big data in chemistry + informatics = chemoinformatics

The **increasing volume of biomedical data** in chemistry and life sciences requires development of **new methods and approaches for their analysis**.

The BIGCHEM project will provide **innovative education in large chemical data analysis**. The innovative research program will be implemented with the target users, **large pharma companies and SMEs**, which generate and analyze large chemical data as well as will promote technology transfer from academy to industrial applications.

Marie Skłodowska-Curie Innovative Training Network
European Industrial Doctorate

BIGCHEM publications

FOLLOW

GET MY OWN PROFILE

Horizon2020 Marie Skłodowska-Curie Innovative Training Network European Industrial Doctorate

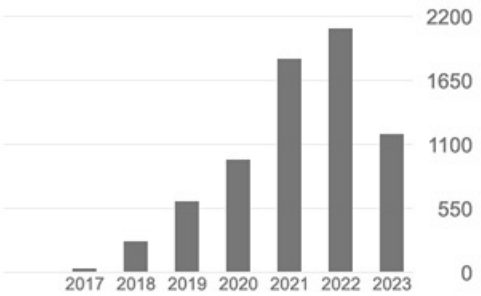
Verified email at bigchem.eu

big data chemoinformatics cheminformatics

TITLE	CITED BY	YEAR
The rise of deep learning in drug discovery H Chen, O Engkvist, Y Wang, M Olivecrona, T Blaschke Drug discovery today 23 (6), 1241-1250	1324	2018
Molecular de-novo design through deep reinforcement learning M Olivecrona, T Blaschke, O Engkvist, H Chen Journal of cheminformatics 9 (1), 1-14	836	2017
Automating drug discovery G Schneider Nature reviews drug discovery 17 (2), 97-113	519	2018
QSAR without borders EN Muratov, J Bajorath, RP Sheridan, IV Tetko, D Filimonov, V Poroikov, ... Chemical Society Reviews 49 (11), 3525-3564	440	2020
Application of Generative Autoencoder in *De Novo* Molecular Design T Blaschke, M Olivecrona, O Engkvist, J Bajorath, H Chen Molecular informatics 37 (1-2), 1700123	357	2018
Molecular representations in AI-driven drug discovery: a review and practical guide L David, A Thakkar, R Mercado, O Engkvist Journal of Cheminformatics 12 (1), 1-22	236	2020
A de novo molecular generation method using latent vector based generative adversarial network O Prykhodko, SV Johansson, PC Kotsias, J Arús-Pous, EJ Bjerrum, ... Journal of Cheminformatics 11 (1), 1-13	216	2019

Cited by

	All	Since 2018
Citations	7086	7022
h-index	33	32
i10-index	55	55



Co-authors

VIEW ALL

- Ola Engkvist, AstraZeneca R&D Gothenburg O…
- Hongming Chen, Bioland Laboratory
- Jürgen Bajorath, Professor of Life Science Inform…
- Thomas Blaschke, Phd student, AstraZeneca/Unive…

BIGCHEM fellows: three years after the project end

Two fellows who joined BICGHEM later

Dr. Raquel Rodríguez-Pérez, University of Bonn/Boehringer Ingelheim - PhD by University of Bonn - now **Principal Scientist at Novartis,** Basel, Switzerland

Dr. Arkadii Lin, University of Strasbourg/Boehringer Ingelheim - PhD by University of Strasbourg -now **Team Leader in Chemoinformatics at Insilico Medicine**, Abu Dhabi

Dr. Dipan Ghosh, Helmholtz Zentrum München/LDC - PhD by Technical University of München - now **Scientist at Lead Discovery Center GmbH**, Dortmund, Germany

Laurianne David, University of Bonn/AstraZeneca - now **Computational Chemist at Evotec**, Toulouse, France

Dr. Josep Arús-Pous, University of Bern/AstraZeneca - PhD by University of Berm - now **Senior Machine Learning Engineer at Roche,** Basel, Switzerland

Dr. Xuejin Zhang, ETH Zürich/Boehringer Ingelheim - PhD by ETH Hönggerberg -now **ML/AI Scientist at VantAI, USA**

Thomas Blaschke, AstraZeneca/University of Bonn (finishing PhD, **working on a start-up in the UK**)

Dr. Amol Thakkar, University of Bern/AstraZeneca - PhD by University of Berm - now **Research Scientist at IBM Research**, Zurich,

Michael Withnall, Helmholtz Zentrum München/AstraZeneca - now **Head of Computational Chemistry at Apheris AI,** Berlin

Oliver Laufkötter, AstraZeneca/University of Bonn (finishing PhD)

Benedict Mutimba,* Boehringer Ingelheim - now **Technical Project Manager & Scrum Master at Klick Health,** Toronto, Canada

Eric March Vila,* University of Modena and Reggio Emilia, now - **Research Associate at Universitat Pompeu Fabra,** Barcelona

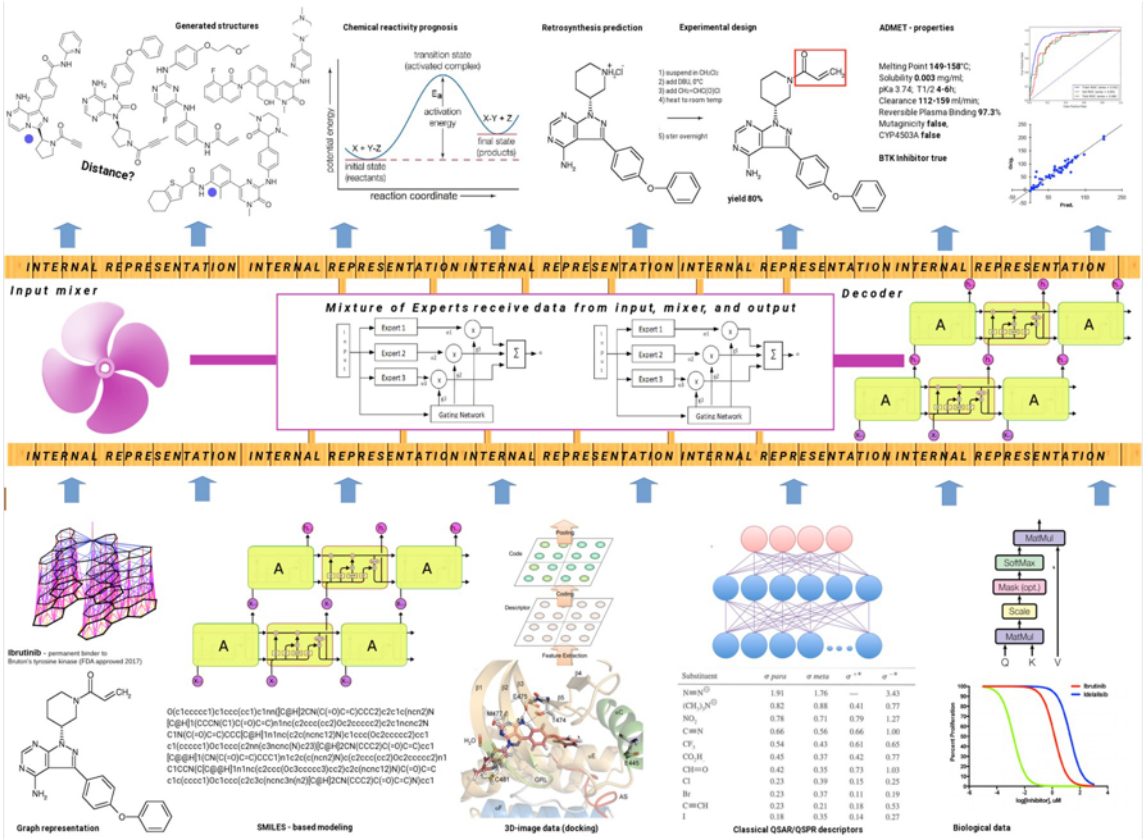
AIDD ITN https://ai-dd.eu

Marie Skłodowska-Curie Actions

Advanced machine Learning for Innovative drug discovery

HELMHOLTZ MUNICH

16 PhDs
17 beneficiaries
13 partners
5 companies
17 countries
3.93 M€



AstraZeneca
janssen
Bayer
Enamine
Pfizer

JKU
RESEARCH TREATMENT
upf. Universitat Pompeu Fabra Barcelona
technische universität dortmund
JG|U
Aalto University
UNIVERSITÄT LUXEMBURG
Scuola universitaria professionale della Svizzera italiana
SUPSI
Freie Universität Berlin
universität wien

https://twitter.com/AiddOne

School at AstraZeneca

Presentations with fellows at conferences, workshops and meetings

2023

All fellows also presented their posters during the AIDD School in AALTO (March 21, Espoo/Helsinki, Finland)

- Presentation of OCHEM model from Kaggle 1st EUOS/SLAS Joint Challenge: Compound Solubility Challenge at during HMGU MTCC workshop (July 7th, 2023)
- Torren-Peraire, P. Enhancing Chemical Synthesis Planning through Combining Single-Step and Multi-Step Retrosynthesis Prediction Strategies. STB seminar (departmental seminar). (May 26th, 2023).
- Mikhail Andronov gave a talk about reagent prediction at the DigiDrug seminar in Berlin organized by Prof. Andreas Bender (May 24th, 2023).
- Invited lecture to present winning model of the Kaggle 1st EUOS/SLAS Joint Challenge: Compound Solubility Challenge at SLAS 2023 conference in Brussels (Peter Hartog is a team member)
- Torren-Peraire, P. AI in the Lab: How Machine Learning Can Transform Chemical Synthesis. Pint of Science Munich. (May 22nd, 2023).
- Ha, S.V. FSL-CP: Few-shot Prediction of small molecule activity using cell microscopy images. AIDD on-line seminar (May 17th, 2023)
- Masood, A., Heinonen, M., Herman, D., Ceulemans, H. Kaski, S. Dos-Time dependent DILI modeling. In Janssen Discovery Data Science meeting. (May 8th, 2023).
- Nahal, Y. Heinonen, M., Engkvist, O. Kaski, S. Human-in-the-loop active learning to improve molecular design and optimization. In AstraZeneca Molecular Design meetings. (April 13th, 2023).
- Torren-Peraire P. Mind the Retrosynthesis Gap: Bridging the divide between Single-step and Multi-step Retrosynthesis Prediction. AIDD on-line seminar (April 5th, 2023)
- Hartog, P., Genheden, S., Tetko, I. Two sides of the same coin: The effect of smiles-based molecular representations on explainability. <Interact> Conference (March 31, 2023)
- Hassen AK., Torren-Peraire P., Genheden S., Verhoeven J, Preuss M., Tetko I. Mind the Retrosynthesis Gap: Bridging the divide between Single-step and Multi-step Retrosynthesis Prediction. <Interact> Conference. (March 30th, 2023).
- Fallani, A. Extending 3D generative modeling of molecules with quantum-mechanical properties. AIDD on-line seminar (March 8, 2023)
- Cremer, J. Equivariant Graph Neural Networks for Toxicity Prediction. AIDD on-line seminar (February 1st, 2023)
- Andronov, M. Reagent prediction with a transformer and its benefits for reaction product prediction. AIDD on-line seminar (February 1st, 2023)

2022

All fellows also presented their posters during the AIDD School in KUL (October 20, Leuven, Belgium)

- Palmacci V. Drug Discovery and Cheminformatics: discovering new drugs in the Big Data era. Seminar at University of Boogna. December 02, 2022.
- Hassen AK., Torren-Peraire P., Genheden S., Verhoeven J, Preuss M., Tetko I. Mind the Retrosynthesis Gap: Bridging the divide between Single-step and Multi-step Retrosynthesis Prediction. NeurIPS 2022 workshop AI for Science: Progress and Promises. December 2, 2022.
- Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Robust task-specific adaption of models for drug-target interaction prediction. In NeurIPS2022 AI4Science Workshop. November 28, 2022.
- Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Task-conditioned modeling of drug-target interactions. In ELLIS Machine Learning for Molecules Discovery Workshop. November 28, 2022.
- Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Task-conditioned modeling of drug-target interactions. In NeurIPS2022 Women in Machine Learning Workshop. November 28, 2022.
- Sanchez-Fernandez, A.; Rumetshofer, E.; Hochreiter, S.; Klambauer, G. Cross-modal Contrastive Learning of Microscopy Image- and Srtructure-Based Representations of Molecules. In NeurIPS2022 Women in Machine Learning Workshop. November 28, 2022.
- Ha, S.V. , Tandon A., Czodrowski, P. Overview of Czodrowski Lab AK-Symposium. Johannes Gutenberg University Mainz. November 10th, 2022.
- Masood, A., Heinonen, M., Herman, D., Ceulemans, H. Kaski, S. Dos-Time dependent DILI modeling. In STB, Helmholtz Zentrum München. November 08, 2022.
- Ha, S.V. Few-shot bioassay prediction with Cell Painting for drug discovery. RdKit UGM 2022. October 13th, 2022.
- Sanchez-Fernandez, A. Cross-Modal Contrastive Learning of Microscopy Image and Structure-Based Representations of Molecules. CytoData Symposium. October 10, 2022.
- Masood, A., Heinonen, M., Herman, D., Ceulemans, H. Kaski, S. Dos-Time dependent DILI modeling. In Finnish Center of Artificial Intelligence Virtual Drug Design Lab seminars. October 10, 2022.
- Andronov, M.; Voinarovska, V.; Wand, M.; Schmidhuber, J. Reagent Prediction With a Molecular Transformer Improves Reaction Data Quality. 23rd EuroQSAR, Heidelberg, Germany, September 2022
- Voinarovska, V.; Dudenko, D.; Torren-Peraire, P.; Tetko, I.; Genheden, S. Addressing the applicability domain in yield prediction, 23rd EuroQSAR, Heidelberg, September 26-30, Germany 2022,
- Nahal, Y. Heinonen, M., Engkvist, O. Kaski, S. Human-in-the-loop active learning to improve molecular design and optimization. In Finnish Center of Artificial Intelligence Virtual Drug Design Lab seminars. September 6, 2022.
- Friesacher H.R., Lewis Mervin L., Engkvist O., Moreau Y., Arany A. Can we trust probabilities in deep drug activity models? A comparative calibration study. 21st European Conference on Computational Biology. September 15, 2022.
- Sanchez-Fernandez, A., Rumetshofer, E., Hochreiter, S., and Klambauer, G. Contrastive learning of image-and structure-based representations in drug discovery. In ICML2022 Workshop on Women in Machine Learning. July 18th, 2022
- Radaeva, M. "Novel CADD-designed Lin28B Inhibitors Suppress Stemness in Neuroendocrine Prostate Cancer" in Strasbourg Chemoinformatics Conference, July 1st, 2022
- Fallani, A., Medrano Sandonas, L. and Tkatchenko, A. Towards the inverse design of molecules with target quantum mechanical properties. In University of Luxembourg Department Workshop. June 2022.
- Svensson, E., Hartog, P. AIDD Codebase: a Framework for Model Integration, Collaboration and Sharing. AIDD on-line seminar (June 22, 2022)
- Svensson, E., Hoedt, P.-J., Hochreiter, S., Klambauer, G. Robust task-specific adaption of drug-target interaction models. In ICML2022 Workshop on Women in Machine Learning. July 18th, 2022.
- Sanchez-Fernandez, A., Contrastive Learning of Image and Structure-Based Representations in Drug Discovery. AIDD on-line seminar (June 8th, 2022)
- Sanchez-Fernandez, A., Rumetshofer, E., Hochreiter, S., and Klambauer, G. Contrastive learning of image-and structure-based representations in drug discovery. In ICLR2022 Workshop on Machine Learning for Drug Discovery. (April 29th, 2022)
- Fallani, A., Medrano Sandonas, L. and Tkatchenko, A. Towards the inverse design of molecules with target quantum mechanical properties. In APS March Meeting. March 2022.
- Nahal, Y. A Survey on Human-in-the-loop Machine Learning on-line AIDD lecture (March 16, 2022)
- Andronov, M. Overview of the methods for chemical reaction prediction, Rising Stars in AI Symposium 2022 in KAUST, Thuwal, Saudi Arabia (March 13, 2022)
- Nahal, Y. Learning from user feedback to improve recommender models and potential applications to molecular design. In Finnish Center of Artificial Intelligence Virtual Drug Design Lab seminars. (March 1st, 2022)
- Fallani, A., Medrano Sandonas, L. and Tkatchenko, A. Towards the inverse design of molecules with target quantum mechanical properties. In DQML Hintertux joint conference. February 2022.
- Masood, A., Heinonen, M., Herman, D., Ceulemans, H. Kaski, S. Dos-Time dependent DILI modeling. In Finnish Center of Artificial Intelligence Virtual Drug Design Lab seminars. (February 8th, 2022)

