Cheminformatics in drug discovery and public health

Progress and challenges ahead

José L Medina-Franco

DIFACQUIM research group

Department of Pharmacy

School of Chemistry, UNAM, Mexico

XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery

May 26th, 2022

Outline

- Introduction
  - UNAM & research group
  - Chemoinformatics
- Progress in drug discovery and public health
  - Case study: Epigenetic drug discovery; inhibitors of DNA methyltransferases
- Cheminformatics: Challenges ahead
- Summary

National Autonomous University of Mexico

UNAM

- 350,000 students
  - 6,836 students: School of Chemistry
- 12,400 full-time professors
- 2,200 buildings in Mexico and abroad

Main campus in Mexico City

School of Medicine

Main Library

Concert Hall

Soccer Stadium

Extension schools abroad

UNAM LOS ÁNGELES

UNAM - Tucson
Centro de Estudios Mexicanos
Tucson

Universidad Nacional Autónoma de México
Chicago, Illinois

Research group DIFACQUIM
Computer-Aided Drug Design at School of Chemistry

Bachelors

Jocelyn Salazar

Alexis Flores

Claudia D. López

Alejandro Ochoa

Hassan Villegas

MSc

Liliam Martínez (UNAM)

Camila Garibay (UMSNH)

Alberto Zavala (UMSNH)

Eva Palma (UMSNH)

Co-tutor: Dr. Víctor M. Baizabal

Ph.D.

Fernanda Saldivar

Diana Prado

Bárbara Díaz

Ana Chávez

Emma Andrade

Alejandro Gonzalez

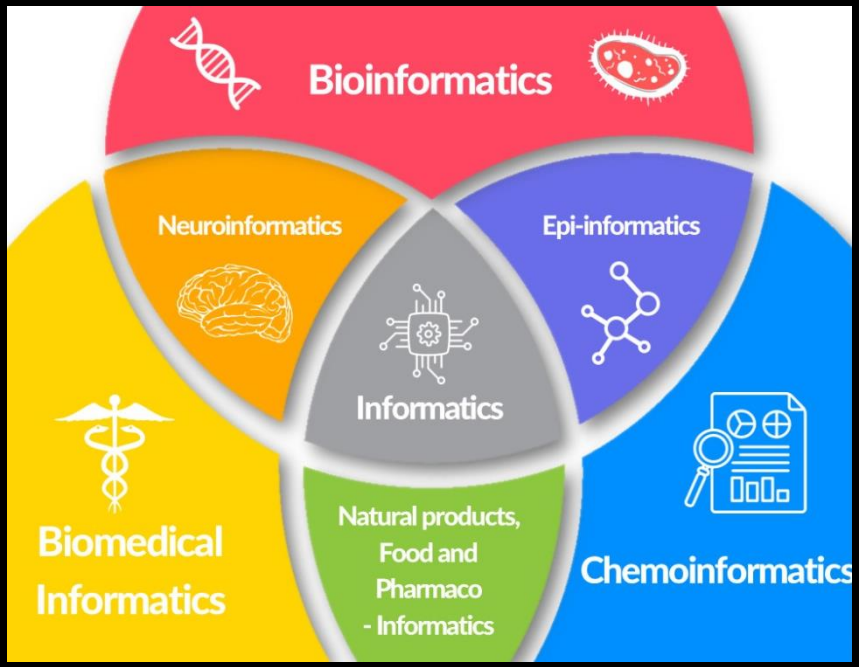
Edgar López (CINVESTAV)

Co-tutor: Dr. Carlos C. Rojas

@difacquim

www.difacquim.com

medinajl@unam.mx

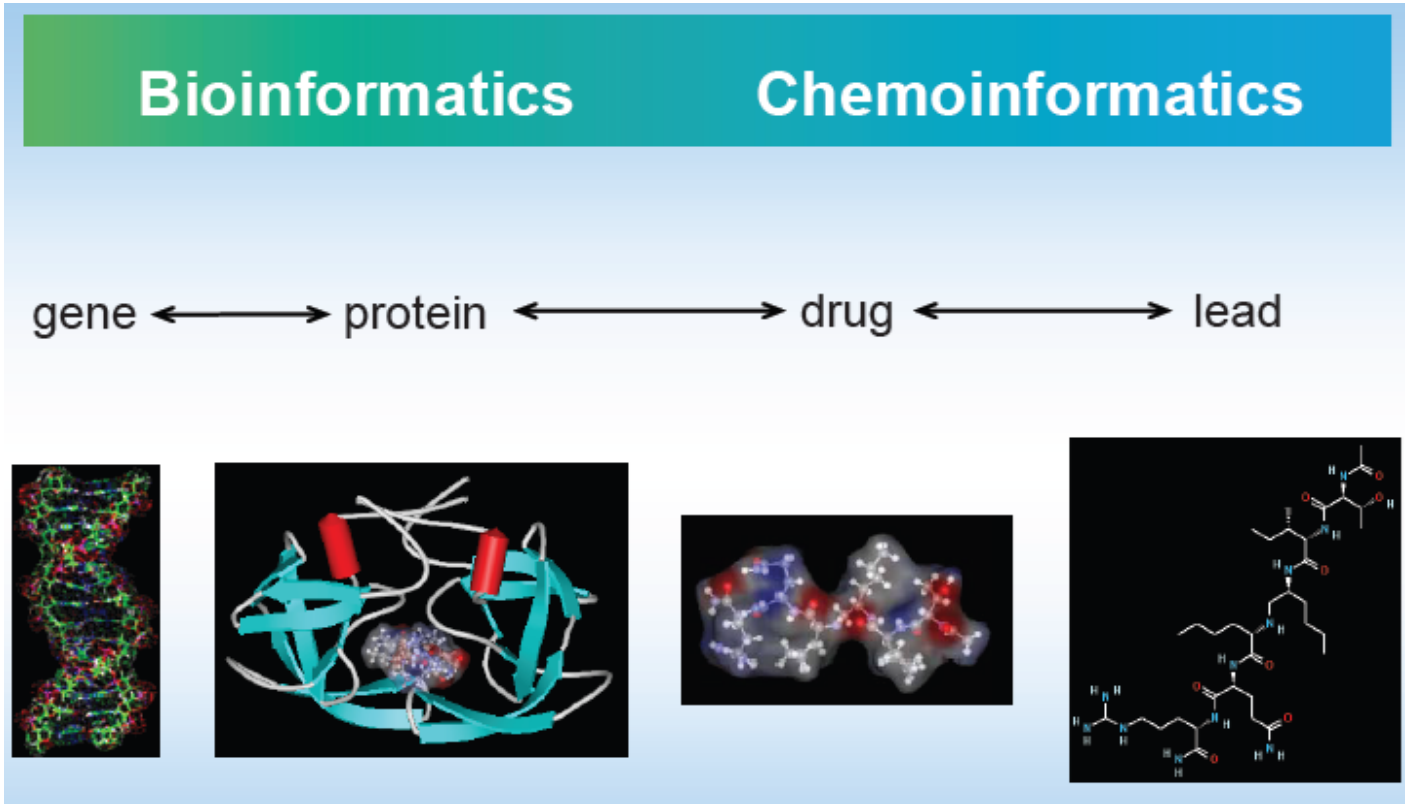


Chemoinformatics

"All concepts and methods that are designed to interface theoretical and experimental efforts involving small molecules"

Drug Discovery Today 2004 9:13–14

Chemoinformatics and bioinformatics

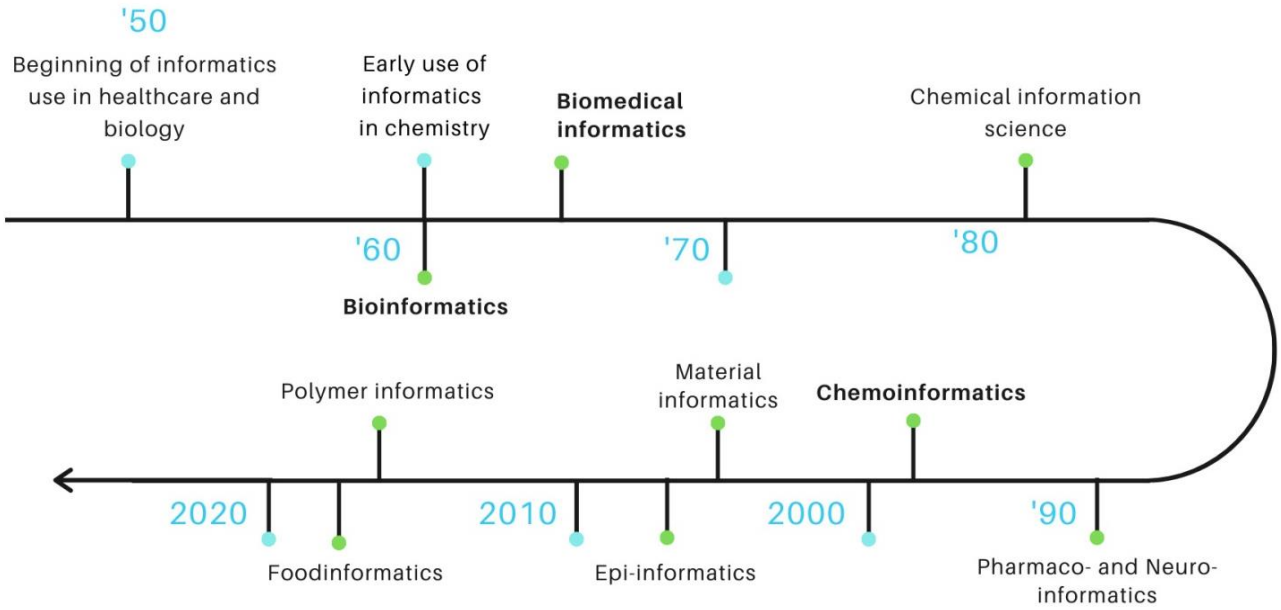


Computational representation

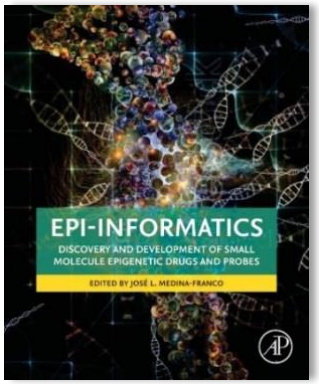
- Sequences
- 3D coordinates

- Molecular fingerprints
- Numerical descriptors.