Program of the First AIDD School in Munich

			Building 57 / Room 052		Building 57 / Room 052		Building 57 / Room 052		B56/ R160a/b	Building 57 / Room 052	
First week	**18.10.21**		**19.10.21**		**20.10.21**		**21.10.21**			**22.10.21**	
Time	**Monday**		**Tuesday**		**Wednesday**		**Thursday**			**Friday**	
10:00-11:00			RDkit: basics	Gregory Landram, Zoom	Workshop on PyTorch	Thomas Viehmann				Reinforcement Learning	Jose Arjona, 45 mins
11:00-12:00			History of AI	Mark Embrechts, live			Transformers for Computer Vision	Thomas Unterthiner, Zoom		Reinforcement Learning	Philipp Renz, 45 mins
12:00-13:00			Lunch		Lunch		Lunch			Lunch	
13:00-14:00			How did we do deep learning in the BIGCHEM: experiences, challenges and opportunities	Josep	Neural networks: tricks of the trade	Pieter-Jan Hoedt, 45 mins, Zoom	Python: advanced (numpy, decorators, network)	Andreas Poehlmann, Zoom	Supervisory Board	SMILES based modelling, tools and models	Esben Jannik Bjerrum, Zoom
14:00-15:00			Introdcution to chemoinformatics	Tropsha	Workshop on PyTorch	Thomas Viehmann	QSAR: substructure analysis	Peter Ertl, Zoom		Introduction to best practices	Samuel Genheden, Zoom
15:00-16:00			Personal presentations of ESRs (5 people)	ESR1-ESR6			Introduction to modeling chem reactions with ML	Marvin Segler, Zoom	General Assembly	Recurrent networks, LSTM, and transformer	Michael Widrich, Zoom
16:00-17:00	Welcome party, SeeHaus in Englischen Garten										

	Building 57 / Room 052		Building 57 / Room 052		Building 57 / Room 052		Building 57 / Room 052			
	25.10.21		**26.10.21**		**27.10.21**		**28.10.21**		**29.10.21**	
Second week	**Monday**		**Tuesday**		**Wednesday**		**Thursday**		**Friday**	
10:00-11:00			Python: testing	Samuel Genheden, Zoom						
11:00-12:00	QMAR: quantitative cell morphology-activity relationships	Adam Arany, live	Python: questions and answers	Andreas Poehlmann, Zoom	Optical chemical structure recognition	Clevert, live	Computational Cancer Pharmacogenomics	Michael Menden	Die Josefsthaler Wasserfälle	
12:00-13:00	Lunch		Lunch		Lunch		Lunch			
13:00-14:00	Atomic Simulation Environment (ASE)	Leonardo Medrano Sandonas (live)	Molecular property prediction	Andreas Mayr	Convolutional Neural Networks	Mark Embrechts, live	Similarity-based data mining with applications to drug discovery	Petra Perner	https://docs.google.com/document/d/1FNgqjcD-x1oiizutOIXzIjJG3sdNmlxdK1tZDVQrD1g/edit?usp=sharing	
14:00-15:00	Introduction to Reaction Informatics	Igor Baskin, Zoo	High Performance Computing / Python (cluster, muli-GPU, etc)	Martijn Oldenhof, live	Placebo	Hans Georg Zimmerman	Personal presentations of ESRs (5 people)	ESR7-ESR11		
	Coffee break		Coffee break							
15:00-16:00	Molecular representations	Rafael Gomez-Bombarelli, Zoom	Structure Query Language: basics	Pavel Karpov, live	Recurrent Neural Networks		Personal presentations of ESRs (5 people)	ESR11-ESR16		
16:00-17:00	Accelerated data science with applications in natural language processing	Anthony (Nvidia)	Generative models and optimization for molecules	Rocío Mercado, zoom, US			Organic reactions modeling: start-of-the-art and perspectives	Connor W. Coley, Zoom		
17:00-18:00	Introduction to CUDA with Python	Christian Hundt (Nvidia)								

Fellows attending School in Lugano

Social activities: boat excursion in Belgium

You Retweeted

BioLearning_KULeuven @BioLearning_KUL · Oct 22, 2022

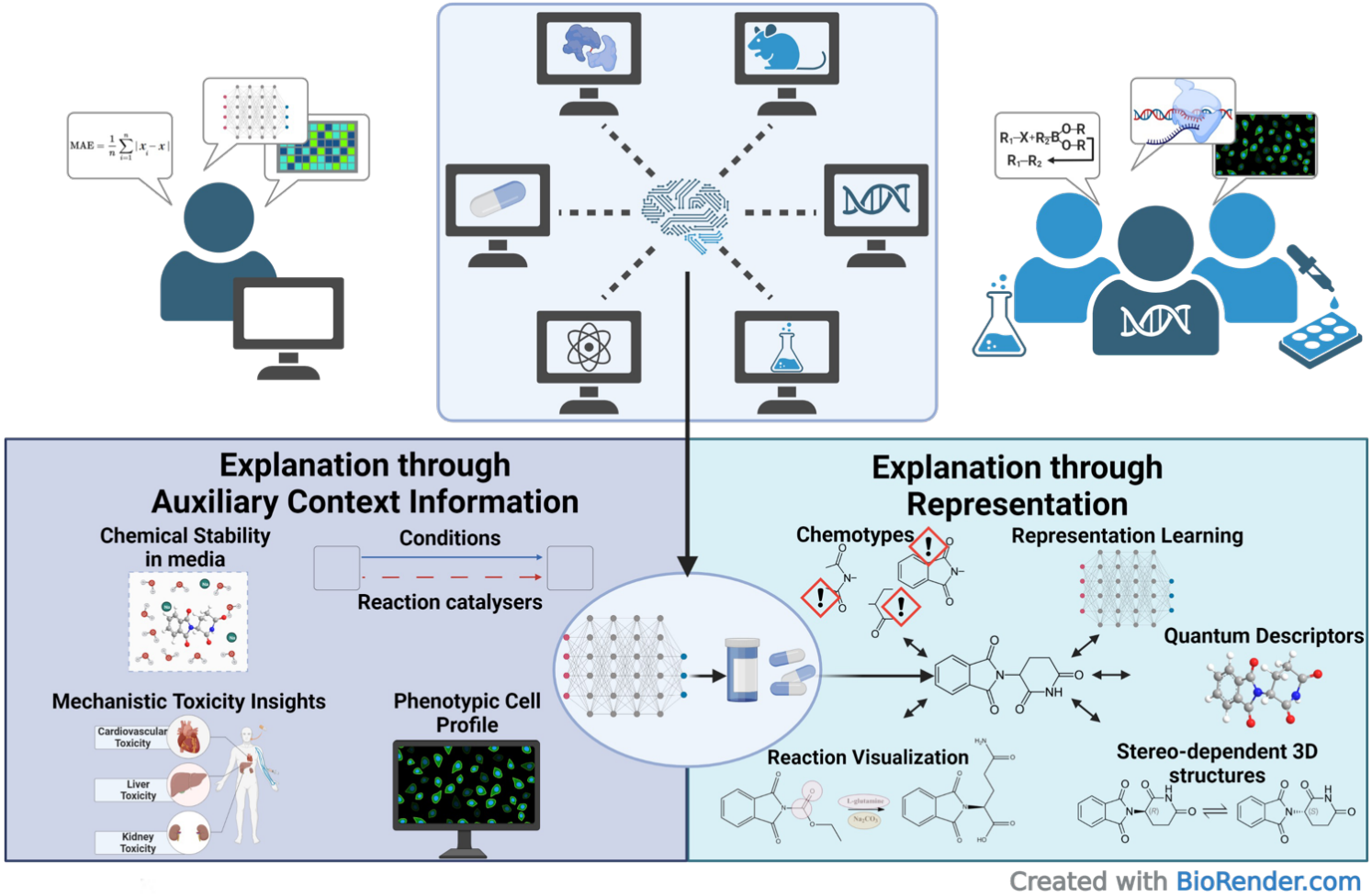
@AiddOne 3rd school social. Ghent visit, and experimenting with herbal extract flavours in beer.

Outreach activities: Pint of Science in Munich pub

AIDD @AiddOne · May 22

Paula @PTorrenPeraire explains the challenges to synthesize new active molecules

Explainable AI for molecules - https://aichemist.eu



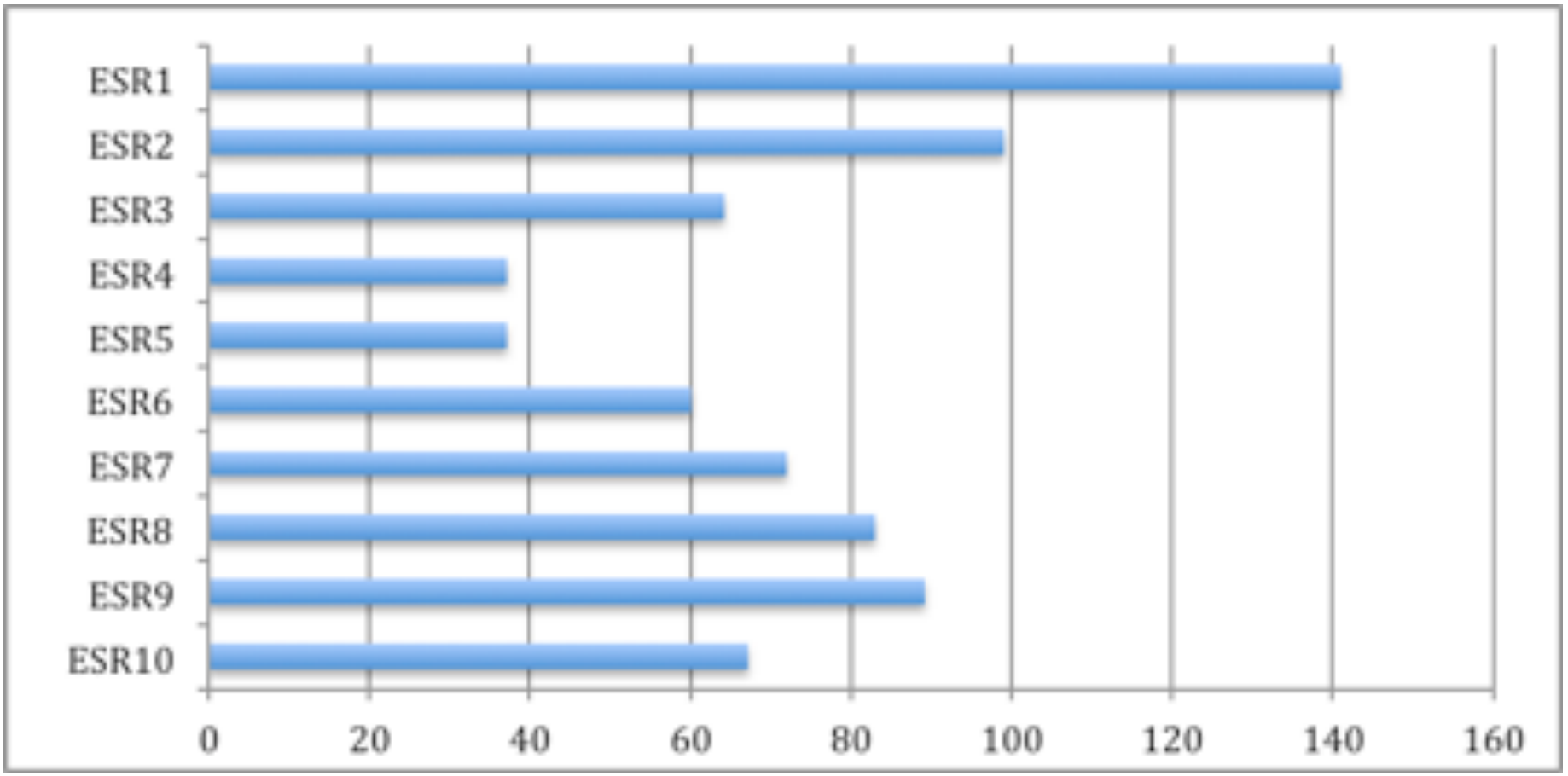
Hyun Kil Shin · 1st

Artificial intelligence in chemistry

South Korea

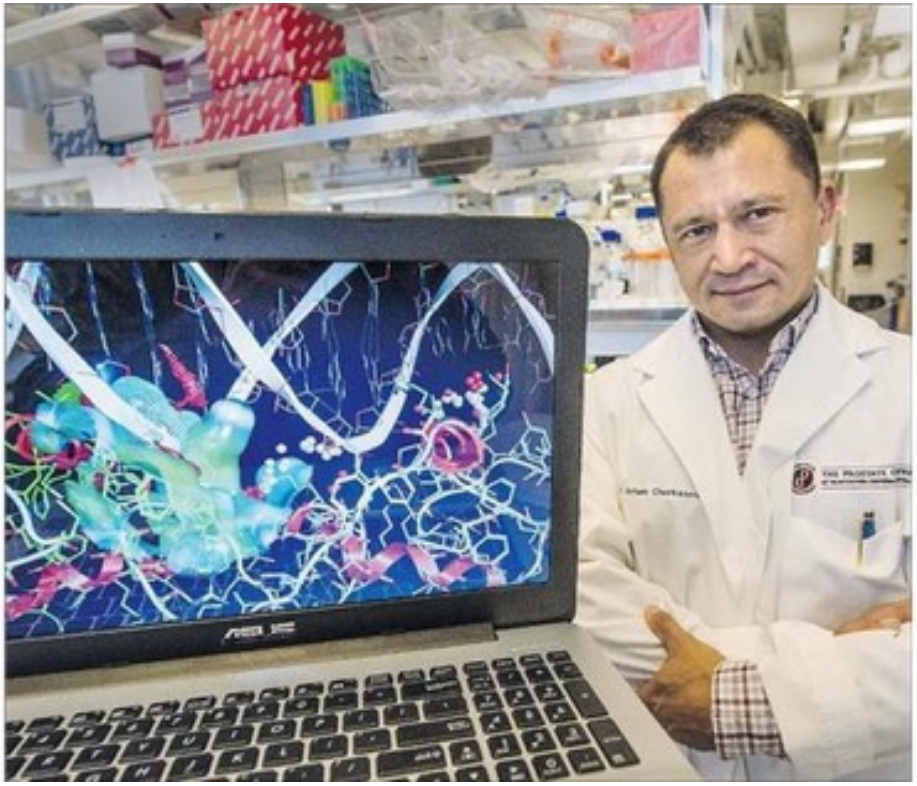
Experience: Korea Institute of Toxicology, University of Science & Technology (UST), and 2 more