Timeline and impact of chemoinformatics



Today



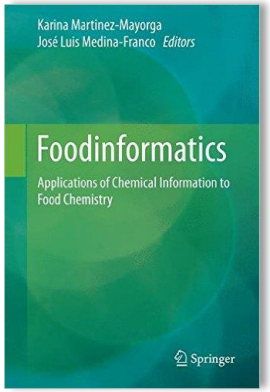
Epigenetic drug discovery

Applications in different areas

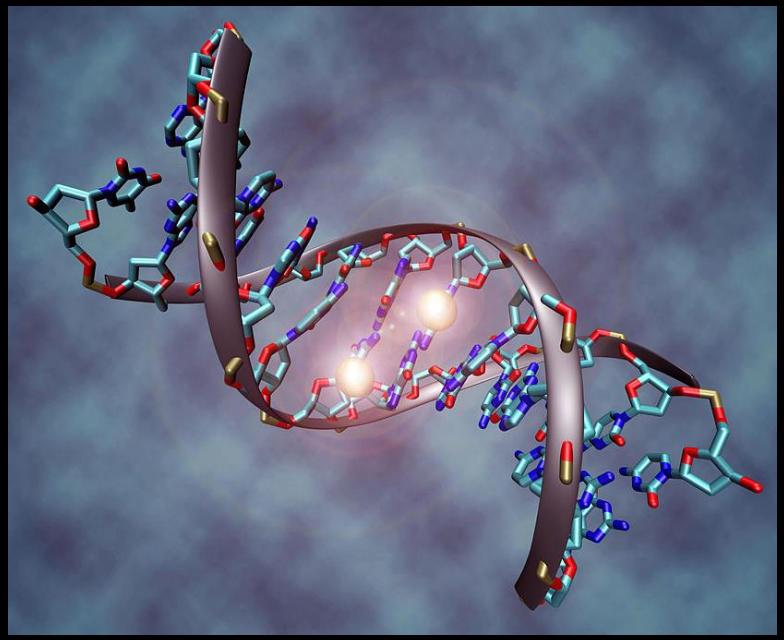
Toxicity

Natural products informatics

Biomedical informatics



Food science



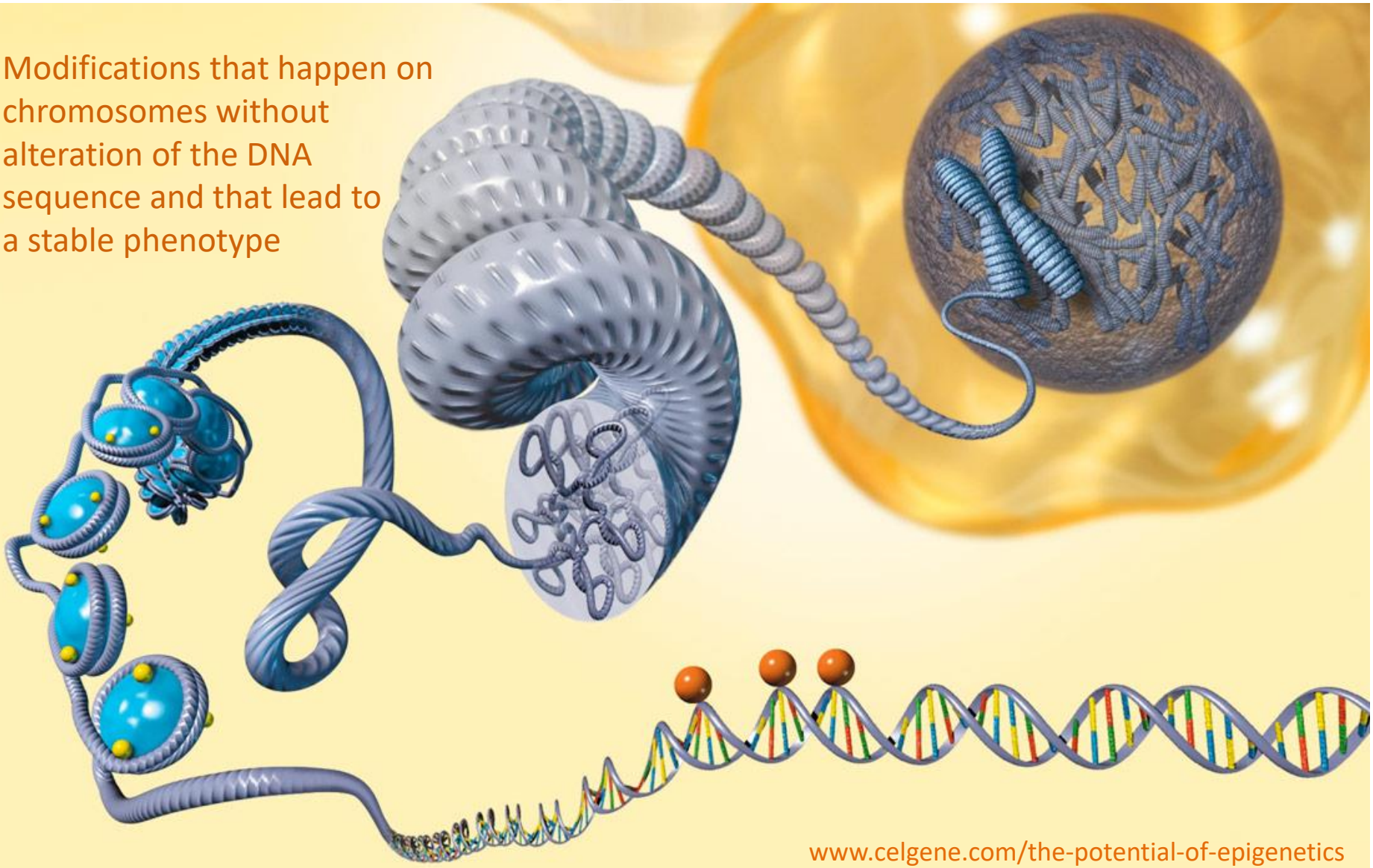
Progress in drug discovery and public health

Epigenetic drug discovery

Inhibitors of DNA methyltransferases

Epigenetics

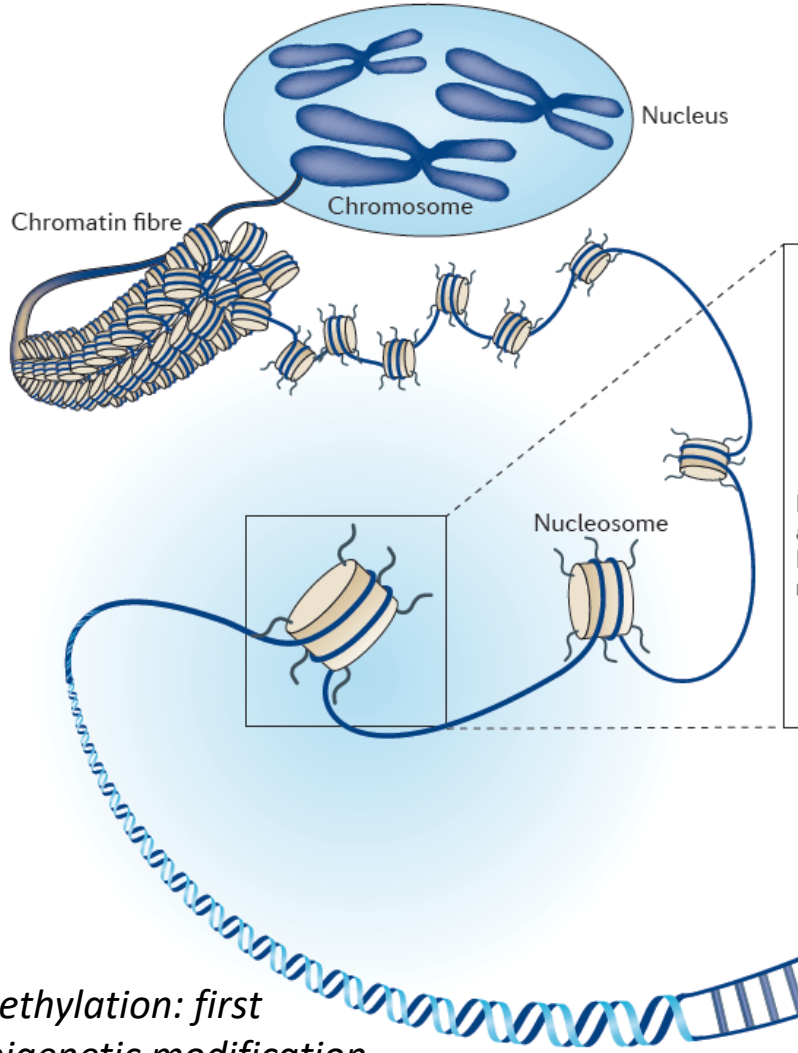
Modifications that happen on chromosomes without alteration of the DNA sequence and that lead to a stable phenotype



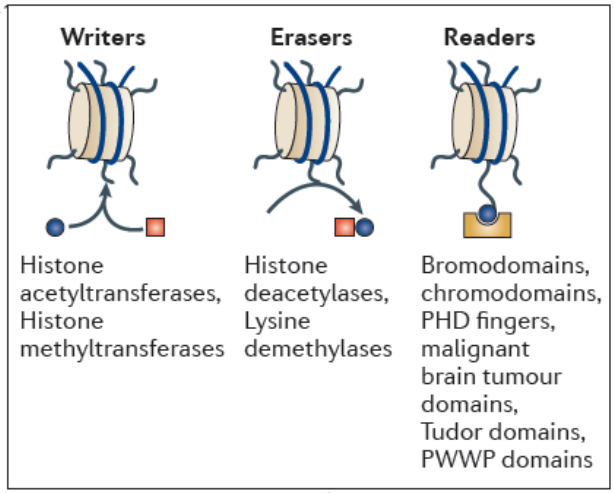
www.celgene.com/the-potential-of-epigenetics

Epigenetic targets

Histone methylation and acetylation are frequently dysregulated in cancer cells and other diseases



Epigenetic protein families



3 active DNMTs in eukaryotes:

- DNMT3A
- DNMT3B

} de novo methylation

- DNMT1

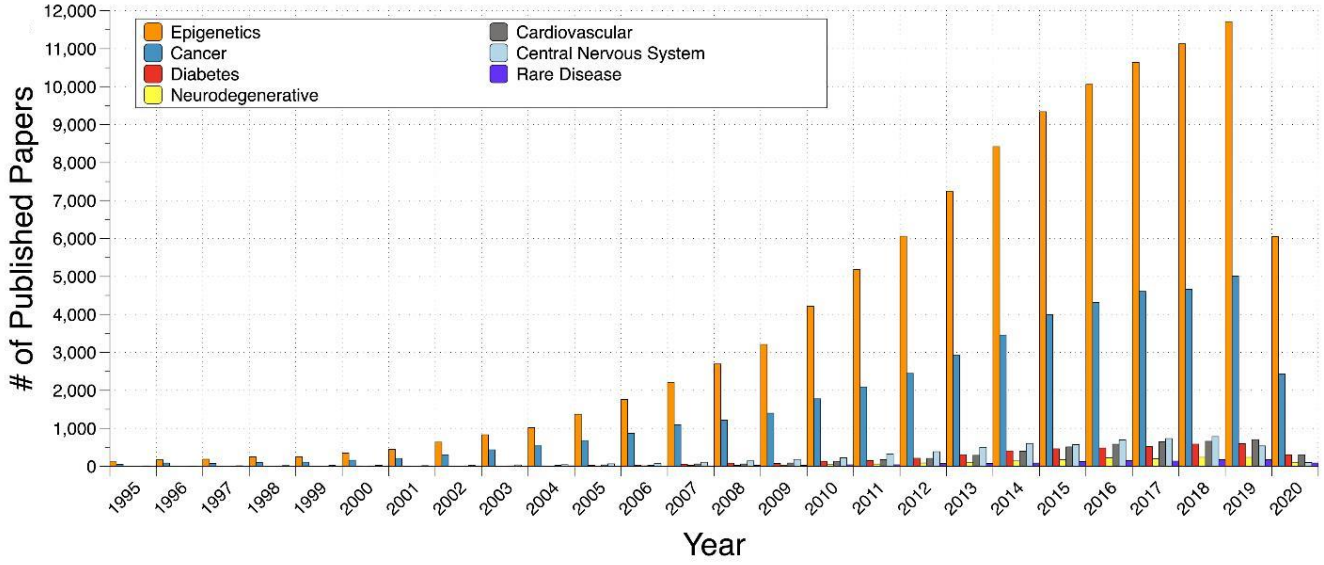
maintains existing methylation patterns

DNA methyltransferases (DNMTs)

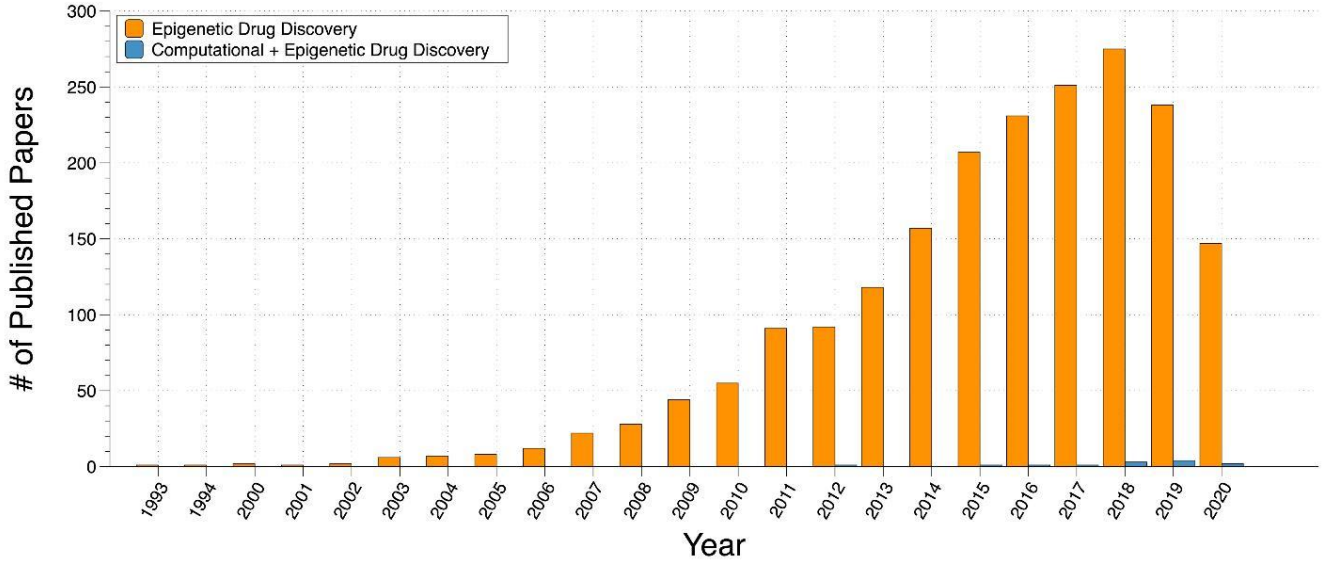
Methylation: first epigenetic modification linked to cancer

Increased interest of epigenetic drug discovery

Epigenetics and therapeutic potential



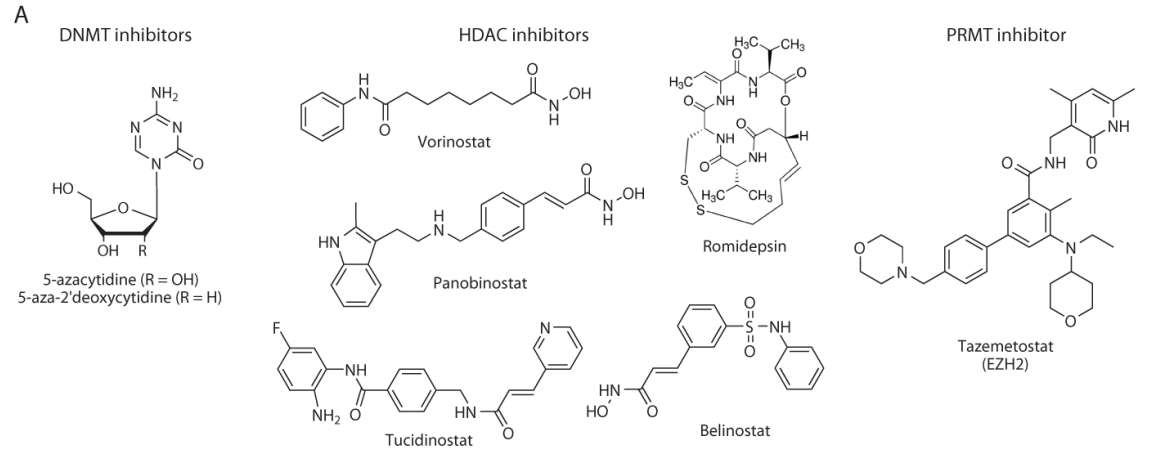
Epigenetic drug discovery

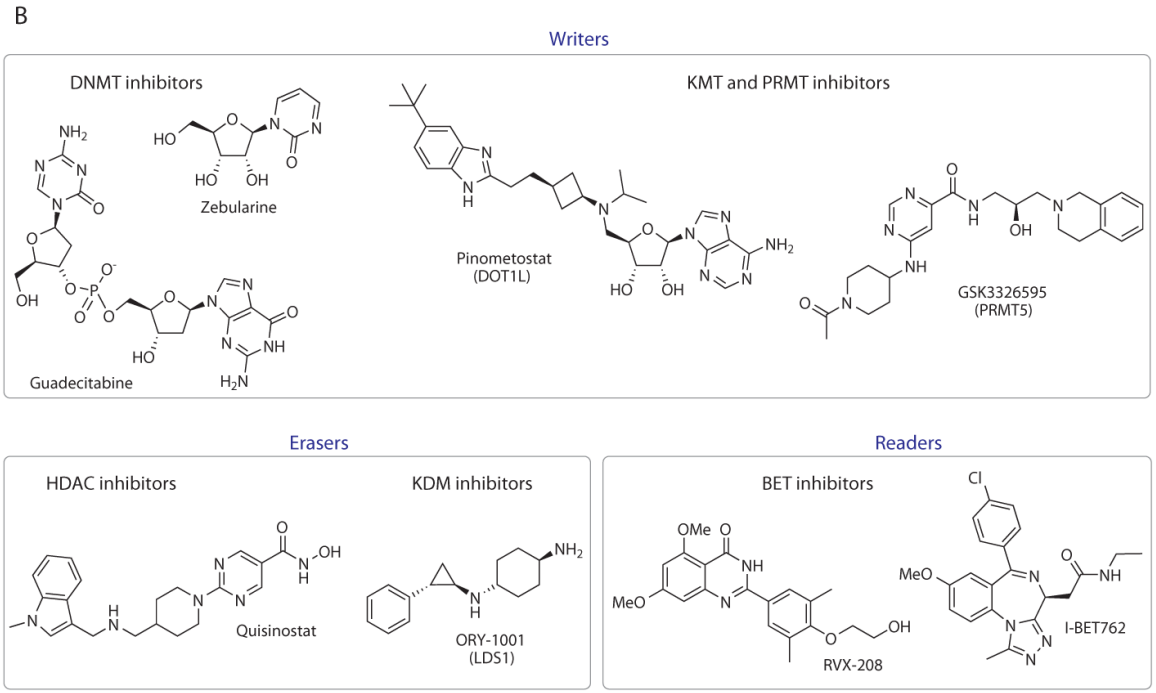


Epi-drugs in clinical use and clinical development

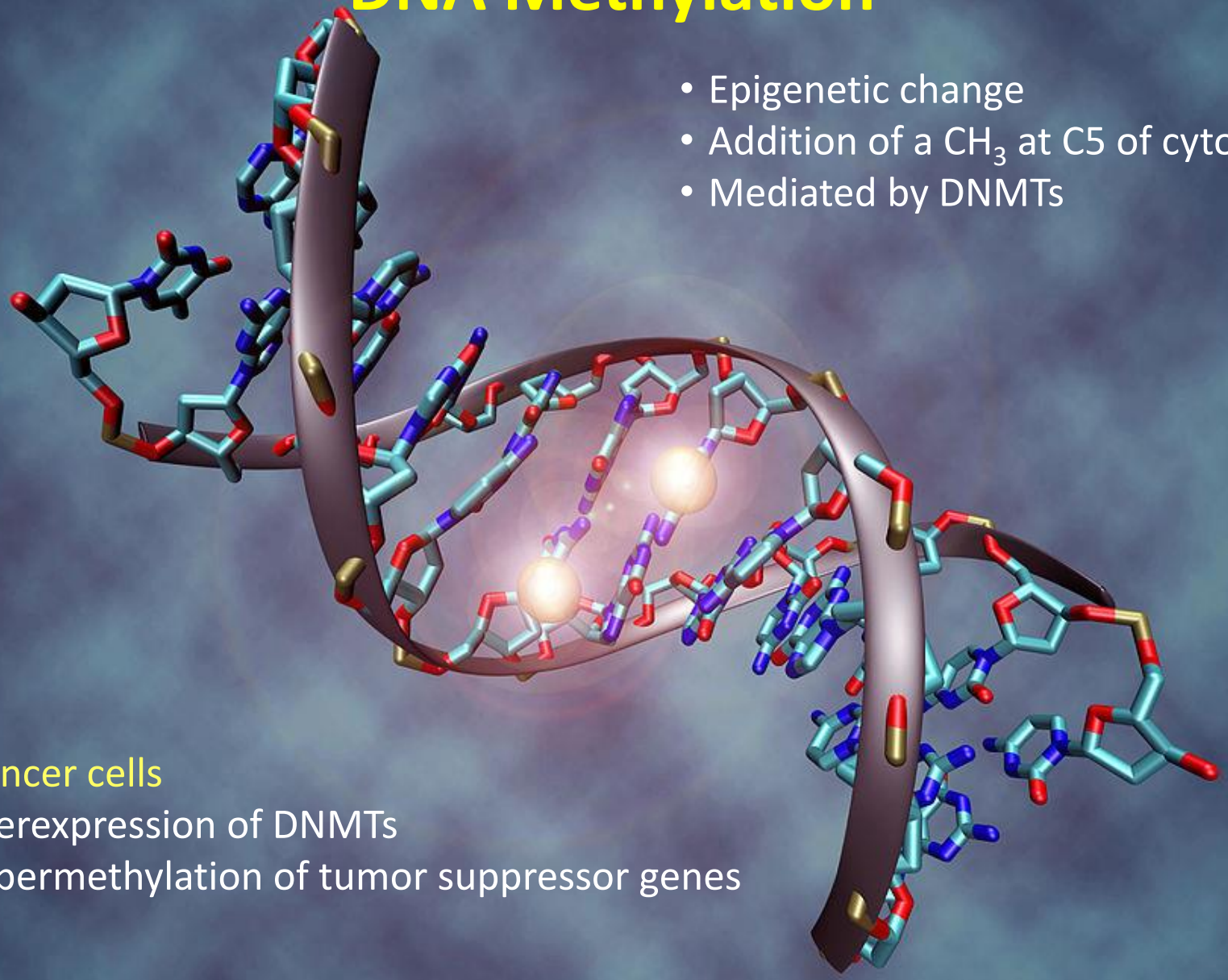
A. Drugs approved for clinical use.

B. Examples of small molecules in clinical development.





DNA Methylation



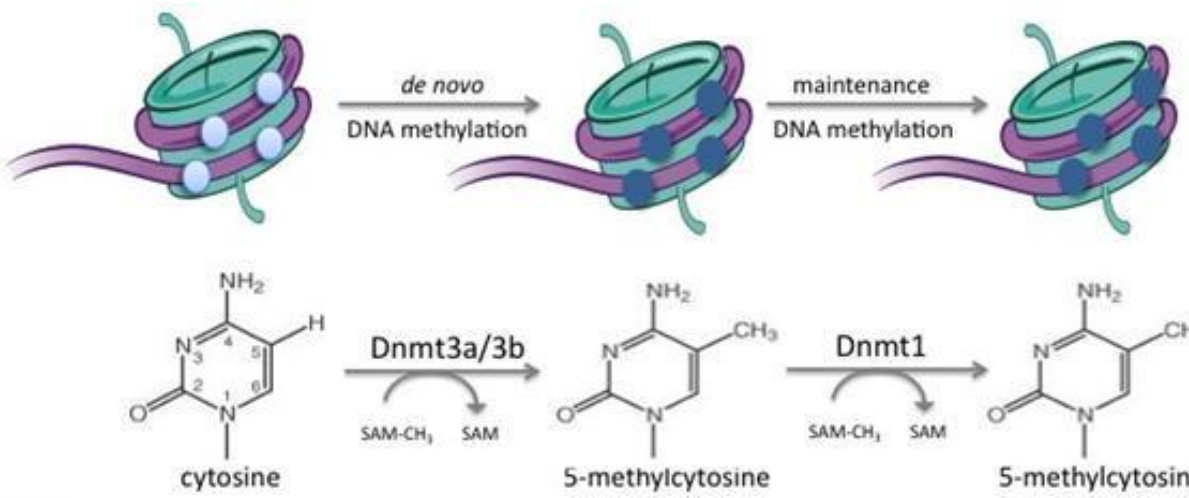
- Epigenetic change
- Addition of a CH_3 at C5 of cytosine
- Mediated by DNMTs