Number of applications per ESR position (BIGCHEM)



Science and entrepreneurship

UBC lands massive licensing deal for cancer drug

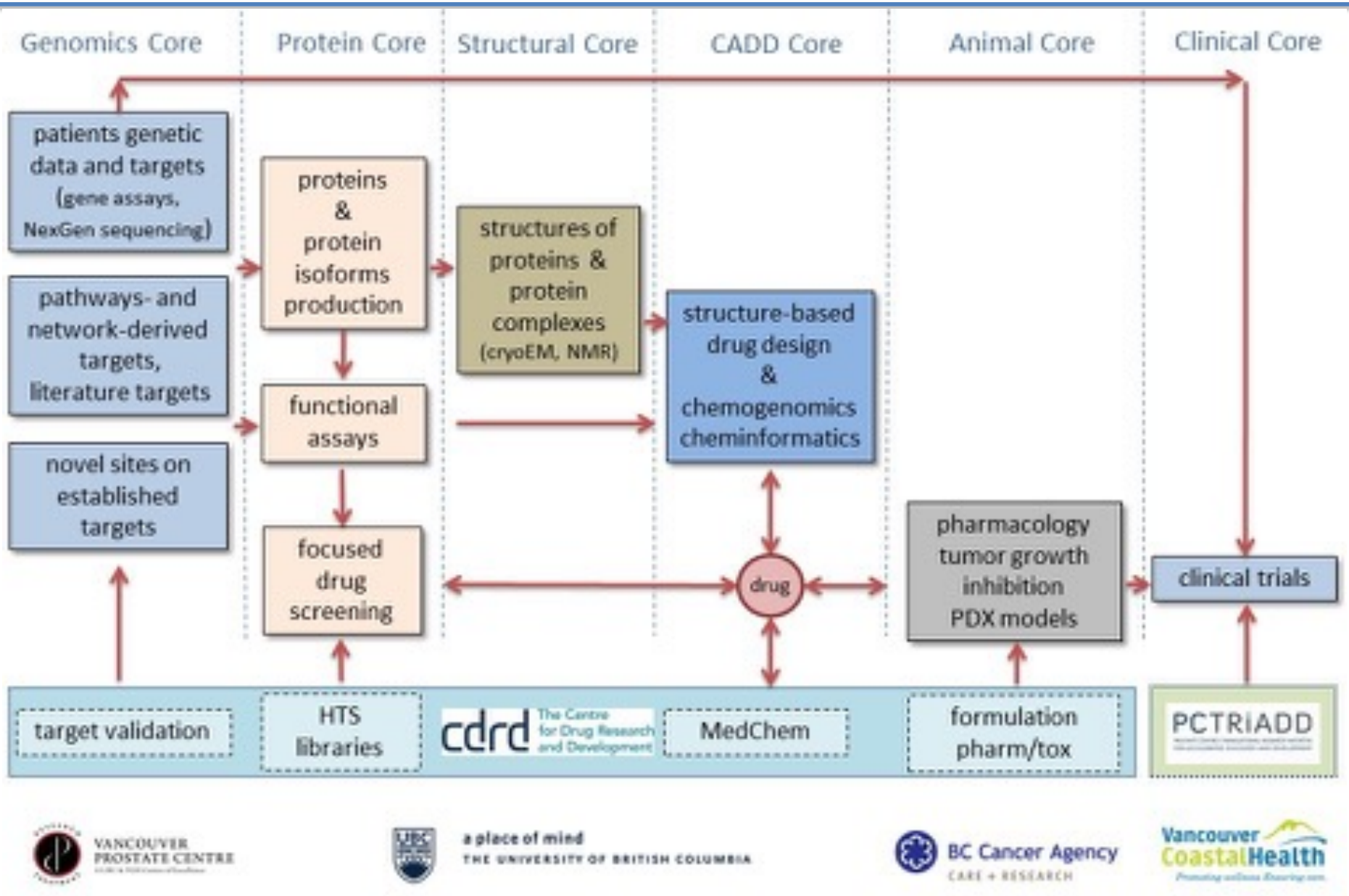


Researcher Artem Cherkasov displays a computer model simulation used to develop a new treatment for drug-resistant prostate cancer at the Vancouver Prostate Centre . '**Using computer simulations, we sometimes go through 50 million compounds to find a molecule that will seat in a precise and accurate way**,' he says. Photograph by: Steve Bosch, PNG

Under the terms of the agreement with Roche, UBC and VCHRI receive an upfront payment and up to $141.7 million US in milestone payments if the drug moves through pre-clinical and clinical trials, regulatory approval and meets sales targets, and then royalties thereafter.

http://vancouversun.com/news/staff-blogs/ubc-lands-massive-licensing-deal-for-cancer-drug

ADDUCT Organization & Infrastructure



Slide courtesy of Prof. A. Cherkasov

First-in-class drug for blood cancers discovered by Ontario researchers receives record-setting industry investment

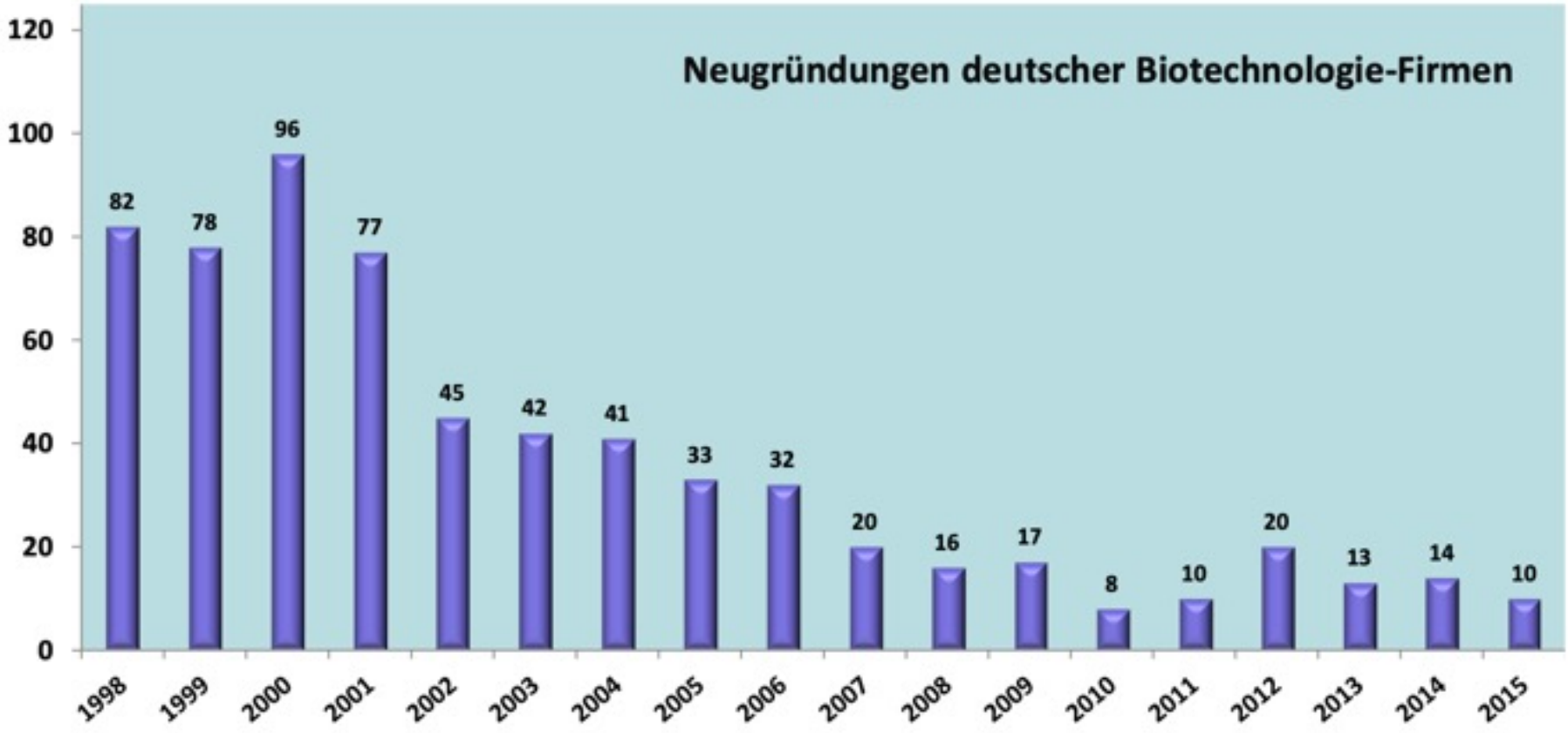
Years of hard work by OICR's Drug Discovery group and Ontario partners moves potential new treatment for leukemia towards clinical trials

On January 29, 2019, Celgene Corporation made an investment of up to US$1 billion that will facilitate further research and development of the potential drug and support clinical trials in Ontario. The potential drug was designed to exploit a weakness in leukemia centred on the protein WDR5 that was discovered by Ontario researchers.

https://news.oicr.on.ca/2019/01

Example of GO-Bio funding in Germany

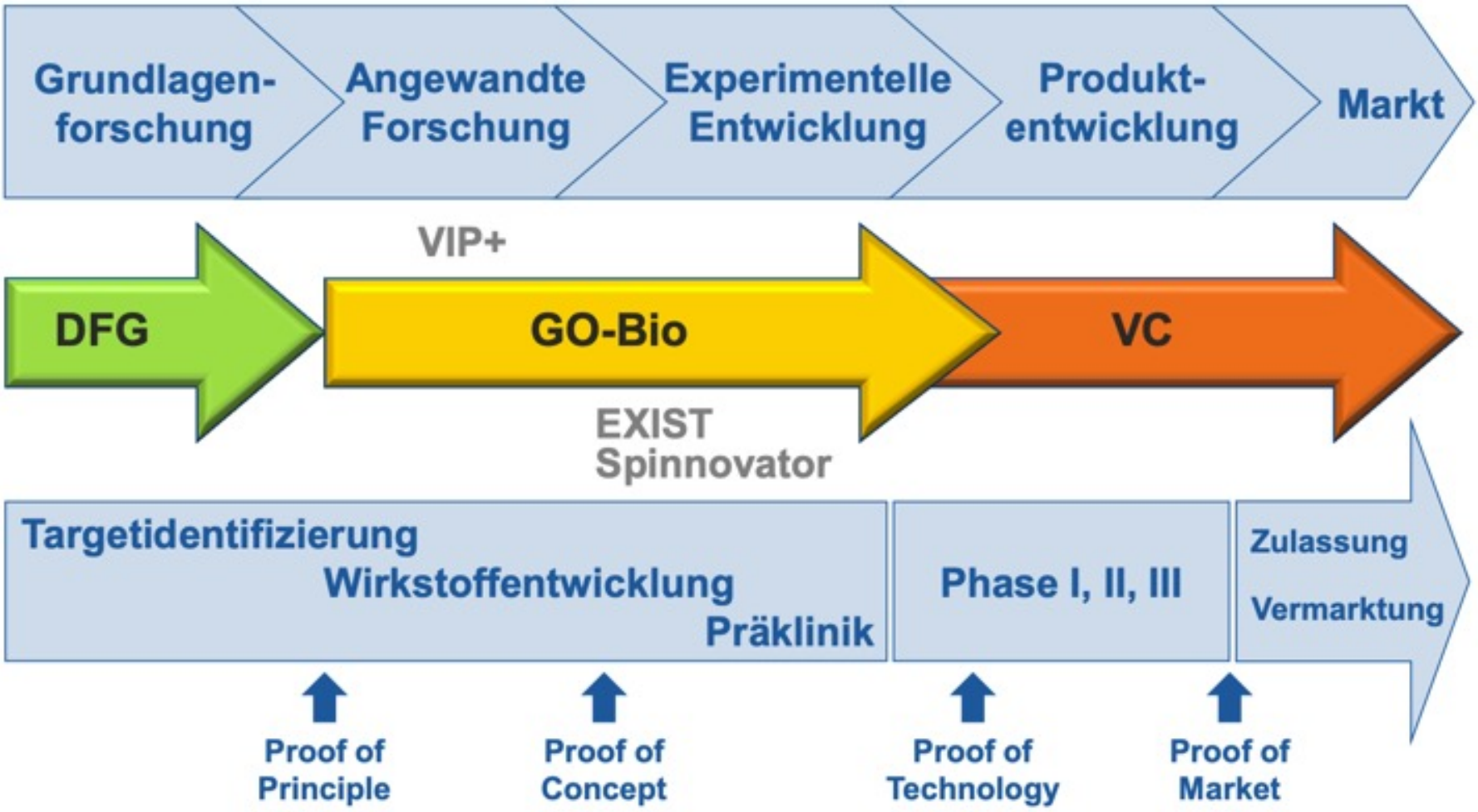
GoBio program



(Quelle: Ex-post-Evaluation der KMU-Förderung des BMBF in der Biotechnologie (ZEW 2012), modifiziert)