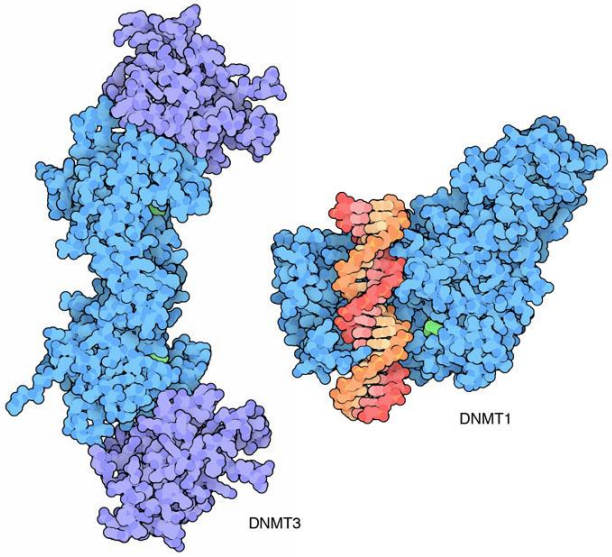
In cancer cells

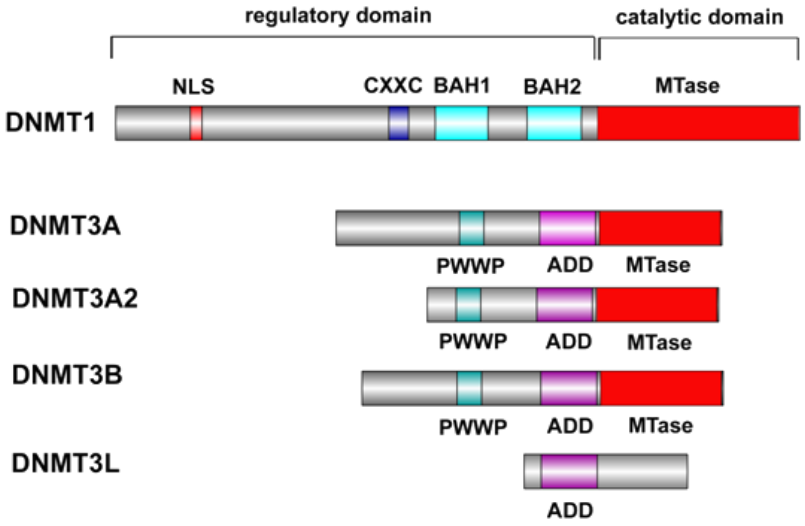
- Overexpression of DNMTs
- Hypermethylation of tumor suppressor genes

FEBS Letters 2011 585:2078

DNA metiltransferases (DNMTs)



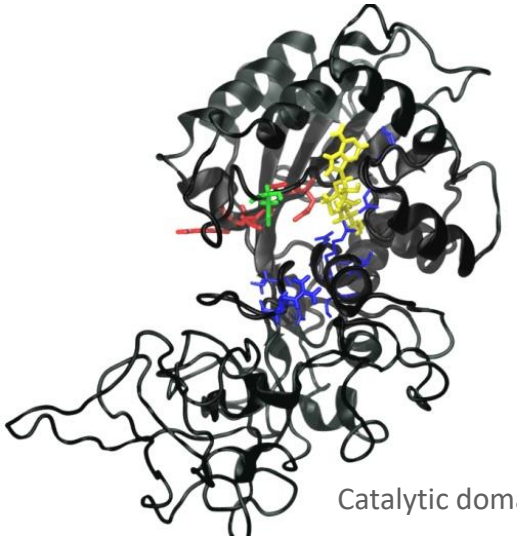




3D structures of DNMTs

Selected crystal structures of DNMTs

PDB code	Type	Region	Res (Å)	Cofactor	Ligand
3SWR	hDNMT1	Autoinhibitory linker, CXXC, BAH1/2, methyltransferase domain	2.49	SFG	
4DA4	mDNMT1	BAH1/2, methyltransferase domain	2.60	SAH	5-methyl-2'-deoxycytidine
3PTA	hDNMT1	Autoinhibitory linker, CXXC, BAH1/2, methyltransferase domain	3.60	SAH	
3PT6	mDNMT1	Autoinhibitory linker, CXXC, BAH1/2, methyltransferase domain	3.00	SAH	
3PT9	**mDNMT1**	**BAH1/2, methyltransferase domain**	**2.50**	**SAH**	



Catalytic domain

nature cancer

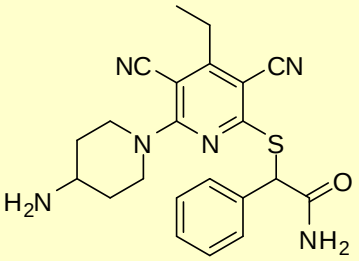
Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › nature cancer › articles › article

Article | Published: 27 September 2021

Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia

Melissa B. Pappalardi, Kathryn Keenan, ... Michael T. McCabe + Show authors

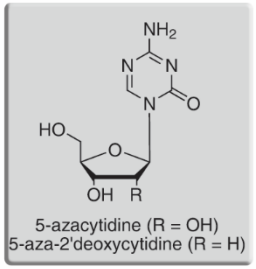


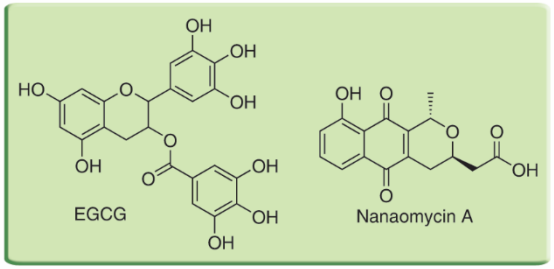
GSK3685032

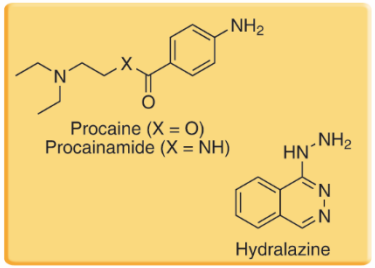
DNMT1 IC_{50} 36 = nM
DNMT3B/3L IC_{50} > 100000 nM
DNMT3A/3L IC_{50} > 100000 nM

Demethylating compounds

Cytosine analogues (citotoxic effects)

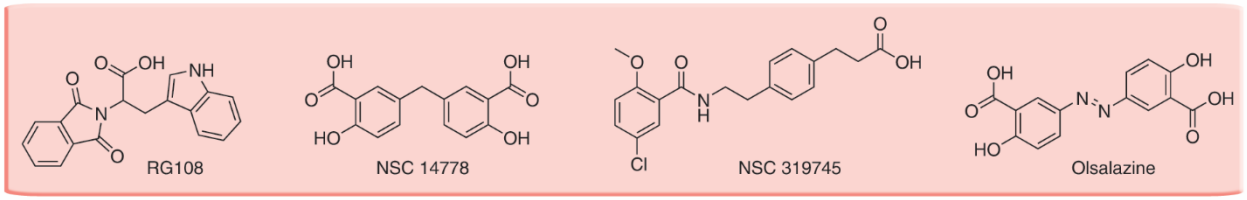




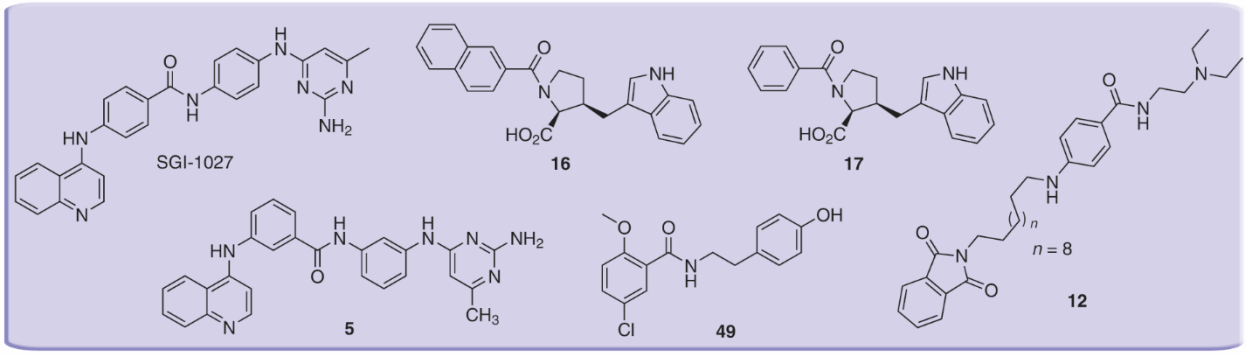


Approved drugs

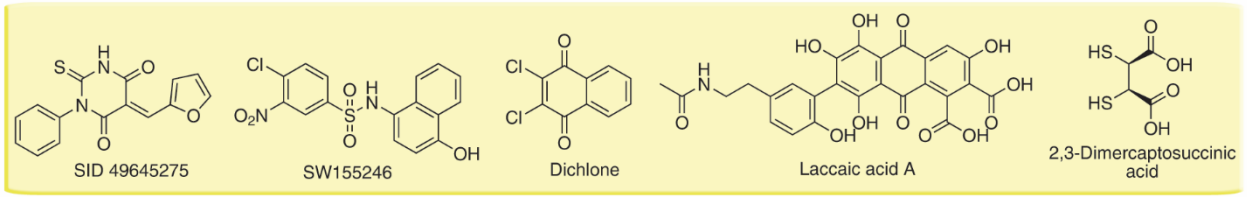
Virtual screening hits

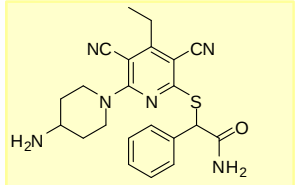


Synthesis



HTS





Goal of the research program

Discovery and development of DNMT inhibitors and other epigenetic targets

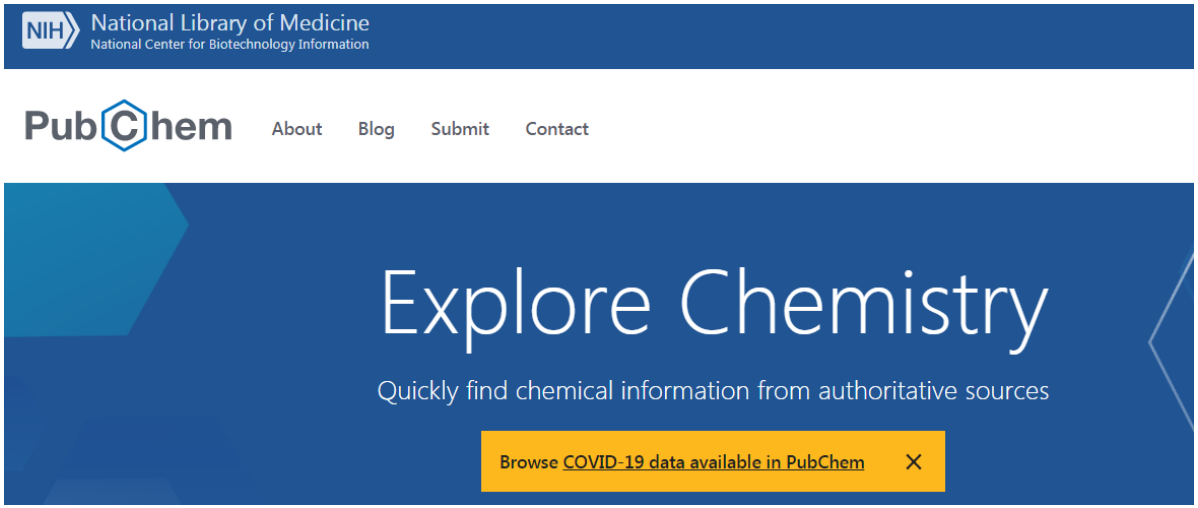
Approach

Computational methods integrated with experimental validation.

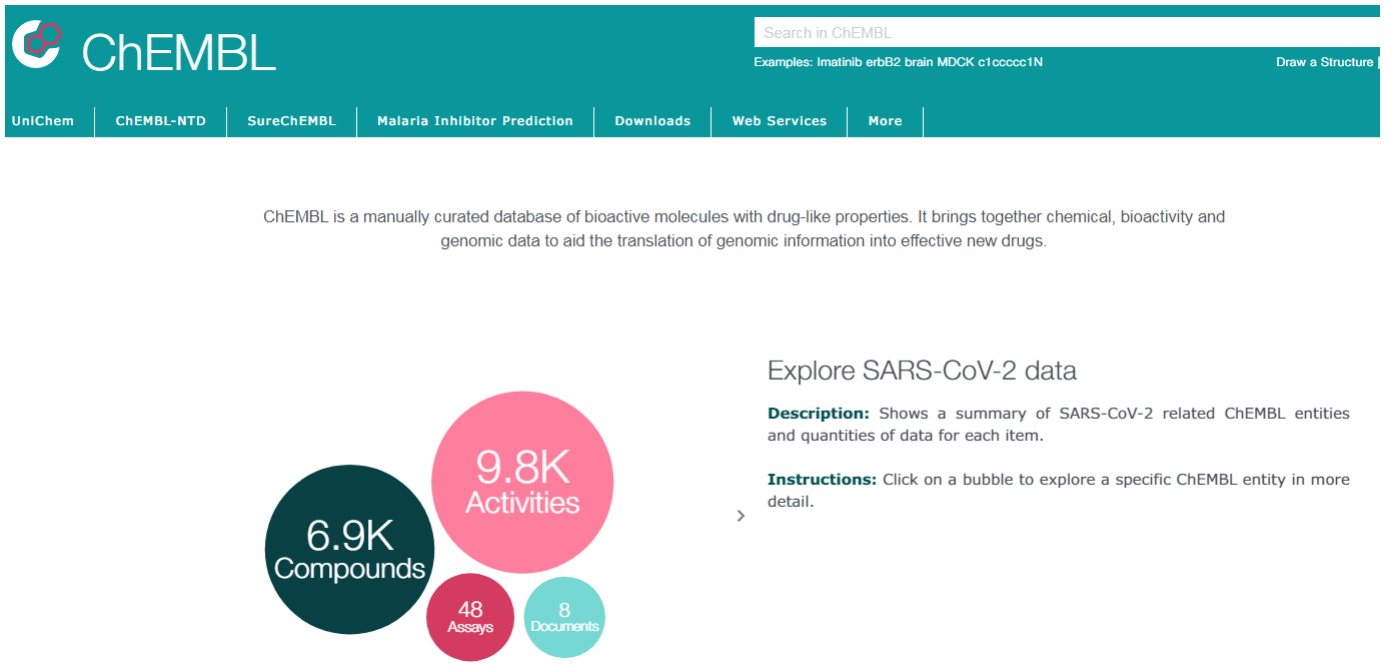
Specific aims

- Development of predictive models.
- Virtual screening of compound libraries.

Databases of compounds with experimental activity



- 103,284,373 compounds.
- 271,135,693 bioactivity data.
- 9,643,220 publications.
- 3,173,654 patents.



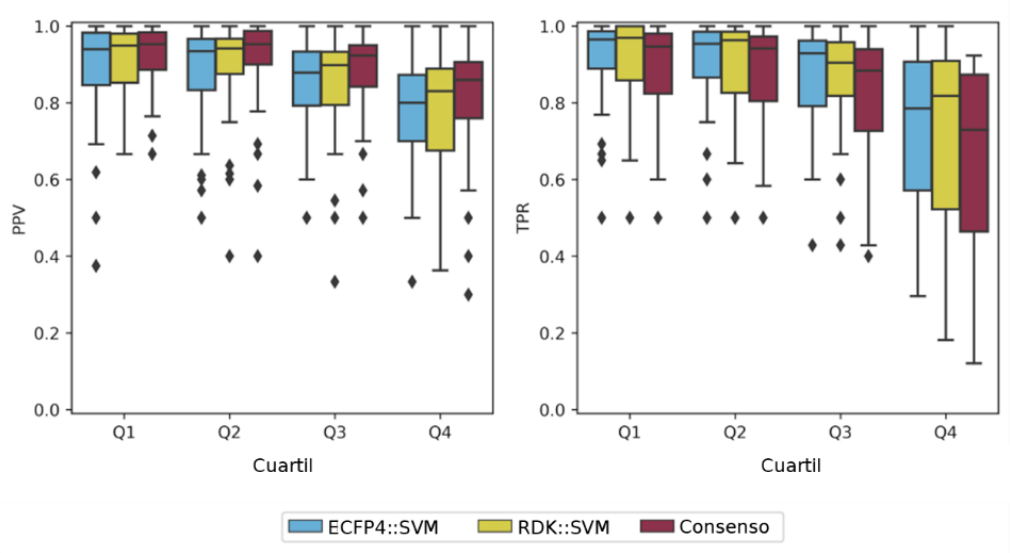
Epigenetic data in ChEMBL

Structure-activity data vs. 55 epigenetic targets

Target	Function	Families (HGNC)	Cluster (manually annotated)	Molecules	Scaffolds	% Active
BAZ2B	Acetylated histone reader	PHD finger proteins, methyl-CpG binding domain containing	BRD	53	27	25
BRD2	Histone PTM reader	NA	BRD	277	91	87
BRD3	Histone PTM reader	NA	BRD	263	89	95
BRD4	Histone PTM reader	NA	BRD	643	259	80
BRD9	Histone PTM reader	NA	BRD	13	9	77
BRPF1	Histone PTM reader	PHD finger proteins, PWWP domain containing	BRD	27	15	89
DNMT1	DNA methyltransferase	Zinc fingers CXXC-type, seven-beta-strand methyltransferase motif containing	DNMT	248	194	60
DNMT3A	DNA methyltransferase	PWWP domain containing	DNMT	47	30	55
DNMT3B	DNA methyltransferase	PWWP domain containing	DNMT	40	22	50
CREBBP	Histone acetyltransferase	Zinc fingers ZZ-type, lysine acetyltransferases	HAT	180	65	64
EP300	Histone acetyltransferase	Zinc fingers ZZ-type, lysine acetyltransferases	HAT	73	52	78
KAT2A	Histone acetyltransferase	Lysine acetyltransferases, ATAC complex, SAGA complex, GCN5 related N-acetyltransferases	HAT	27	20	41
KAT2B	Histone acetyltransferase	Lysine acetyltransferases, ATAC complex, SAGA complex, GCN5 related N-acetyltransferases	HAT	121	40	61
NCOA1	Histone acetyltransferase	Basic helix-loop-helix proteins, lysine acetyltransferases	HAT	634	568	22
NCOA3	Histone acetyltransferase	Basic helix-loop-helix proteins, lysine acetyltransferases, trinucleotide repeat containing	HAT	564	517	32
HDAC1	Histone deacetylase	Histone deacetylases class I, EMSY complex, NuRD complex, SIN3 histone deacetylase complex	HDAC	3304	1418	90
HDAC2	Histone deacetylase	Histone deacetylases class I, EMSY complex, NuRD complex, SIN3 histone deacetylase complex	HDAC	942	427	84
HDAC3	Histone deacetylase	Histone deacetylases class I	HDAC	854	395	80
HDAC4	Histone deacetylase	Histone deacetylases class IIA	HDAC	704	348	69
HDAC5	Histone deacetylase	Histone deacetylases class IIA	HDAC	235	150	58