https://www.go-bio.de/gobio/shareddocs/downloads/de/praesentation_gb8170302.pdf

Financing rounds



https://www.go-bio.de/gobio/shareddocs/downloads/de/praesentation_gb8170302.pdf

Projektträger Jülich
Forschungszentrum Jülich

Bundesministerium
für Bildung
und Forschung

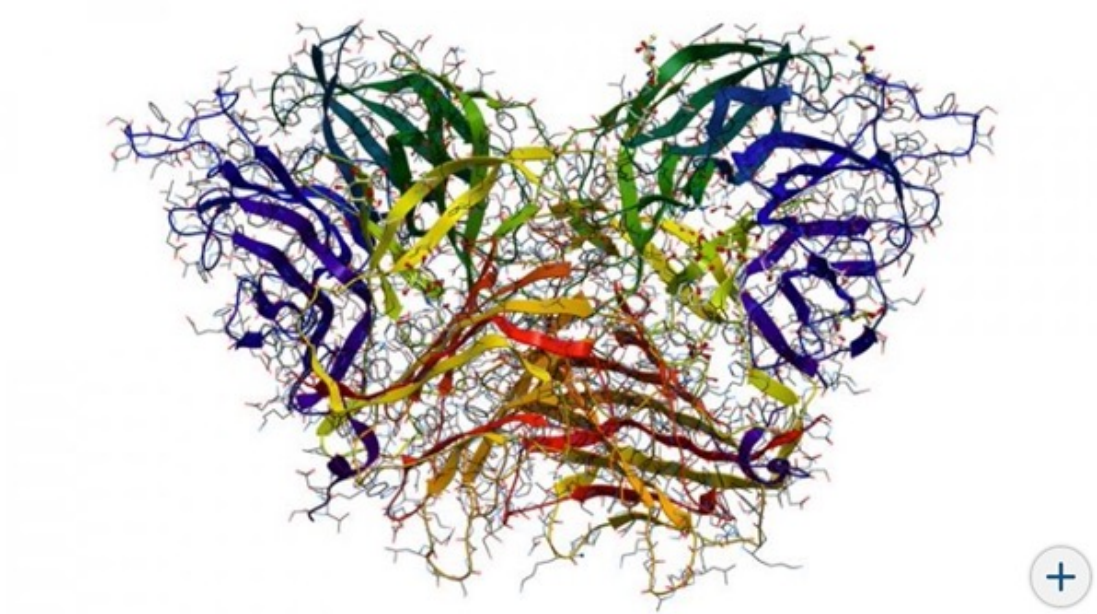
Firma	Sitz	Gründung
Corimmun	Martinsried	2006
pluriSelect	Leipzig	2006
Elara Pharmaceuticals	Heidelberg	2006
ChromoTek	Martinsried	2008
Riboxx	Radebeul	2009
BioNTech RNA Pharmaceuticals	Mainz	2009
FreiBiotics	Freiburg	2009
ethris	Martinsried	2009
iThera Medical	Oberschleißheim	2010
TissUse	Berlin	2010
SYNIMMUNE	Tübingen	2010
Eupheria Biotech	Dresden	2010
Cortec	Freiburg	2010
eADMET	Oberschleißheim	2010
AgroProtect	Aachen	2010
PSites Pharma	Frankfurt	2011
CCRP Therapeutics	Berlin	2011
Dynamic Biosensors	München	2012
BianoScience	Zwickau	2012
ImevaX	München	2014
Rigontec	Siegburg	2014
Zellkraftwerk	Leipzig	2014
SenseUp	Jülich	2015
Signatope	Reutlingen	2016

https://www.go-bio.de/gobio/shareddocs/downloads/de/praesentation_gb8170302.pdf

16.06.2020 | GO-BIO

Neues Therapie- und Diagnoseverfahren zur Behandlung von autoimmunbedingter Herzschwäche

GO-Bio Runde 1 – PD Dr. med. Roland Jahns – Medizinische Klinik und Poliklinik I, Universitätsklinikum Würzburg | corimmun GmbH



© Leonid Andronov - Fotolia

Zuwendungsempfänger: Julius-Maximilians-Universität Würzburg
Förderung: GO-Bio Phase I (01.04.2007 bis 30.09.2008, 970.397 Euro)

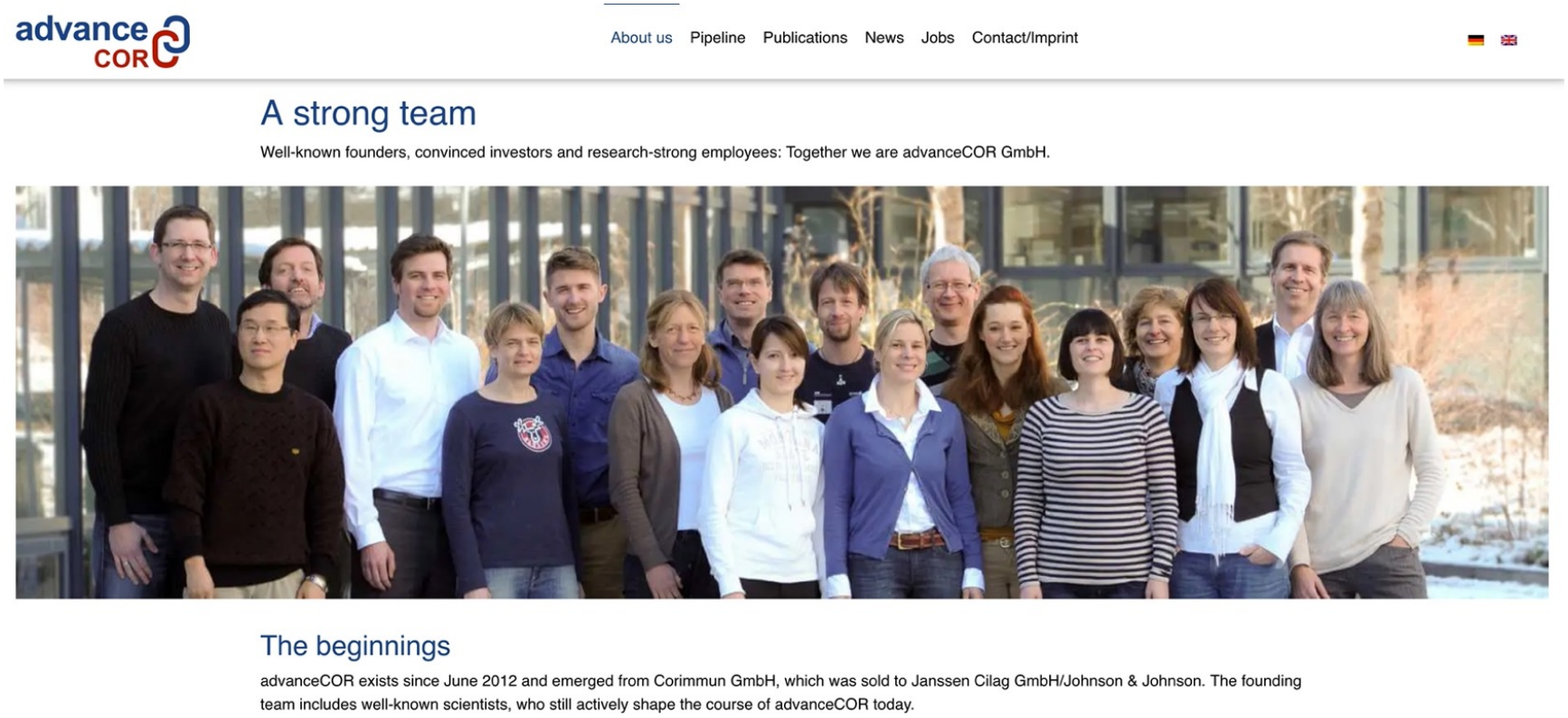
Zuwendungsempfänger: Corimmun GmbH
Förderung: GO-Bio Phase II (01.10.2008 bis 31.03.2011, 2.056.802 Euro)

Weitere Informationen

↗ Medizinische Klinik und Poliklinik I, Universitätsklinikum Würzburg
↗ Positive Phase-II-Ergebnisse für Revacept bei AdvanceCor

Example of a successful exist of a Go-Bio project

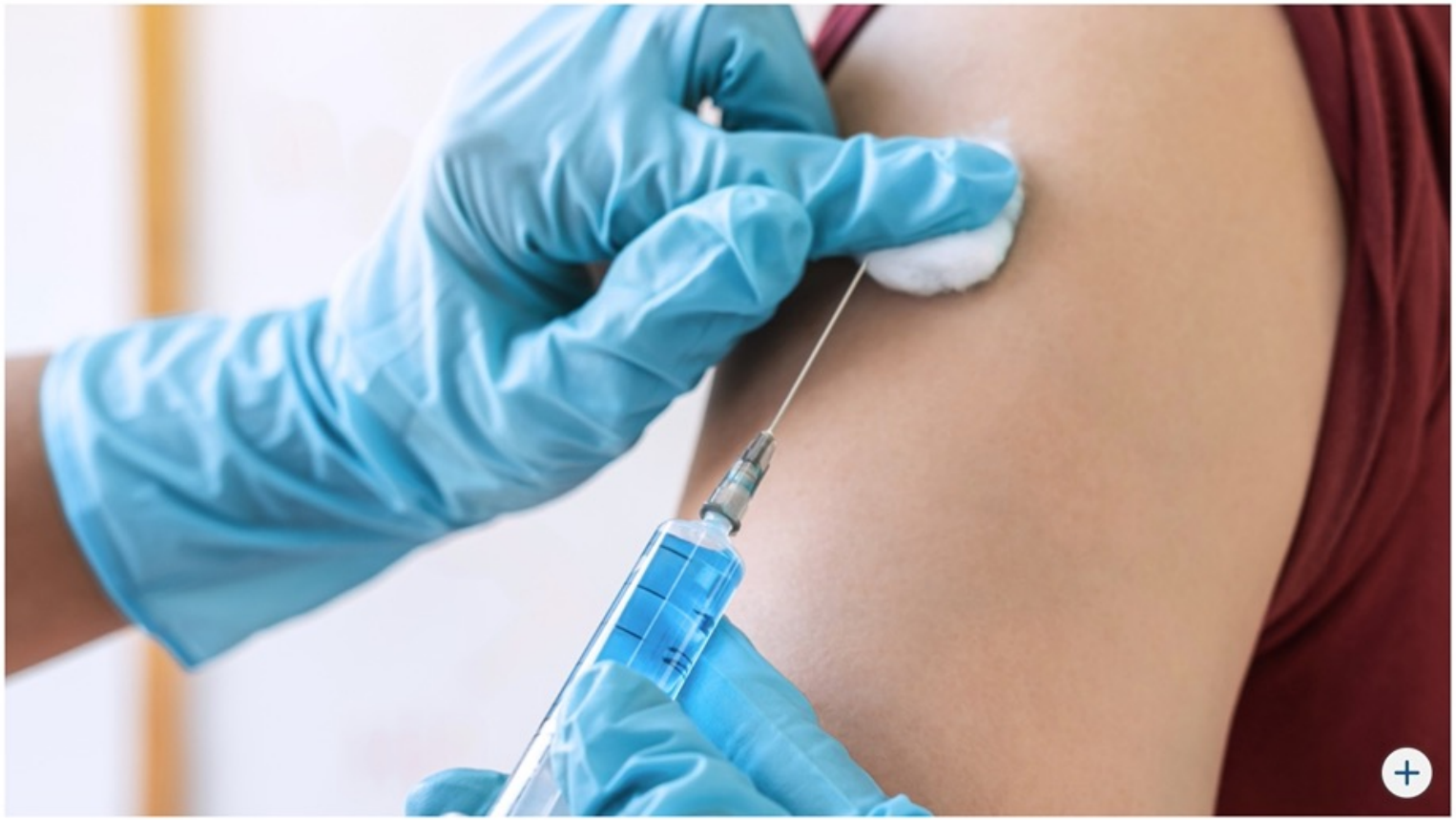
AdvanceCor emerged from Corimmun GmbH after **100 million US dollars (in 2012) investment by Janssen.** The founding of Corimmun was supported by the BMBF as part of the GO-Bio funding.



12.11.2020 | GO-BIO

Entwicklung innovativer Impfstoffe gegen Krebserkrankungen

GO-Bio Runde 1 – Prof. Dr. med. Ugur Sahin – Experimentelle Onkologie, III. Medizinische Klinik, Johannes Gutenberg-Universität Mainz | BioNTech RNA Pharmaceuticals GmbH



© Freedomz - stock.adobe.com

Zuwendungsempfänger: Universitätsmedizin der Johannes Gutenberg-Universität Mainz
Förderung: GO-Bio Phase I (1.3.2007 bis 28.2.2010, 1.242.119 Euro)

Zuwendungsempfänger: BioNTech RNA Pharmaceuticals GmbH
Förderung: GO-Bio Phase II (1.3.2010 bis 31.12.2013, 2.892.862 Euro)

Weitere Informationen

↗ III. Medizinische Klinik, Johannes Gutenberg-Universität Mainz
↗ BioNTech RNA Pharmaceuticals GmbH

Ugur Sahin · 1st

Professor for Translational Oncology and Immunology at University Medical Center Mainz

University Medical Center Mainz

University of Cologne

Top Voice

Mainz, Rhineland-Palatinate, Germany · Contact info

204,857 followers

Dr. Philipp Julian Köster, MBA, Raymond Barlow, and 54 other mutual connections

Message More

10.10.2019

BioNTech an der Nasdaq

Der Mainzer BioNTech SE gelang der Sprung an die amerikanische Technologiebörse Nasdaq. Das Unternehmen nahm mit seinem Börsendebüt 150 Mio. US-Dollar ein.

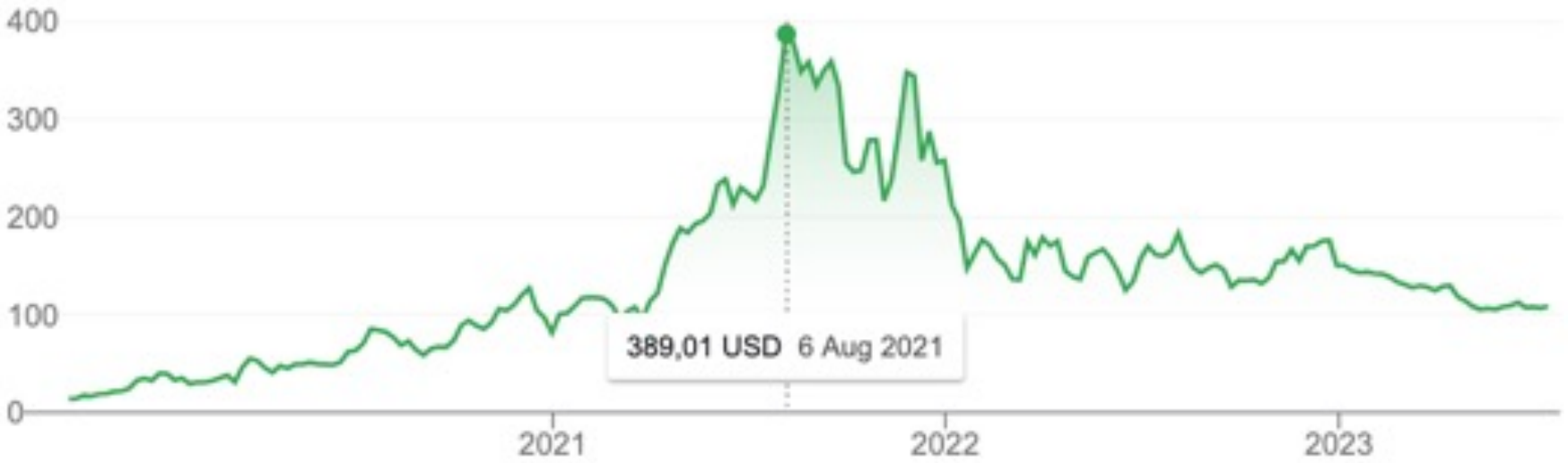
Market Summary > **BioNTech SE - ADR**

108,57 USD

+94.75 (685.60%) ↑ past 5 years

10 Jul, 10:10 GMT-4 • Disclaimer

1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max



Open	107,62	Mkt cap	26,23B	52-wk high	188,98
High	108,90	P/E ratio	3,99	52-wk low	100,08
Low	107,27	Div yield	-		

04.10.2006 | GO-BIO

Die Toxizität von Wirkstoffen und Chemikalien berechnen

GO-Bio Runde 1 – Dr. Igor Tetko – Institut für Bioinformatik und Systembiologie (IBIS), Helmholtz Zentrum München | eADMET GmbH