Jesús Naveja

Development of predictive models



Binary classification

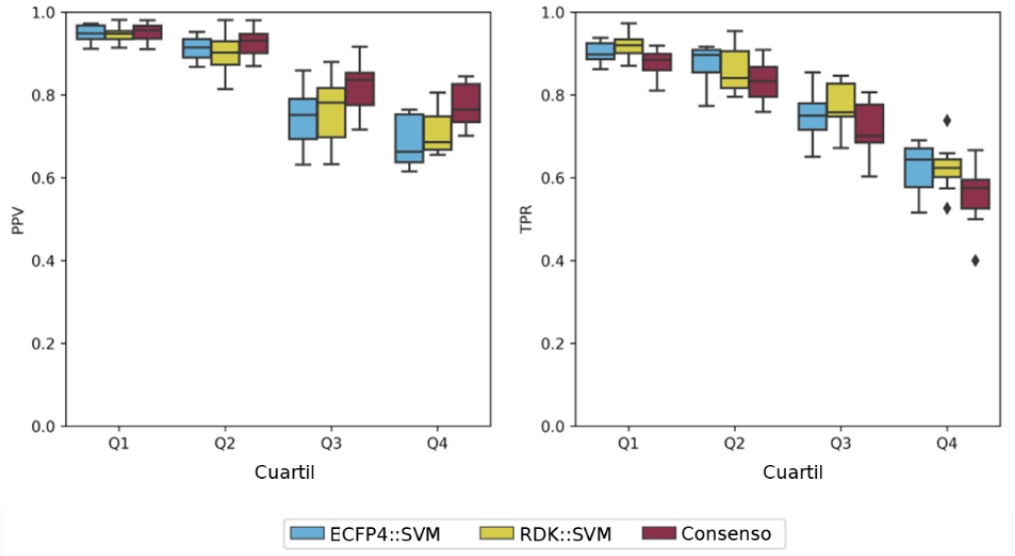
Precision: 0.92 - 0.81

Sensibility: 0.89 - 0.65

Target prediction

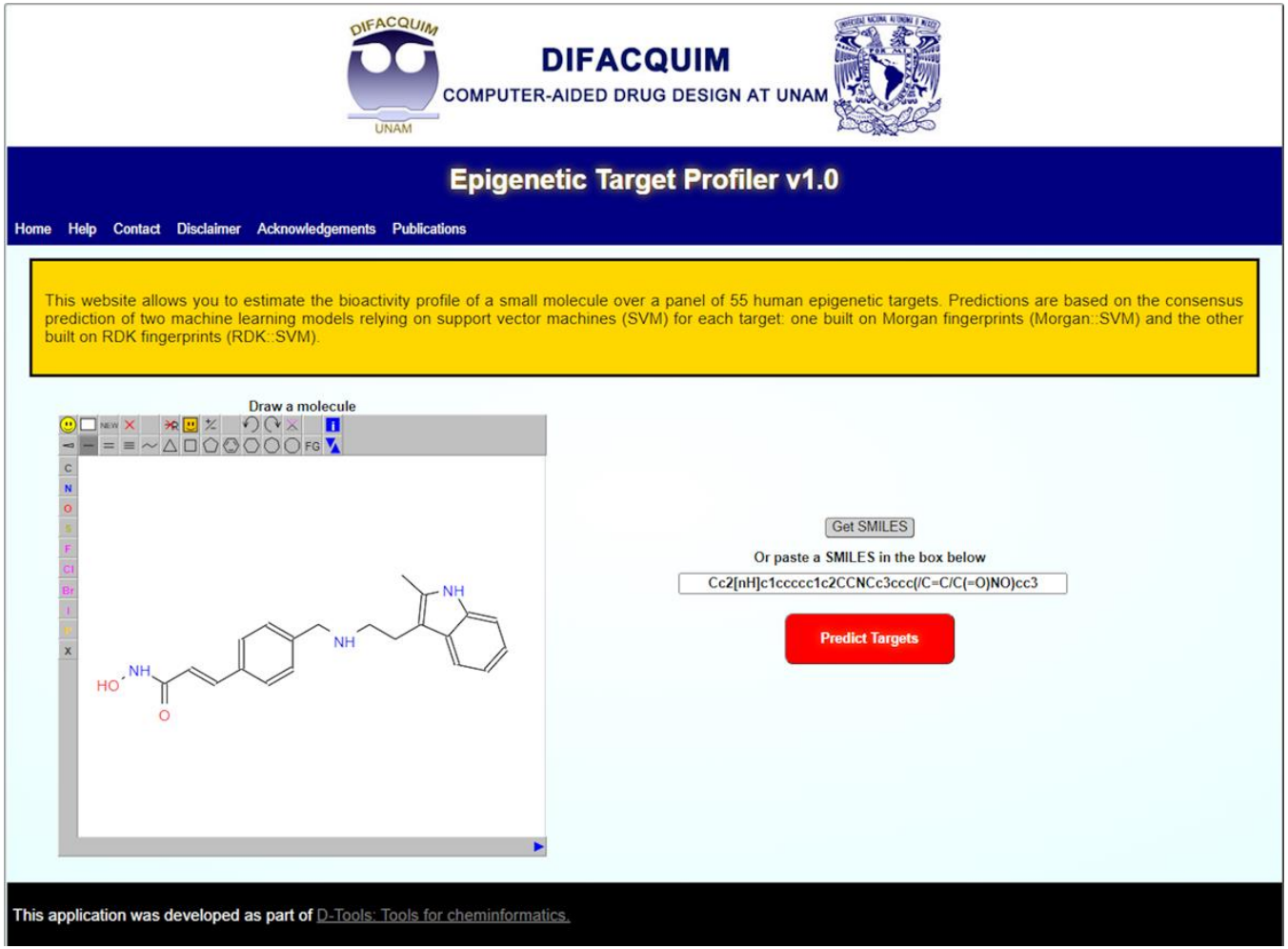
Precision: 0.95 - 0.77

Sensibility: 0.89 - 0.56



Norberto Sánchez

Epigenetic target fishing



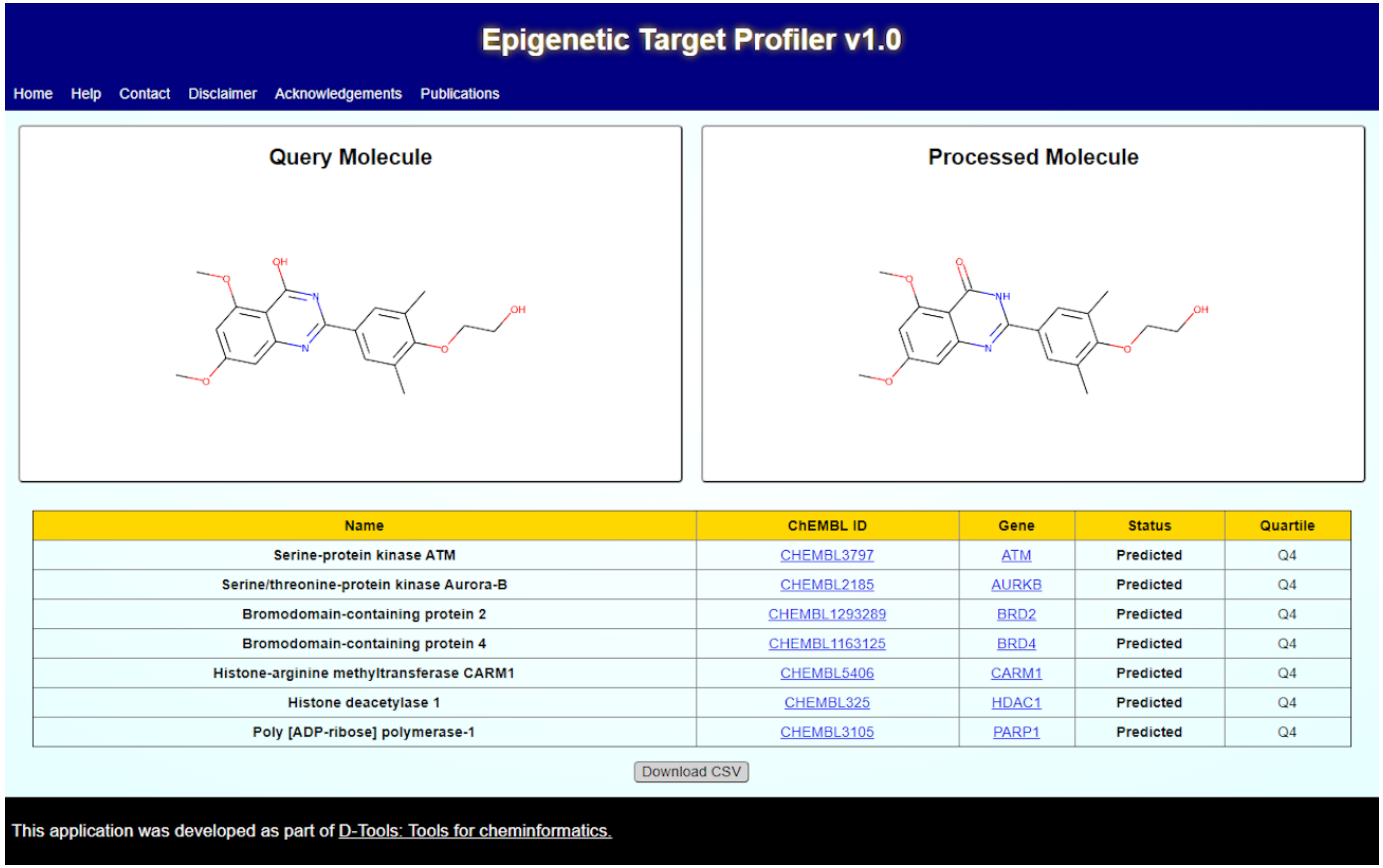
- **Implementation of predictive models.**
- Free webserver to predict the activity of small organic molecules with 55 epigenetic targets.

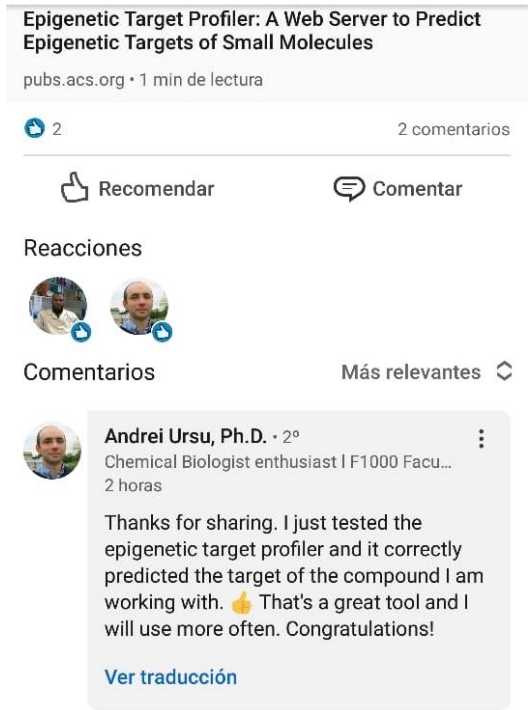
Norberto Sánchez

www.epigenetictargetprofiler.com

Epigenetic target fishing

Inverse virtual screening with epigenetic targets

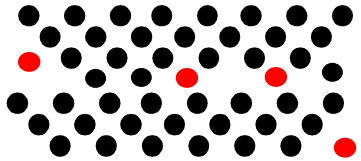




Norberto Sánchez

Screening of compound databases

1. *In silico* characterization and "purification" (data curation) of compound libraries



Compound libraries
- Natural products
- Food chemicals
- In-house from collaborators
- Commercial synthetic
- Focused libraries

2. Computational screening

Strategies
- Similarity searching
- Molecular docking & dynamics
- Machine learning: Epigenetic Target Profiler

Selected molecules for testing (virtual screening hits)

3. Experimental screening of selected compounds

Approaches
- Enzymatic inhibition assays: collaboration and contract services.
- Additional experiments: Open for collaborations!

Active compounds

4. Optimization

Strategy

Chemical synthesis and testing of analogs.

Compound libraries

Natural products & food chemicals

~0.5 million compounds

~530 natural products

~22K food chemicals

Data sets from collaborators

- Synthetic analogs of caffeic acid — Dra. Laura Alvarez (UAEM), Dr. Mayra Antúnez (UAEM)
- Small molecules — Dr. Alexander Gagnon (UQAM, Canada)

Commercial general synthetic libraries

ZINC15

Welcome to ZINC, a free database of commercially-available compounds for virtual screening. ZINC contains over 230 million purchasable compounds in ready-to-dock, 3D formats. ZINC also contains over 750 million purchasable compounds you can search for analogs in under a minute.

Focused libraries

Name	Number of compounds
Bromodomain-containing protein 4 (BRD4) inhibitor library	401
CREB binding protein (CREBBP) targeted library	576
DNA (cytosine-5)-methyltransferase 1 (DNMT1) inhibitor library	466
DNA (cytosine-5)-methyltransferase 3 beta (DNMT3b) inhibitor library	1261
DOT1-like histone H3 methyltransferase (DOT1L) inhibitor library	622
Enhancer of Zeste Homolog 2 (EZH2) Targeted Library	979
Histone acetyltransferase (HAT) inhibitor library	814
Histone deacetylase (HDAC) inhibitor library	803

Epigenetic focused libraries: characterization

11 focused libraries.

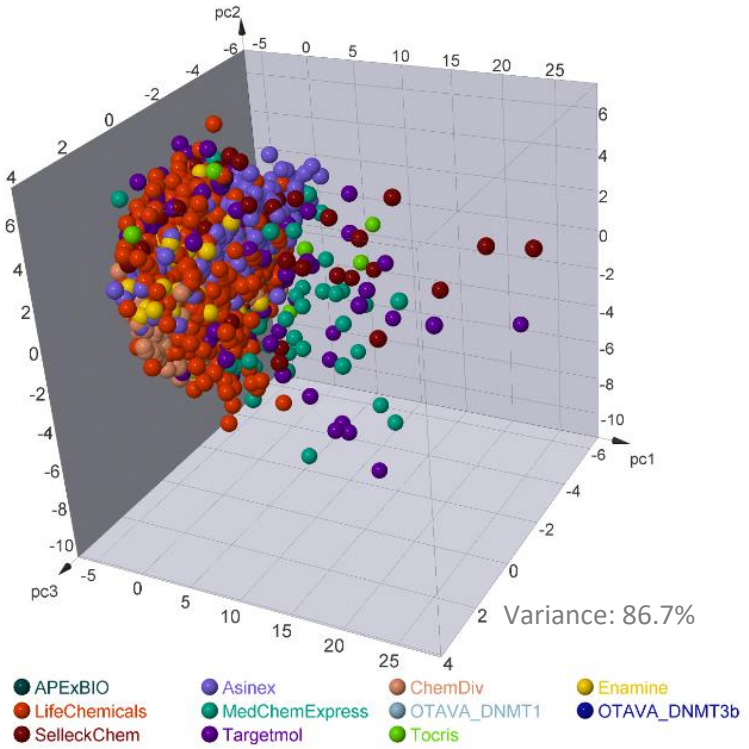
- 56,680 initial compounds.
- **53,443** compounds after data curation.

Company	Library size	
	Initial	After curation
ApeXBio	328	310
Asinex	5 391	5 313
ChemDiv	30 431	27 543
Enamine	9 352	9 352
Life Chemicals	7 019	7 011
MedChemExpress	700	650
OTAVA DNMT1	466	399
OTAVA DNMT3B	1 261	1 230
Targetmol	932	859
Tocris	101	99
SelleckChem	699	677

Alexis Padilla

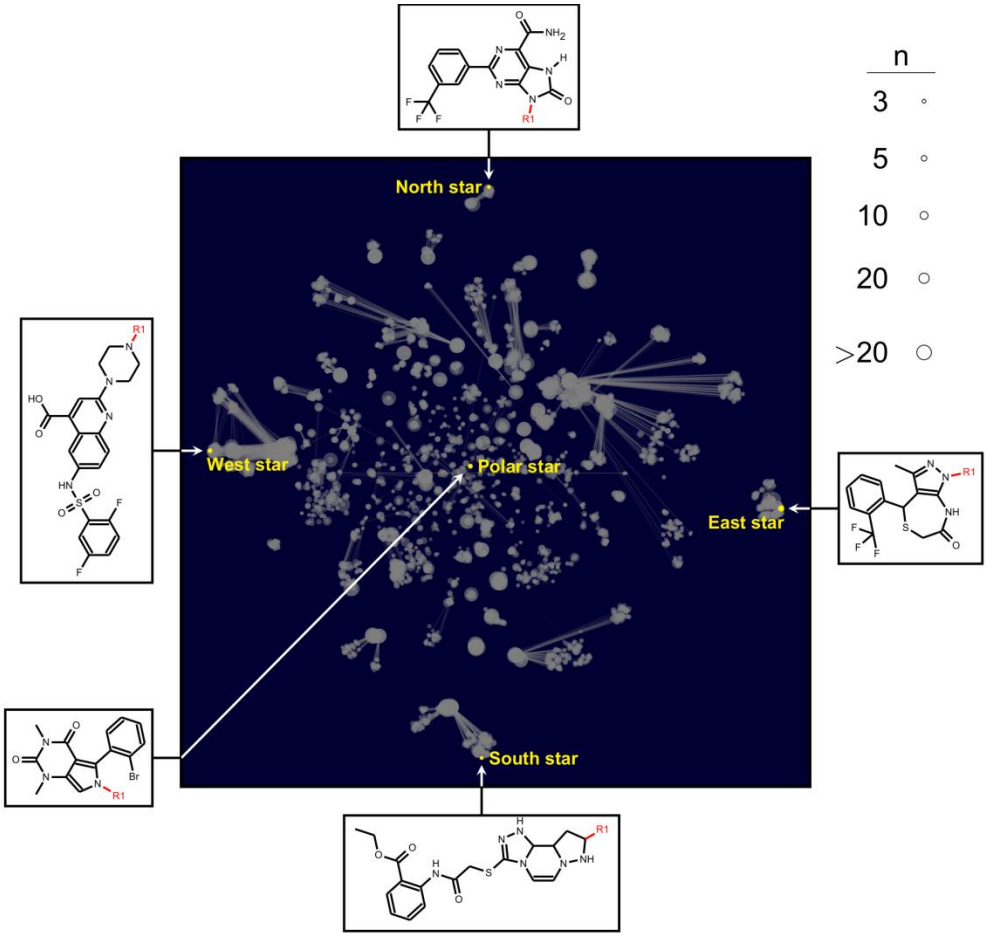
Visual representation of the **chemical space**

Principal component analysis of 6 properties of pharmaceutical interest.



- Compound libraries have drug-like properties.
- Have different diversity (MedChemExpress, Targetmol, SelleckChem).

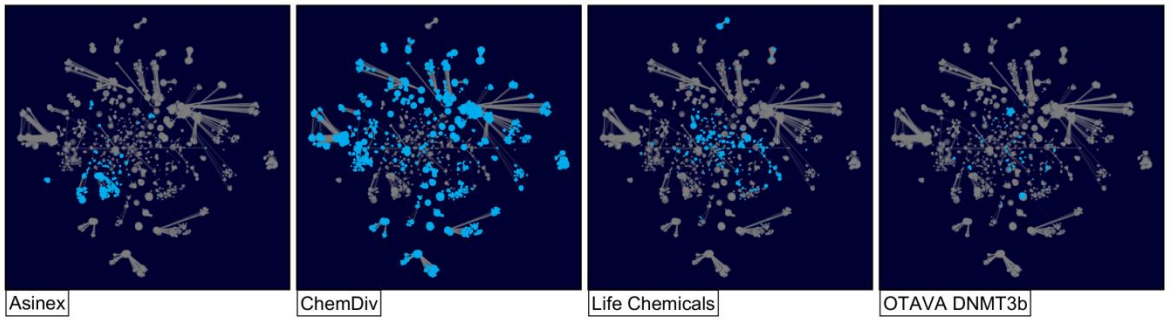
Constellation maps **Epigenetic focused libraries**



Representative molecular scaffolds of each "constellation"

Molecular libraries have:

- Different chemical structures.
- Cover different regions of the chemical space.
- Different diversity.



Computational screening of focused libraries

Approaches

1. **Epigenetic Target Profiler (ETP)**
2. **Docking** with MOE
3. Docking AutoDock VINA
4. **Re-scoring**: Extended connectivity interaction features (ECIF).*

53,443 compounds (initial library size)

ETP

119 compounds with best ETP predictions.

Consensus docking and re-scoring

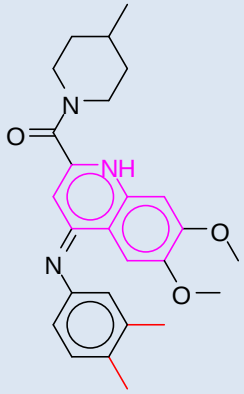
20 compounds selected for enzymatic inhibition assays.

Proveedor	ID	AutodockVINA (kcal/mol)	ECIF	MOE (kcal/mol)
Enamine	Z991906684	-8.5	5.36	-8.02
Tocris	TC25	-9.8	5.3	-9.95
Targetmol	T2354	-9.4	5.36	-9.74
ChemDiv	C191-0147	-9	5.19	-8.57
ChemDiv	C769-0077	-8.8	5.28	-8.19
ChemDiv	F477-3331	-9.3	5.26	-8.33
ChemDiv	G119-0160	-8.7	5.37	-7.29
ChemDiv	L485-2681	-8.9	5.4	-8.64
ChemDiv	L485-2718	-9.4	5.48	-8.26
ChemDiv	L485-2735	-8.9	5.45	-8.12
ChemDiv	L485-2754	-10	5.56	-7.95
ChemDiv	L485-2759	-10.1	5.42	-8.16
MedChemEx	HY-10128	-7.7	5.52	-9.83
ChemDiv	L485-2761	-9.6	5.43	-8.48
ChemDiv	L485-2767	-9.8	5.23	-8.84

Alexis Padilla

Focused libraries: experimental screening

% of activity of DNMT1 of virtual screening hits at 100 μM*

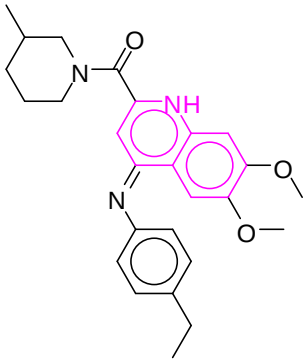


MolPort-009-710-005
ChemDiv
58.7 (5.35) %

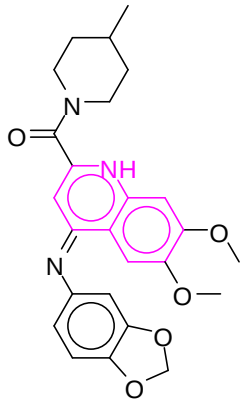
MolPort-010-672-527
ChemDiv
1078.9 (65) %

Quinolines

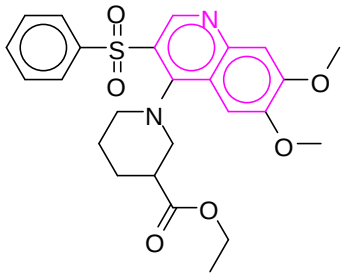
- **Inhibition 10-40 %**
- Cases of activation and *activity cliffs*



MolPort-009-700-568
ChemDiv
86.9 (1.11) %



MolPort-010-672-537
ChemDiv
88.7 (2.26) %

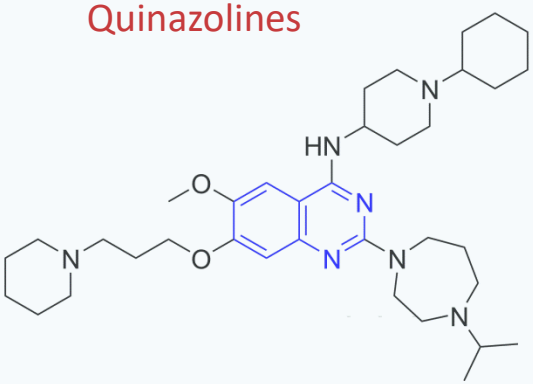


MolPort-003-071-343
ChemDiv
90.5 (3.25) %

REACTION BIOLOGY CORP.

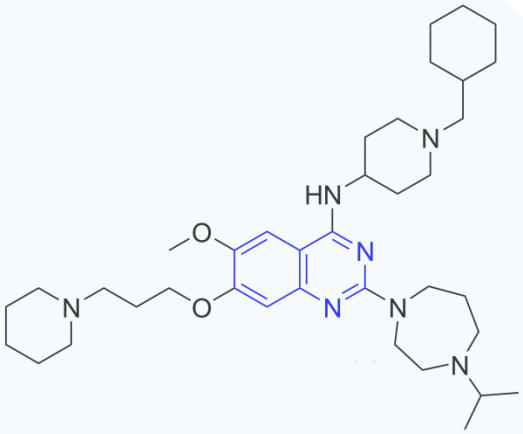
Focused libraries: experimental screening

Quinazolines



MolPort-023-277-153

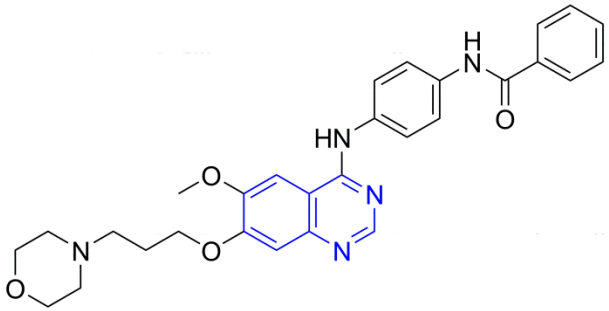
DNMT1: IC_{50} = 30 nM
DNMT3A: IC_{50} = 4870 nM
DNMT3B: > 100000 nM



MolPort-035-789-726

DNMT1: IC_{50} = 81 nM
DNMT3A: IC_{50} = 14690 nM
DNMT3B: > 100000 nM

- Nanomolar inhibition of DNMT1.
- ~ 10 times more potent than the control SAH.
- Selective towards DNMT1.
  - Low or no activity with DNMT3A and DNMT3B



MolPort-006-396-396

DNMT1: % inhibition = 7.4 +/- 2.4
DNMT1: IC_{50} NA
DNMT3A, DNMT3B: NA

Reference:
SAH: IC_{50} = 0.34 μM

REACTION BIOLOGY CORP.

Survey in the literature…

Quinazolines are also inhibitors of the epigenetic reader G9a

Journal of Medicinal Chemistry

ARTICLE

pubs.acs.org/jmc

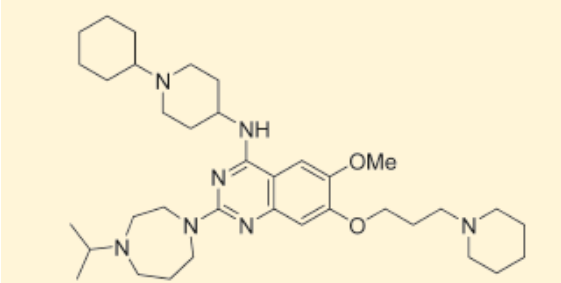
Optimization of Cellular Activity of G9a Inhibitors 7-Aminoalkoxy-quinazolines

Feng Liu,[†,∥] Dalia Barsyte-Lovejoy,[‡,∥] Abdellah Allali-Hassani,[‡] Yunlong He,[§] J. Martin Herold,[†] Xin Chen,[†] Christopher M. Yates,[⊥] Stephen V. Frye,[†] Peter J. Brown,[‡] Jing Huang,[§] Masoud Vedadi,[‡] Cheryl H. Arrowsmith,[‡] and Jian Jin[*,†]

[†]Center for Integrative Chemical Biology and Drug Discovery, Division of Medicinal Chemistry and Natural Products, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States

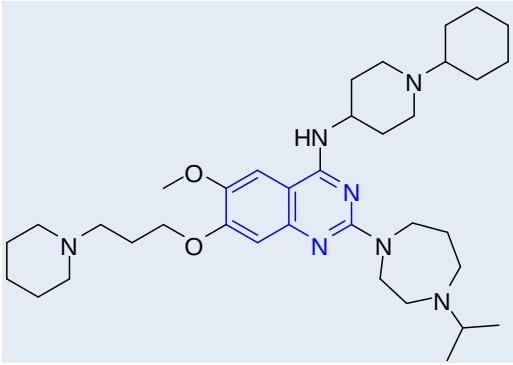
[‡]Structural Genomics Consortium, University of Toronto, Toronto, Ontario, M5G 1L7, Ontario, Canada

[§]Laboratory of Cancer Biology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States



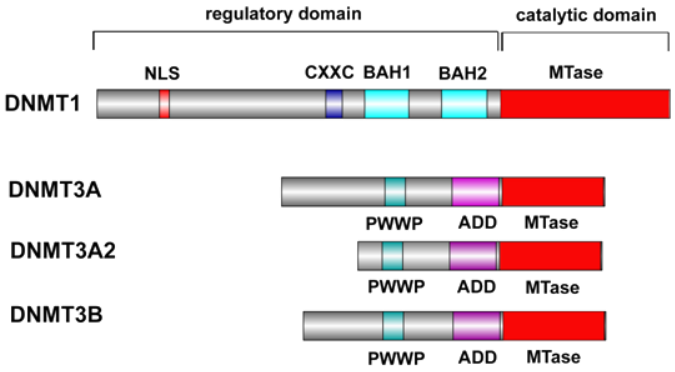
G9a IC_{50} = 6 nM
Reduction of H3K9me2 in MCF7 cells: IC_{50} = 10 nM
Cell toxicity (MCF7 cells): EC_{50} = 4,700 nM

Progress to elucidate the mechanism of inhibition and selectivity

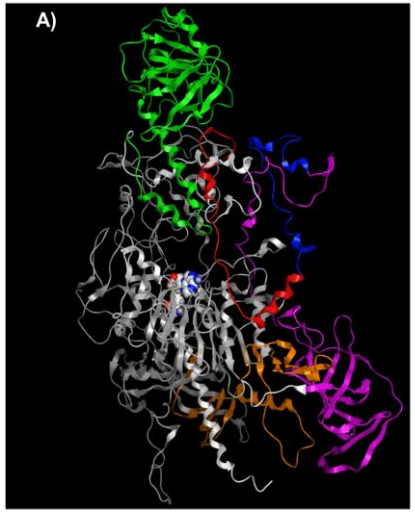


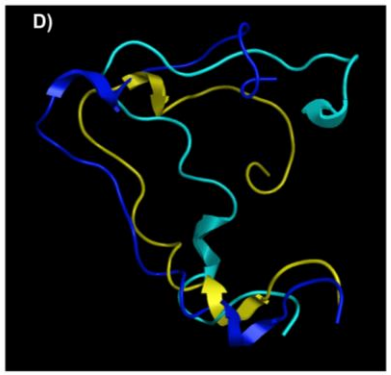
MolPort-023-277-153

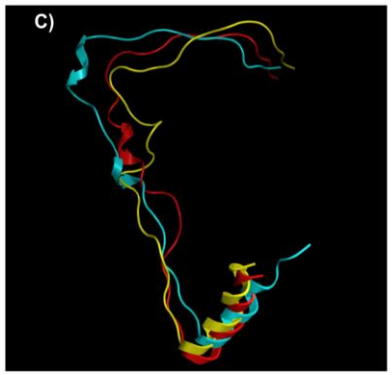
DNMT1: IC_{50} = 30 nM
DNMT3A: IC_{50} = 4870 nM
DNMT3B: > 100000 nM



Computational
(docking + dynamics)
Interaction with the CXXX domain.