© Evgeny Rannev - Fotolia

Zuwendungsempfänger: Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)
Förderung: GO-Bio Phase I (01.04.2007 bis 31.03.2011, 705.742 Euro)

Zuwendungsempfänger: eADMET GmbH
Förderung: GO-Bio Phase II (01.04.2011 bis 31.03.2014, 809.844 Euro)

Weitere Informationen

↗ Institut für Bioinformatik und Systembiologie (IBIS)
↗ eADMET GmbH

Go-Bio 1st round, Award ceremony, 2006

Özlem Türeci, Roland Jahns, Annette Schavan

Topics from the Go-Bio GründungsGespräche

GG1 - Nov. 2007 - IP-strategy and coaching
GG2 – March 2008 - Business plan and preparations of financing negotiations
GG3 - Nov. 2008 - The first round of financing - requirements and factors of success
GG4 - March 2009 - Clinical studies – an introduction
GG5 - Nov. 2009 - Management
GG6 - March 2010 - Quality management
GG7 - Nov. 2010 - Business game "Risky Business" (development of pharmaceuticals)
GG8 - May 2011 - Negotiation management
GG9 - Dec. 2011 - Contract design
GG10 - May 2012 - Presentation technique, customer and investor approach
GG11 - Dec. 2012 - IP and company takeover
GG12 - Apr. 2013 - Experience reports from GO-Bio founders
GG13 - Dec. 2013 - Team development and HR management
GG14 - May 2014 -Coaching
GG15 - Dec. 2014 - Financing
GG16 - May 2015 - Governance, risk management & compliance (GRC)
GG17 - Nov. 2015 - Company simulation
GG18 - June 2016 - Contract design and team pitch
GG19 - Dec. 2016 - Experience reports
GG20 - May 2017 - IPO
GG21 - Dec. 2017 - Corporate Communications
GG22 - June 2018 - Clinical Studies
GG23 - Dec. 2018 - Business simulation game
GG24 - May 2019 - Coaching, business etiquette, team development
GG25 - Nov. 2019 - Contract law
GG26 - Nov. 2020 - IP-strategy
GG27 – June 2021 - Financing
GG28 – Dec. 2021 - Business plan
GG29 - Jun 2022 – Contracts and contract negotiations
GG30 - Sep 2022 – Business simulation
GG31 - May 2023 - Team, Team, Team!

GründerGespräche

Know-how: The 'GründungsGespräche' provide the GO-Bio teams with helpful and practical information to prepare them for setting up a
running their company.

A chronology of entrepreneurship

2007 Bundesministerium für Bildung und Forschung — GO-Bio

2009 Helmholtz Enterprise

2010 eADMET — THE REFERENCE IN CHEMOINFORMATICS

2011 HelmholtzZentrum münchen — German Research Center for Environmental Health; Bundesministerium für Bildung und Forschung — GO-Bio; High-Tech Gründerfonds; BayernKapital — Venture Capital für Bayern

2012 - 14 

2014 **Bankruptcy…**

Google, Roche, TUM School of Management Technische Universität München

2015 big chem

2016 HelmholtzZentrum münchen — German Research Center for Environmental Health; ascenion

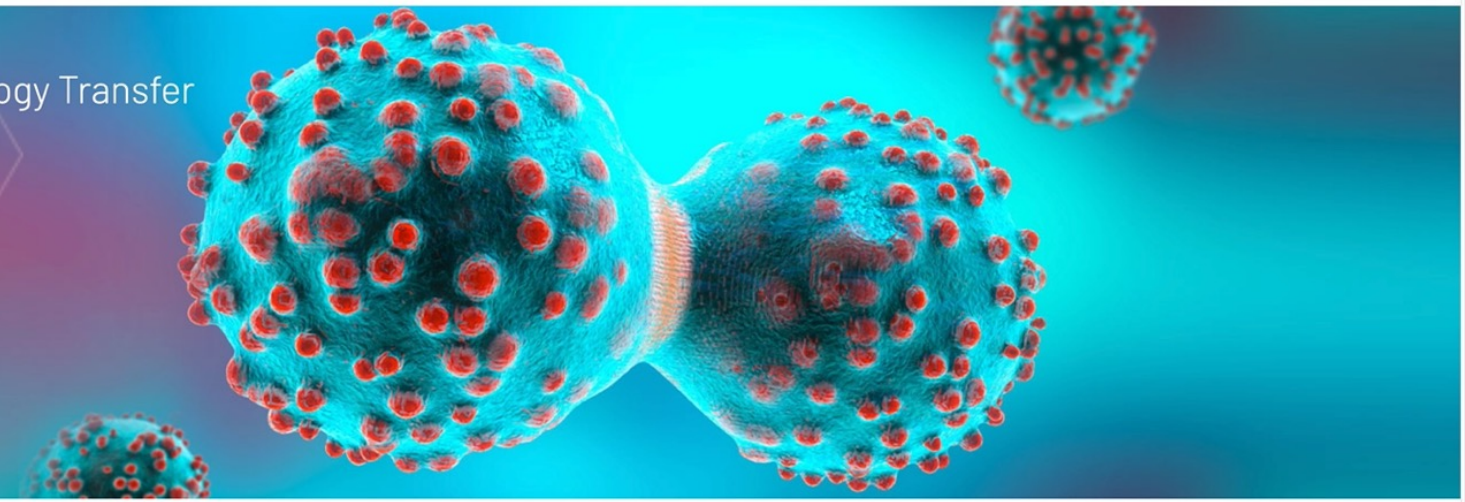
2017+ 

ascenion
A company of the LifeScience Foundation

DE

ABOUT US | PARTNERS | SERVICES | TECHNOLOGY OFFERS | START-UPS | IMPACT | NEWS & EVENTS | CONTACT

Knowledge and Technology Transfer for Academic Life-Science Research



NEWS

28.04.2023

BioVaria Startup Pitch & Partner Awards go to Ceridwen Oncology and Lactabico

28 April 2023, Munich – A record number of over 270 innovators from academia and industry

Read more »

INVENTORS

Protect, develop and commercialize ideas

FOUNDERS

Initiate and finance start-ups

INDUSTRY & INVESTORS

Discover innovation opportunities

OUR IP-PORTFOLIO

Search over 700 technologies and research materials

START-UPS

Start-ups

Getting start-ups off the ground

Supporting start-ups is a particular focus of our work. We have already accompanied over 100 start-up projects. You too can benefit from our experience!

As a company founder:

Together, we create favourable starting conditions, avoiding the mistakes that others have made before. We also connect you to the right people: investors, partners, potential customers.

As a partner institute:

Offer your entrepreneurial scientists optimal support – without having to build up internal resources and know-how. You profit from the proceeds that come from our equity in successful start-ups. These are distributed by the LifeScience Foundation, mainly as grants to be used as desired by the originating institutes.

As an investor:

Discover our attractive start-up projects with a strong IP footing – before they become visible to a broader public at regular industry events.

eADMET core team

Sergii Novotarskyi · 1st
Software Engineer at Facebook
London, England, United Kingdom · **Contact info**
175 connections

Joachim Vogt · 1st
Director, Search and Evaluation International at Abl
Ludwigshafen Am Rhein, Rhineland-Palatinate, Germany ·
500+ connections

Yurii Sushko · 1st
Software Engineering Manager at Google
Greater Munich Metropolitan Area · **Contact info**
365 connections

Dr. Ahmed Sayed, MBA (He/Him) · 1st
EMEA Leader - Prof. Services Emerging Tech (AI/ML, Analytics, IoT, Blockchain, ...)
Greater Munich Metropolitan Area · **Contact info**

Ascenion, eADMET and BIGCHEM (all GmbH) are spin-offs of the center

HELMHOLTZ MUNICH

ENG | DE

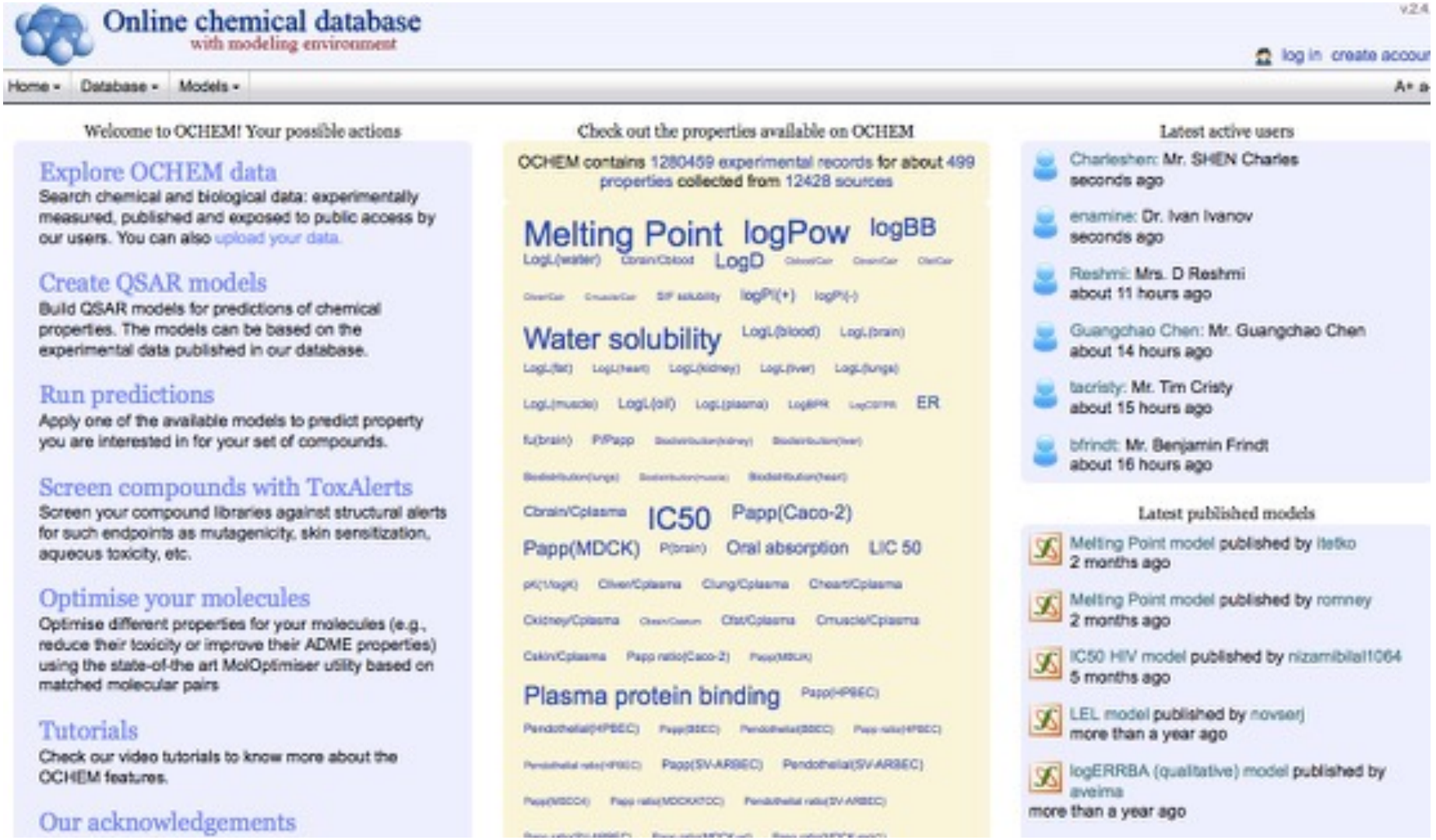
ENG | DE

Mission & Vision | Management | Numbers | Helmholtz Member | Locations | How we work | Organization | Tenders

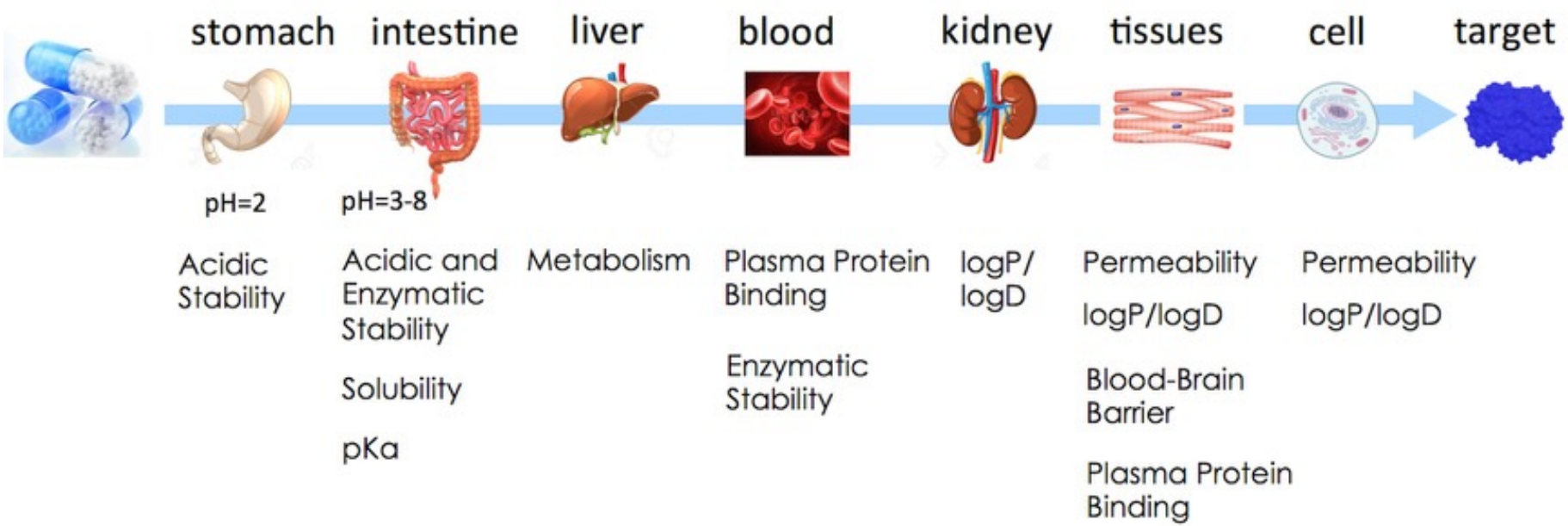
2.475	85	1.604	650
Employees	Nations	Publications in 2021	PhD students

27	48
Spin-Offs	ERC Grants

Data storage and model development: http://ochem.eu



Physiological and physical–chemical barriers affecting a drug bioavailability.



Adapted from Kerns, E. H.; Di, L. Pharmaceutical Profiling in Drug Discovery. Drug Discov. Today 2003, 8, 316–323. Copyright (2003), with permission from Elsevier.

See also Ratkova, E. L. et al Empirical and Physics-Based Calculations of Physical–Chemical Properties. In *Comprehensive Medicinal Chemistry III*, Chackalamannil, S.; Rotella, D. P.; Ward, S., Eds.; Elsevier: Oxford, 2017; Vol. 3, pp 393-428.

ADMETox filters in Bayer

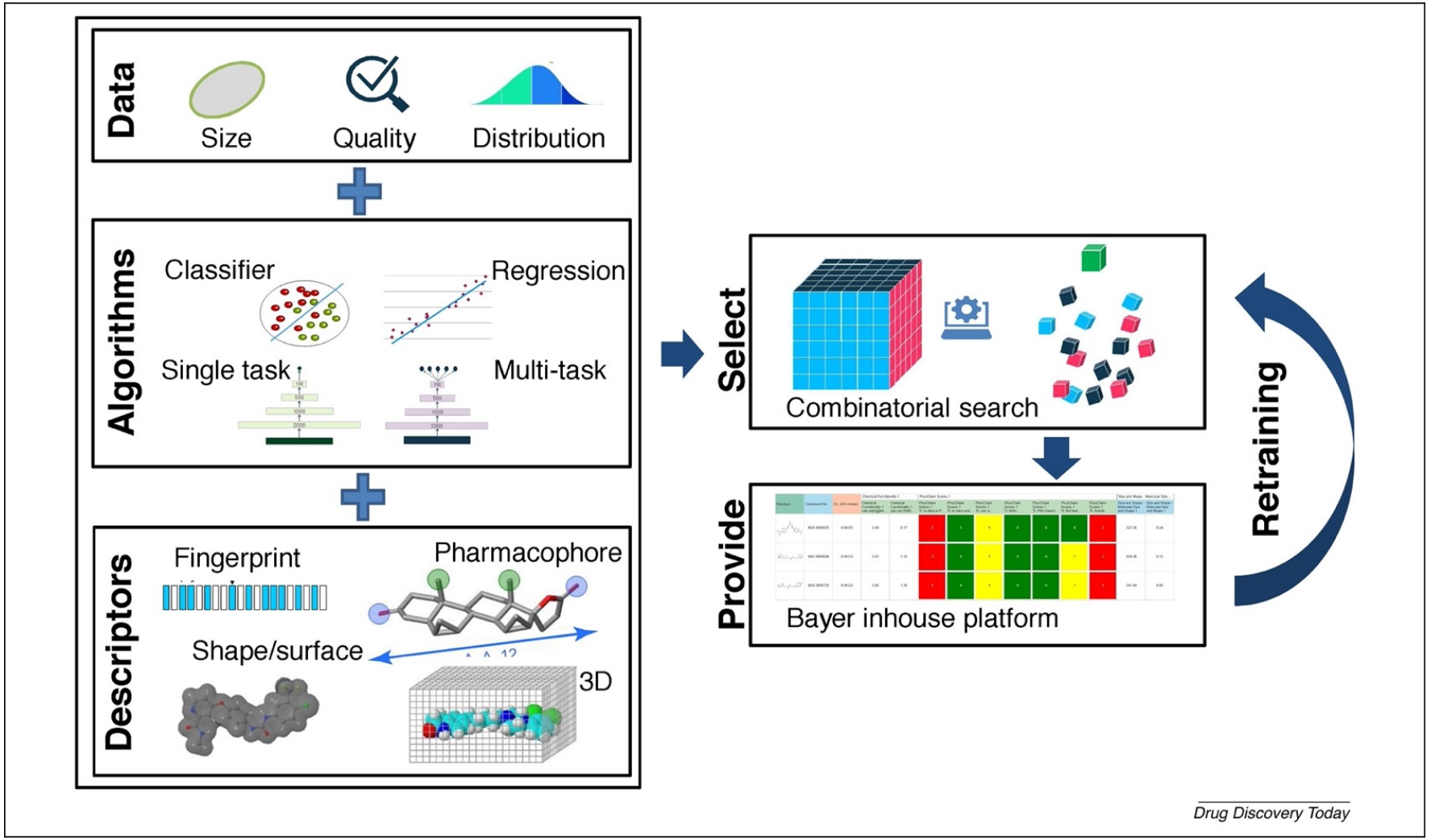
Insufficient quality | First approach | Medium model | Good model | Robust model

Endpoint		Model type	Data set size	2005	2009	2014	2019	Retraining
Absorption	Caco-2 permeation	C (N)	>10 000			RF	SVR	Weekly
	Caco-2 efflux	C (N)	>10 000			RF	SVR	Weekly
	Bioavailability (rat)	C	~2000				RF	On demand
Distribution	Human serum albumin	N	>30 000			PLS	MTNN	On demand
	Fraction unbound	N	>1000			PLS	MTNN	On demand
Metabolism	Microsomal stability (hum)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (mouse)	C (N)	>10 000			RF	RF	Weekly
	Microsomal stability (rat)	C (N)	>10 000			RF	RF	Weekly
	Hepatocyte stability (rat)	C (N)	>30 000			RF	RF	Weekly
Toxicity	hERG inhibition	C	>10 000			RF	SVM	Weekly
	Ames mutagenicity	C	>10 000			RF	RF	On demand
	CYP inhibition isoforms	C	>10 000			RF	RF	On demand
	Phospholipidosis	C	<1000			SVM	SVM	On demand
	Structure filter tool	Score	n.a.	-	-	-	-	On demand
PhysChem	Solubility (DMSO)	N	>30 ,000			PLS	MTNN	On demand
	Solubility (Powder)	N	<10 000				MTNN	On demand
	logD @ pH 7.5	N	>70 000			PLS	MTNN	On demand
	Membrane affinity	N	<10 000			PLS	MTNN	On demand
	pKa	N	>10 000			ANN	ANN	On demand
	Oral PhysChem score	Score	n.a.	-	-	-	-	On demand
	i.v. PhysChem score	Score	n.a.	-	-	-	-	On demand

Drug Discovery Today

Göller, AH et al *Drug Discov. Today* **2020,** *25* (9), 1702-1709.

Bayer workflow for model life cycle



Göller, A.H. et al. *Drug Discov. Today* **2020,** *25* (9), 1702-1709.

eADMET
THE REFERENCE IN CHEMOINFORMATICS

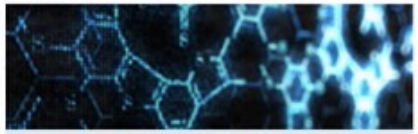
Home Company Products Services News Contact *Support*

eADMET GmbH offers and innovative IT solutions to predict important properties of chemicals and drugs - physical and so-called ADME/Tox- ("drug-like") characteristics. The central product of eADMET is "OCHEM", a Web-based database and modeling system that ...

read more

Meet us at

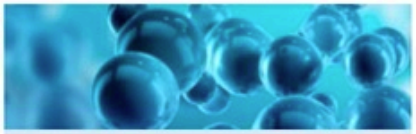
- **lecture "On-line Chemical Modelling Environment" at Enamine Ltd.**
  November 1, 2013; Kyiv, Ukraine
- **9th German Conference on Chemoinformatics**
  November 10-12, 2013 ; Fulda, Germany
- **Big data in Life Science and ICT**
  November 12, 2013; Munich, Germany
- **Third investment Forum**
  November 19, 2013; Berlin, Germany
- **2nd Workshop Structural Biology in the Helmholtz Association**
  November 28-29, 2013; ; Munich, Germany



OCHEM

OCHEM is our platform for the creation of in silico ADME / Tox prediction models. We offer a free version of the modeling development environment for academic use.

Free Version



ePhysChem

Using ePhysChem the physical properties of substances can be assessed quickly. ePhysChem contains AlogPS 3.01 for logP and logS prediction and many other models. Try our free version!

Free Version

Training

Our next in silico ADME/T training courses start in Winter 2014. If you are interested in a course or have specific questions, please use our contact form or call us. Independant from our scheduled courses we can offer you and your team with individualized training content about QSPR, QSAR and in silico ADME / Tox ...

Consulting

We create models for you and carry out literature and database searches. Get models built on with thousand of external and internal measurements - maximizing the information content of your data and getting a clear view on the achievable accuracy of in silico predictions for your compounds. Quickly and inexpensively.

Key issues for *in-silico* ADME/Tox

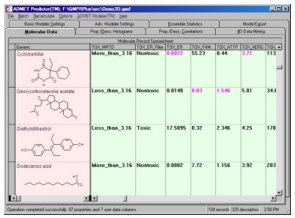
eADMET *in-silico* tools

- Need structure information only, Identify poor predictions.
- Saves time and costs in R&D
- Fail early, fail fast, fail cheap

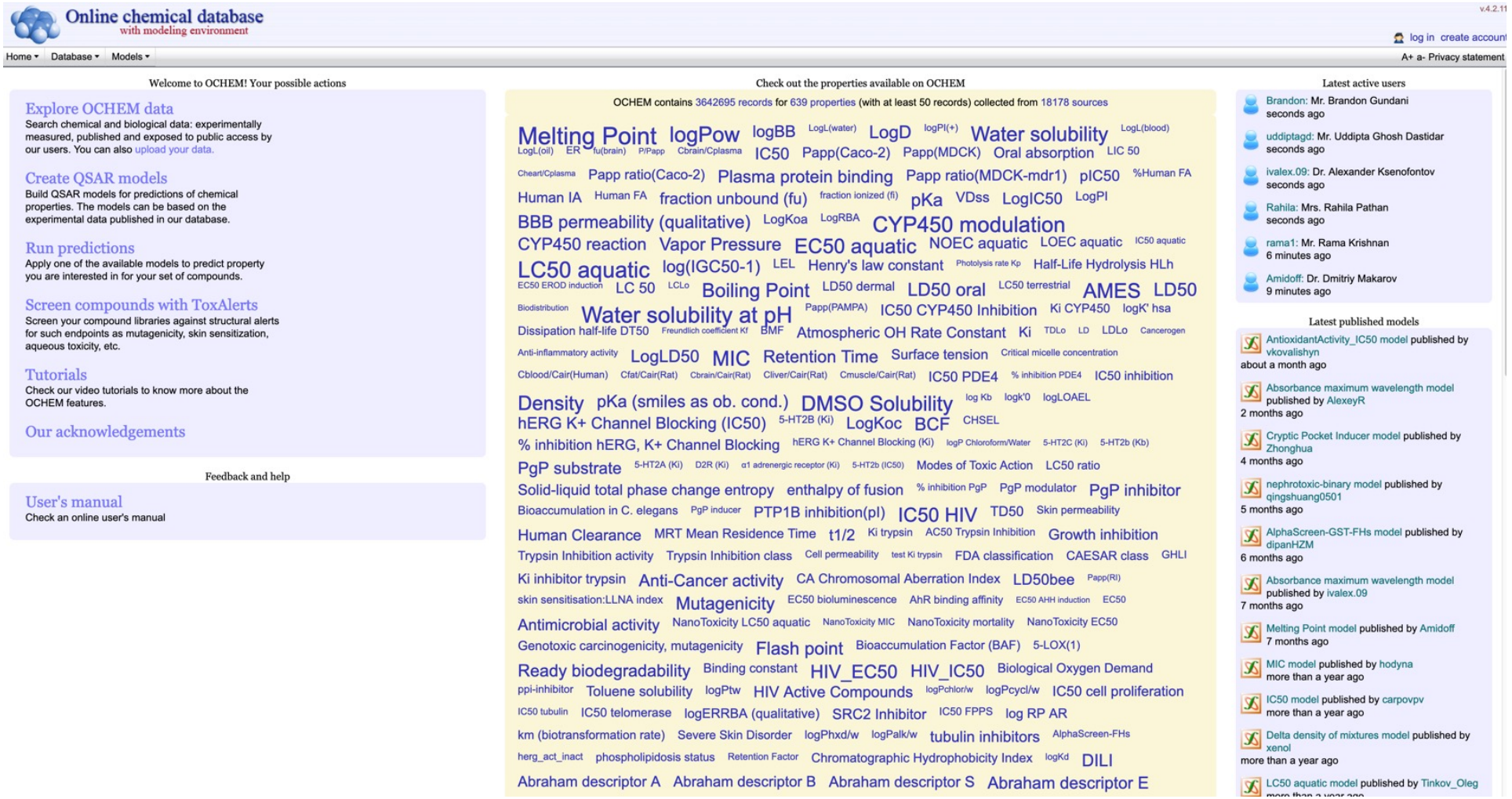
but

In-vitro and *in-vivo* testing of actual compounds are considered more reliable than *in-silico* tools

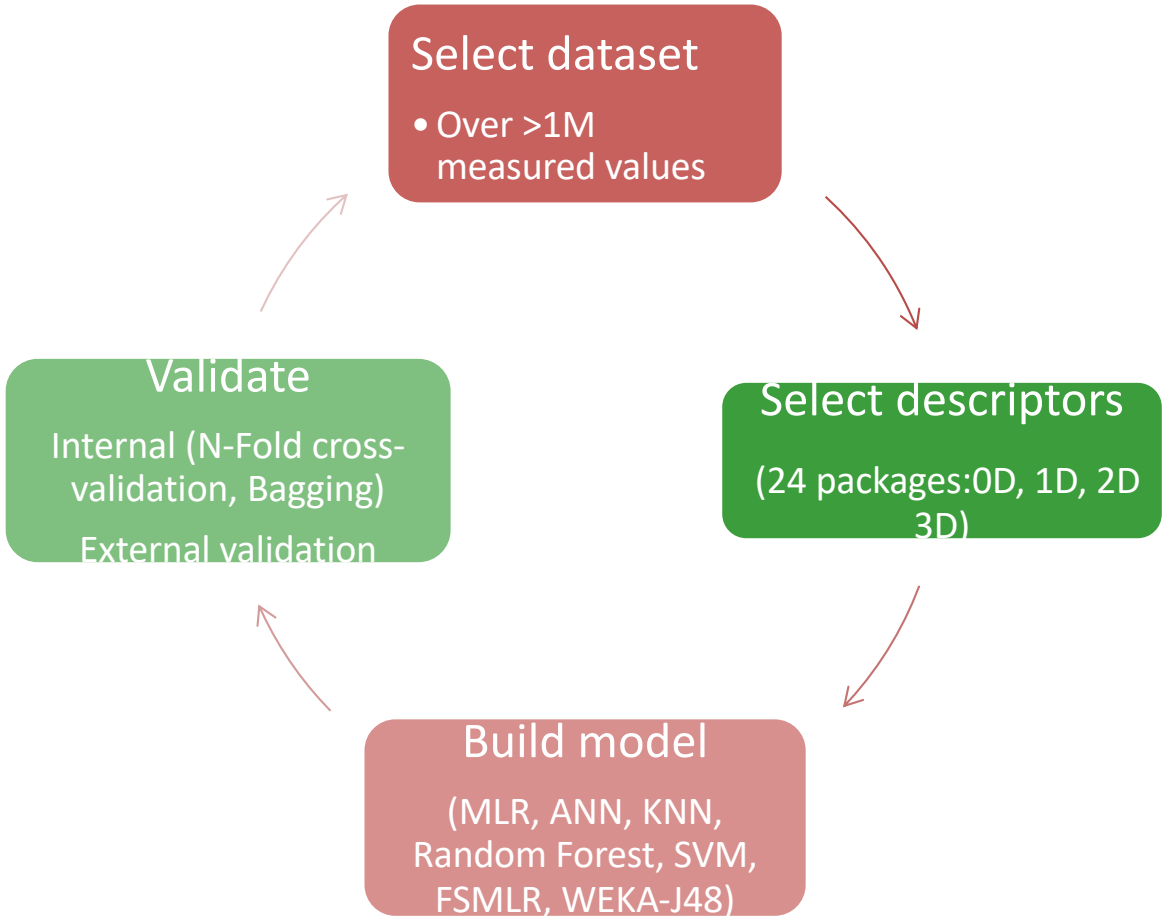
Key feature: accuracy estimation

Method	Costs	Time	Quality
Blind *in-vitro* testing of all compounds 			
Blind *in-silico* prediction of all compounds 			
eADMET: *in-silico* prediction plus intelligent testing of selected compounds  + 			

Data storage and model development: http://ochem.eu



Modeling iterative workflow



Traditional representation of chemical structures

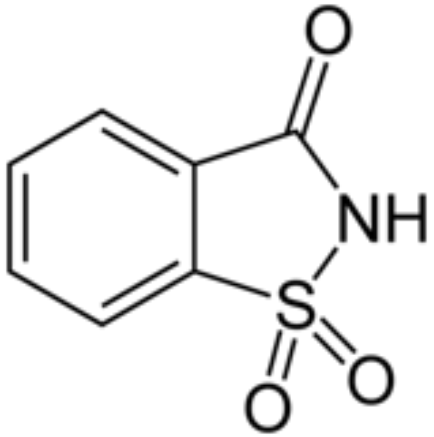
Saccharin

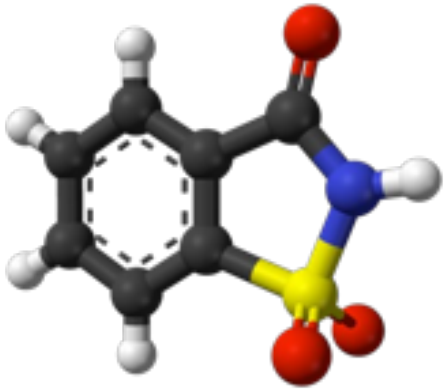
1D

2D

3D

$C_7H_5NO_3S$





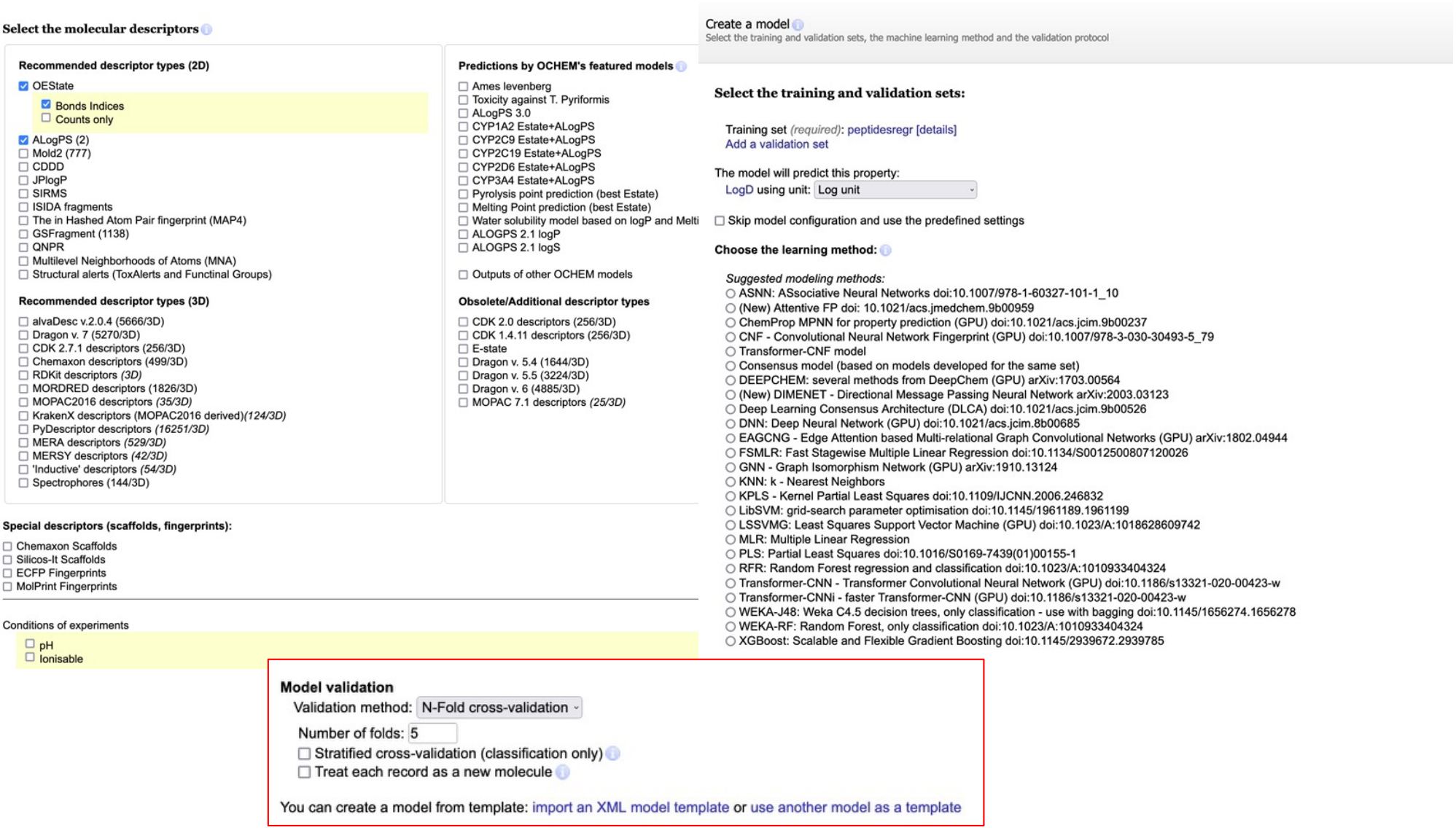
Examples of descriptors

- [x] alvaDesc v.2.0.4 (5666/3D)

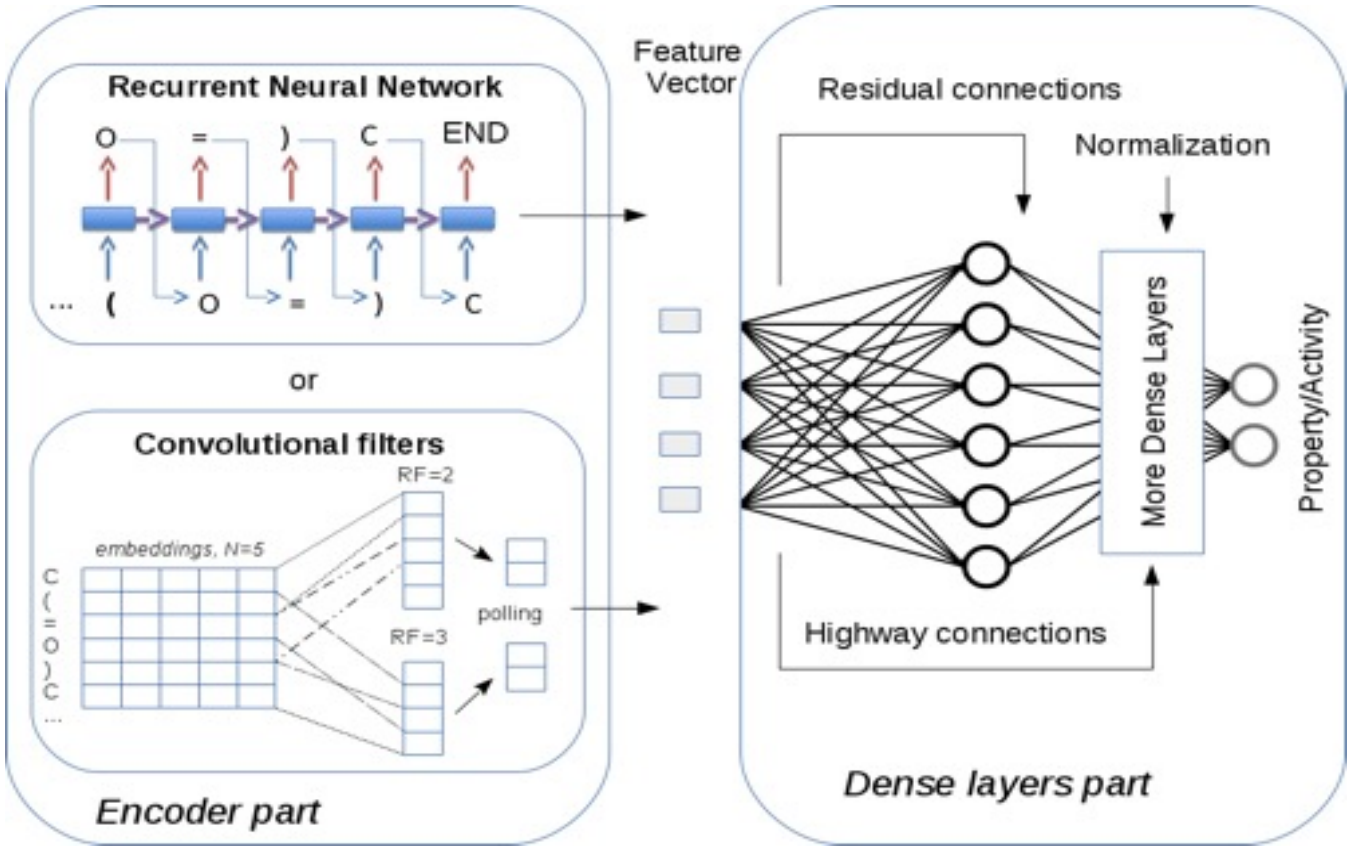
[select all] [select none] [select 3D] [unselect 3D]

- [x] Constitutional descriptors (50)
- [x] Topological indices (79)
- [x] Connectivity indices (37)
- [x] 2D matrix-based descriptors (608)
- [x] Burden eigenvalues (96)
- [x] ETA indices (40)
- [x] Geometrical descriptors (3D, 38)
- [x] 3D autocorrelations (3D, 80)
- [x] 3D-MoRSE descriptors (3D, 224)
- [x] GETAWAY descriptors (3D, 273)
- [x] Functional group counts (3D, 154)
- [x] Atom-type E-state indices (346)
- [x] 2D Atom Pairs (1596)
- [x] Charge descriptors (3D, 15)
- [x] Drug-like indices (30)
- [x] WHALES (3D, 33)
- [x] Chirality (70)
- [x] Ring descriptors (35)
- [x] Walk and path counts (46)
- [x] Information indices (51)
- [x] 2D autocorrelations (213)
- [x] P_VSA-like descriptors (69)
- [x] Edge adjacency indices (324)
- [x] 3D matrix-based descriptors (3D, 132)
- [x] RDF descriptors (3D, 210)
- [x] WHIM descriptors (3D, 114)
- [x] Randic molecular profiles (3D, 41)
- [x] Atom-centred fragments (115)
- [x] Pharmacophore descriptors (165)
- [x] 3D Atom Pairs (3D, 36)
- [x] Molecular properties (3D, 27)
- [x] CATS 3D (3D, 300)
- [x] MDE (19)

QSPR/QSAR modelling in OCHEM



Machine Learning directly from chemical structures



P. Karpov, G. Godin, I. V. Tetko, *J. Cheminform.* **2020**, *12*, 17.

https://github.com/bigchem/transformer-cnn

Winning model: OCHEM-generated consensus model

Andrea Kopp

SLAS Europe 2023

25.05.2023

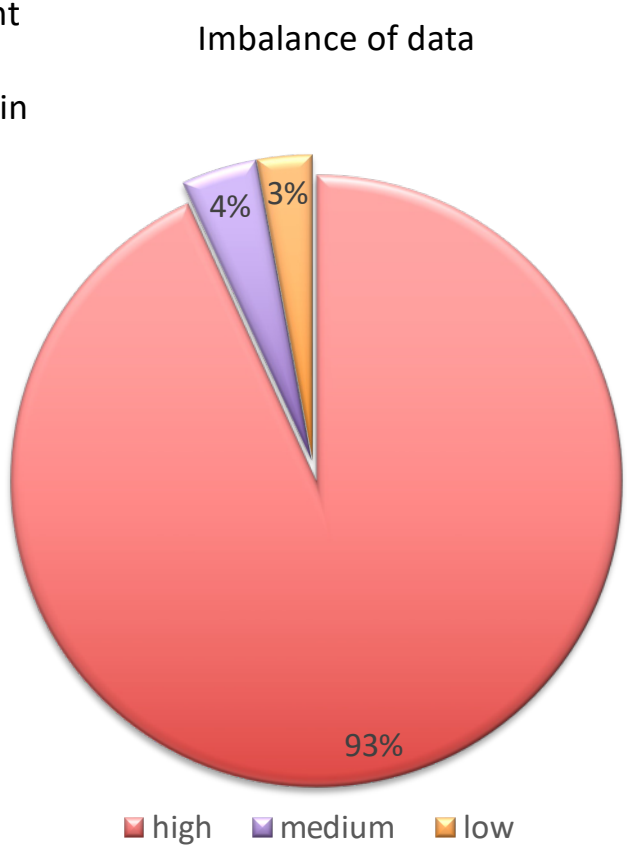
HELMHOLTZ MUNICH

Team of Igor Tetko with Peter Hartog, Martin Šícho and Guillaume Godin

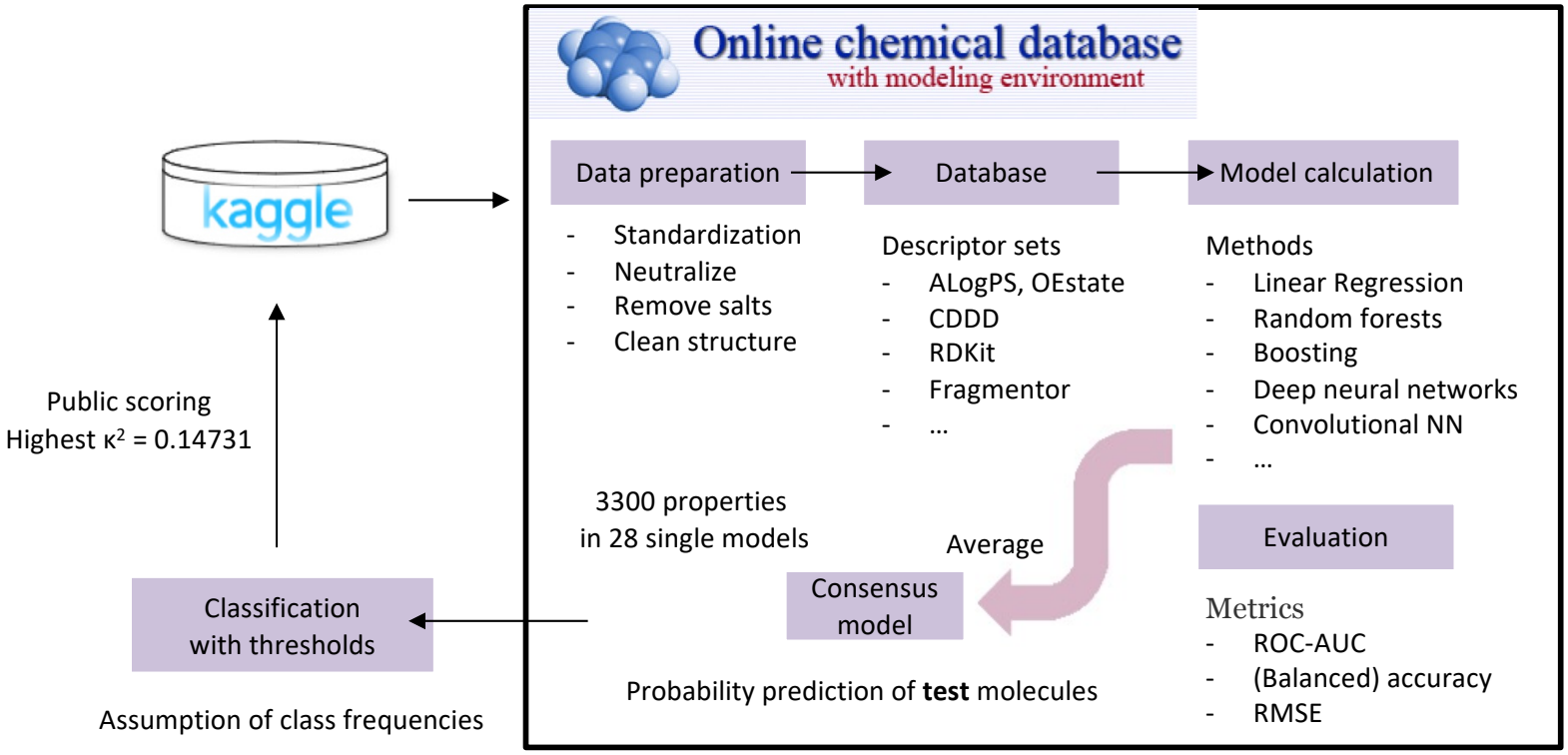
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Challenge set-up

- Experimentally: Nephelometer measures undissolved sediment
- Classification into *low*, *medium* and *high* soluble with phenytoin and amiodarone as thresholds
- 70k training datapoints, 15k public leaderboard, 15k private leaderboard
- Stratified random sampling



Workflow with OCHEM



OCHEM for modeling

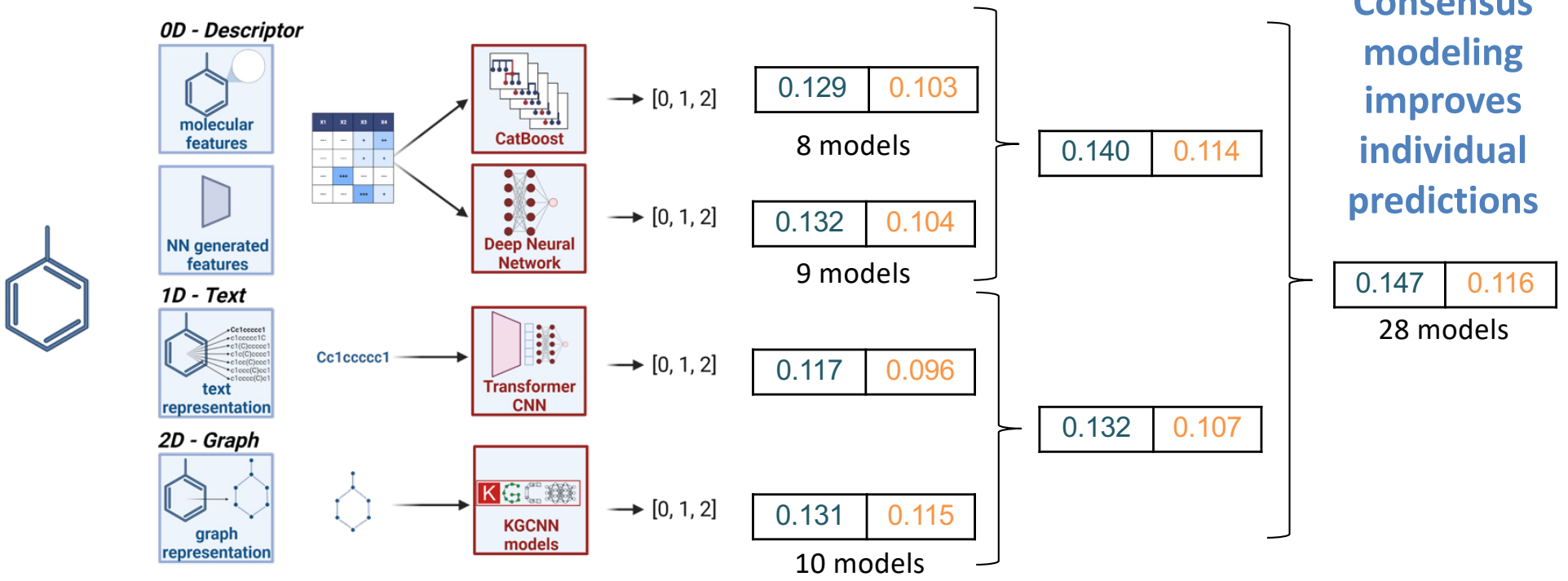
- Graphical interface allows comprehensive modeling without explicit coding
- Implementation for GPU and CPU use
- Consensus models:
  - Average over multiple models to improve prediction
  - Orthogonal models
  - Various descriptor sets/ molecular representations

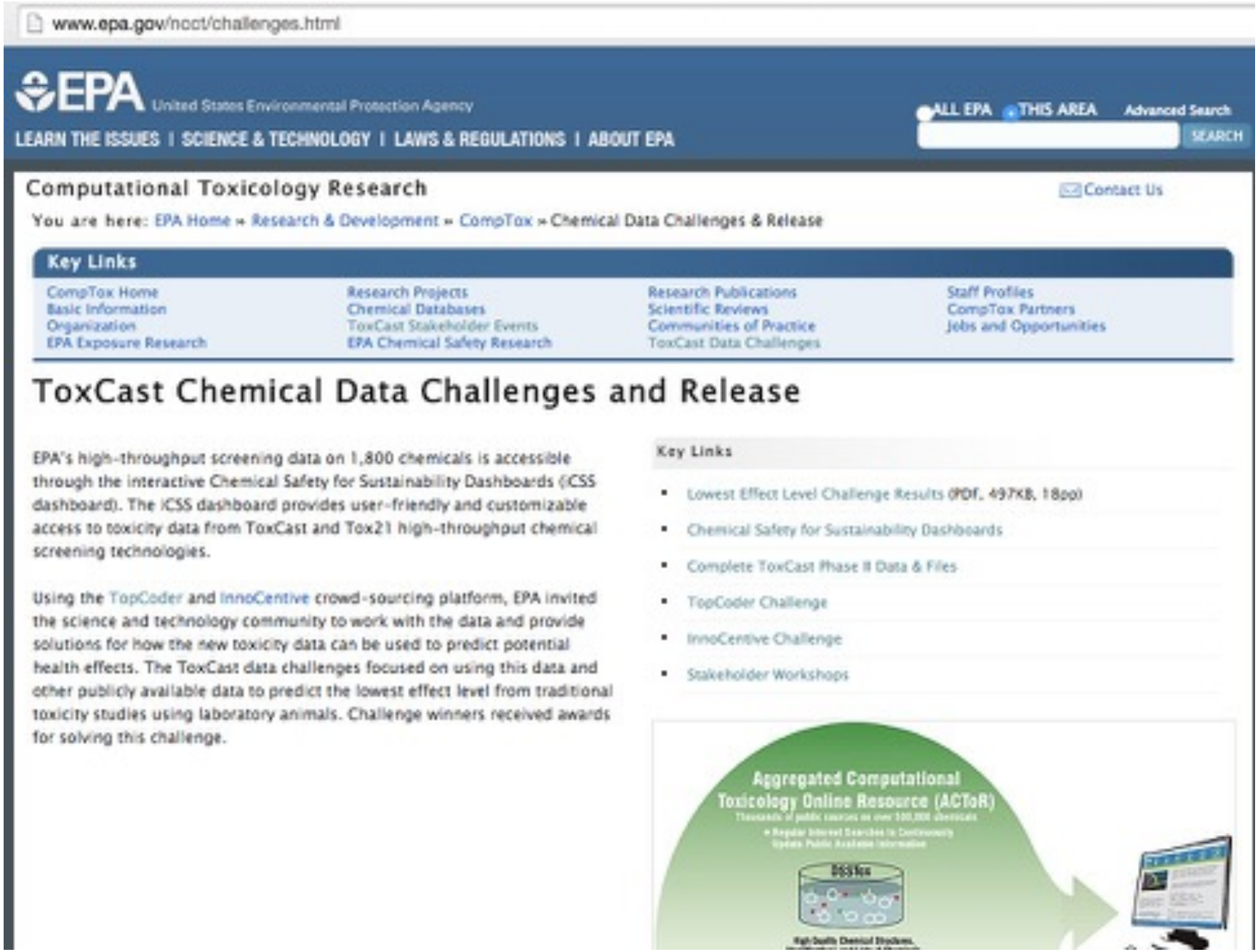
Metrics: AUC for Training set Validation: Cross-Validation (84 models)

	LSSVMG	ASNN	PLS	KNN
ALogPS, OEstate	0.74	0.68	0.61	0.64
CDDD	0.8	0.74	0.75	0.71
CDK2 (cons,topol,geom,elec,hybrid) 3D:corina	0.75	0.71	0.56	0.71
ChemaxonDescriptors (pH 0 - 14:1) 3D:corina	0.76	0.7	0.59	0.68
Dragon6 (2D blocks: 1 28)	0.64	0.66	0.59	0.65
Dragon6 (3D blocks: 1-29) 3D:corina	0.76	0.72	0.57	0.65
Fragmentor (length:2 - 4)	0.72	0.7	0.59	0.63
GSFrag (F + L)	0.69	0.69	0.61	0.61
InductiveDescriptors 3D:corina	0.69	0.71	0.57	0.67
JPlogP	0.73	0.74	0.59	0.67
MAP4	0.71	0.65	0.59	0.67
MORDRED (All) 3D:corina	0.77	0.73	0.57	0.68
Mera, Mersy 3D:corina	0.73	0.69	0.55	0.67
OEstate	0.74	0.67	0.63	0.68
PyDescriptor 3D:corina	0.71	0.71	0.7	0.67
QNPR (length:1 - 3)	0.68	0.62	0.58	0.58
RDKIT (3D blocks: 1-11 15-16) 3D:corina	0.77	0.72	0.56	0.65
SIRMS (labels:charge+logp+hb+refractivity)	0.76	0.73	0.59	0.67
Spectrophores (accuracy=20) 3D:corina	0.68	0.6	0.52	0.6
StructuralAlerts	0.67	0.64	0.58	0.51
alvaDesc (3D blocks: (only) 1-30) 3D:corina	0.75	0.71	0.57	0.68

Quadratic kappa metric scores →

Public leaderboard	Private leaderboard





Novotarskyi, S. et al. *Chem. Res. Toxicol*. 2016, 29, 768-75.

National Center for Advancing Translational Sciences

Tox21 Data Challenge 2014

Login

Home | About | Registration | Data/Resources | Submissions | Discussion | Leaderboard | Contact Us | Survey NEW

Open call ends: November 14, 2014

About the Data

The Challenge

The 2014 Tox21 data challenge is designed to help scientists understand the potential of the chemicals and compounds being tested through the Toxicology in the 21st Century initiative to disrupt biological pathways in ways that may result in toxic effects.

The goal of the challenge is to "crowdsource"

All challenge winners will receive the opportunity to submit a paper for publication in a special thematic issue of *Frontiers in Environmental Science* and recognition on the NCATS website and via social media.

openOCHEM https://github.com/openochem

Overview | Repositories 3 | Projects | Packages | Stars

openochem / README.md

Open OCHEM -- AI models for drug discovery and environmental chemistry

The Open OCHEM is open source version of the On-line Chemical database and Modelling Environment Platform (http://ochem.eu)

It is a user-contributed repository of referenced experimental data, computational tools and models of ADMET properties of chemical compounds. The OCHEM algorithms can reliably identify compounds predicted with experimental accuracy: there is no need to test them in a lab. The OCHEM can be used for timely and low-cost identification of scaffolds with lower risks of failure due to the unfavorable physico-chemical and/or biological properties. The free open source of OCHEM is a reference system for academic users thus accumulating data and knowledge produced in academia. The developed OCHEM workflow allows an unbiased comparison of different existing and new machine learning algorithms which can be easily integrated in OCHEM by its users.

OCHEM software can be used to develop QSPR and QSAR models for various biological and physico-chemical projects. It can work with millions of molecules and can be configured to use hundreds of CPUs or GPUs. Open OCHEM allows you to install the fully functional version of the software and analyse your data privately. The closed source version is also available from BIGCHEM GmBH and provides several additional optimized software packages which were contributed by the company or its partners.

The open OCHEM currently supports tens methods and descriptors packages, which were developed and contributed by different providers and are distributed under the open source or respective license agreements (most of them are free of charge for academic, educational, recreational or evaluation purposes - check each respective license agreement).

See installation instructions how to install and run open the OCHEM.

We wish you a happy computing!

We sincerely thank Yuriy Sushko, Sergey Novotarskyi, Pavel Karpov, Mark Embrechts, Ivan Khokhlov, Robert Körner, Anil Kumar Pandey, Elena Salmina, Stefan Brandmaier, Larisa Charochkina, Vasyl Kovalishyn, Ahmed Abdelaziz, Matthias Rupp, Dipan Ghosh, Zhonghua Xia, Alli Keys as well as many other current and former members of Tetko's group and eADMET and BIGCHEM GmbH companies for their contributions to the development, testing, use and the feedback.

We also thank developers of CDK, MOPAC2016, KGCNN, OpenBabel, Xemistry, BALLOON, WEKA as well as Vsevolod Tanchuk, Sergey Sosnin, Maxim Fedorov, Peter Ertl, Bruno Bienfait, Ruud van Deursen, Gilles Marcou, Igor Baskin, Artem Cherkasov, Pavel Polishchuk, Eugene Radchenko, Vladimir Palyulin, Vijay Masand, Vishweh Venkatraman, Andrea Mauri, Weida Tong, Huixiao Hong, Todd Martin, Peter Jarowski, Vladimir Poroikov, Dmitriy Filimonov, Atif Raza and many others who contributed modules that are used in the OCHEM.

Questions

1) What kind of difficulties that you have encountered? How did you overcome?
2) How did you develop your career skills(such as technical skills for your job)?
3) How we can learn to think as a researcher? Do you have any tips for that?
4) Have you ever give up anything during your journey as a researcher?
5) Do you have any recommendations to the students who are doing their masters degree before starting to pHd?

Acknowledgements

Andi Kopp
Peter Hartog
Fabian Krüger
Paula Torren-Peraire
Varvara Voinarovska
Katya Ahmad
Marchela Pandelova
Nesma Mousa
Mark Embrechts

Emilio Benfenati
and all colleagues from
CONCERT REACH

Martin Šícho (Uni Leiden)
Guillaume Godin (Firmenich)
Ruud van Deursen (Firmenich)

Michael Sattler (HMGU)

Alexander von Humboldt
Stiftung/Foundation

CONCERTREACH
CONCERTING EXPERIMENTAL DATA
AND IN SILICO MODELS FOR REACH

eeco
environmental chemoinformatics
marie curie initial training network

BigChem