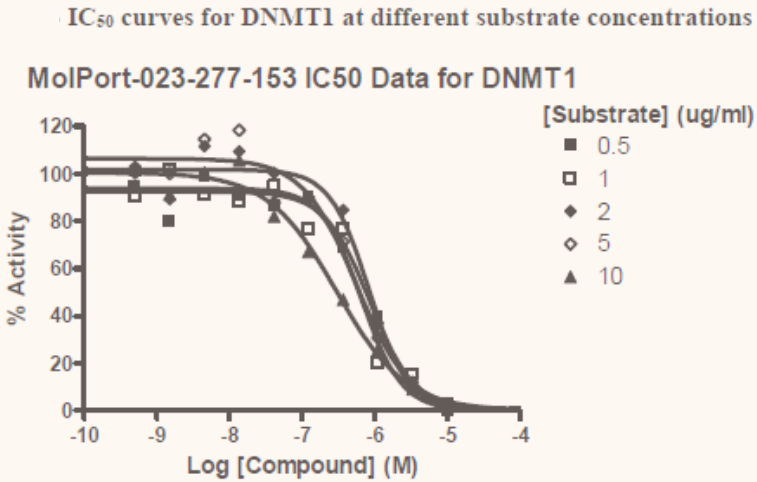






M. en C. Edgar López

Experimental
It does not compete with DNA



Towards multitarget epigenetic drug discovery

Artificial Intelligence in the Life Sciences 1 (2021) 100008

Contents lists available at ScienceDirect

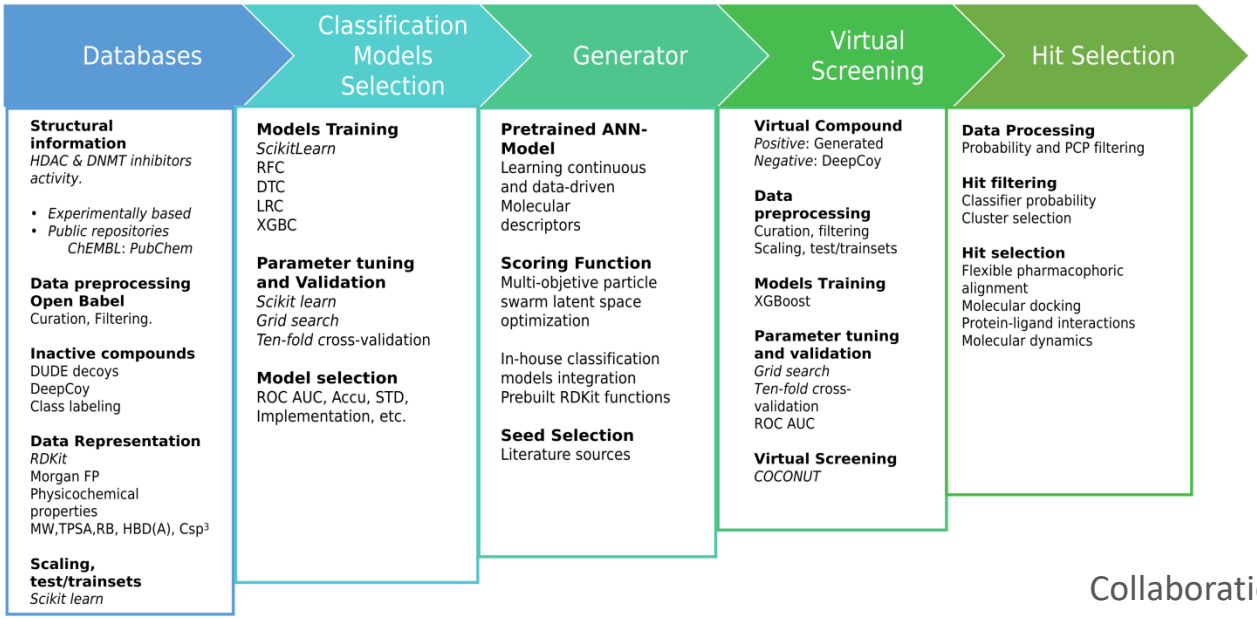
Artificial Intelligence in the Life Sciences

journal homepage: www.elsevier.com/locate/ailsci

Methods & Protocols

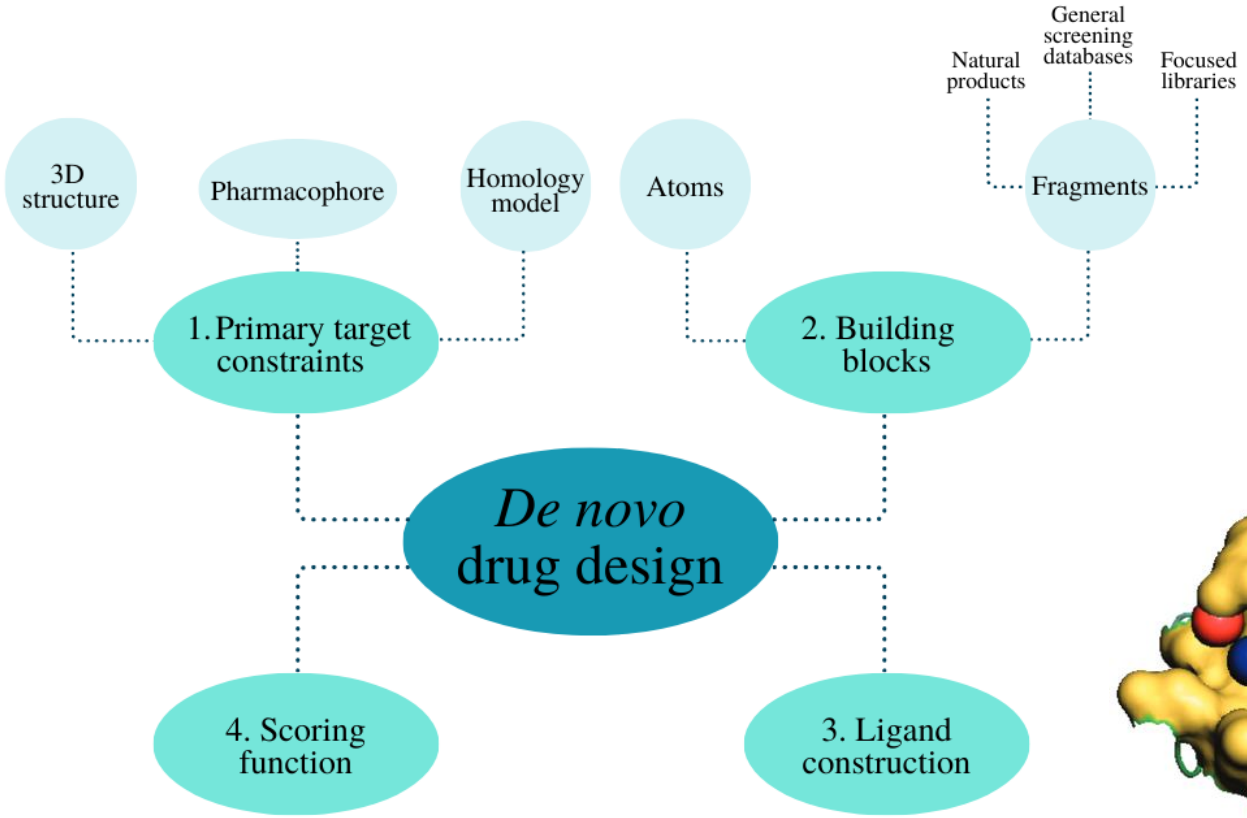
An *in silico* pipeline for the discovery of multitarget ligands: A case study for epi-polypharmacology based on DNMT1/HDAC2 inhibition

Fernando D. Prieto-Martínez[a], Eli Fernández-de Gortari[b,*], José L. Medina-Franco[c], L. Michel Espinoza-Fonseca[d,*]

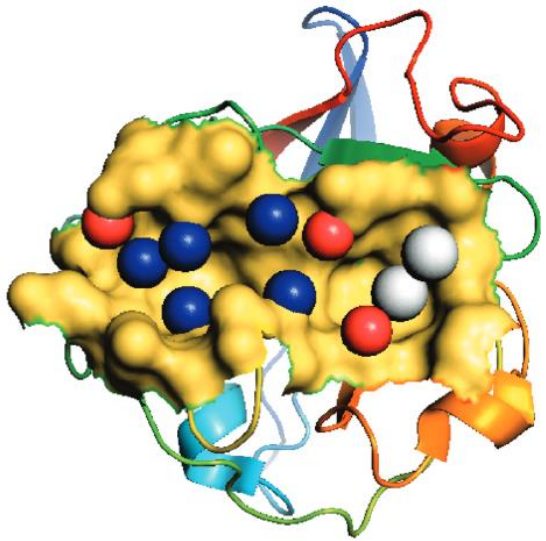


Collaboration with Dr. Eli Fernandez

Perspective: *De novo* design



MSc. Diana Prado



Design of small molecules from scratch based on anchor points in the binding site

Chemical space

Biological space

Methods

Human and others

frontiers in Drug Discovery

SPECIALTY GRAND CHALLENGE
published: 28 July 2021
doi: 10.3389/fddsv.2021.728551

Grand Challenges of Computer-Aided Drug Design: The Road Ahead

José L. Medina-Franco *

DIFACQUIM Research Group, Department of Pharmacy, School of Chemistry, National Autonomous University of Mexico, Mexico City, Mexico

Chemical space

- Expand the medicinally relevant chemical space.
- Rational design and screen of ultra-large chemical libraries.
- Repurpose existing libraries (drugs and *in-house* collections).
- Rescue missing hits and lead compounds from screening libraries.
- Explore neglected regions of chemical space.

Biological space

- Improve multi-target drug design and polypharmacology.
- Explore “dark” targets and identify novel promising regions in the genome.
- Improve targeting protein-protein interactions.
- Continue investigating targets associated with rare and neglected diseases.

Methodological challenges

How to conduct the search for new and better drugs at the intersection of the chemical and biological spaces?

- Computational chemogenomics.
- Automated *de novo* design and computational fragment screening.
- Improve property prediction, including ADME and toxicity.
- Modeling large and complex systems.
- Continue to improve molecular docking and scoring.
- Improve the hit rate of virtual screening and strategies to automatically propose high quality hits.
- Synergize with other methods: consensus approaches.
- Ensure **data curation** and **quality**.

Human factor and other challenges

- Communication and human interaction.
  - Improve multidisciplinary research: reach common objectives from different perspectives.
  - Enhance communication across research teams; avoid duplicating efforts.
- Dissemination and data sharing.
  - Rigorous dissemination of information and high-quality data.
  - Transparency and reproducibility.
  - Open science vs. securing intellectual property.
- Education and training.
  - Individuals and teams.
  - Set up realistic expectations of computational methods.

Grand challenge

Use cheminformatics rationally beyond the hype

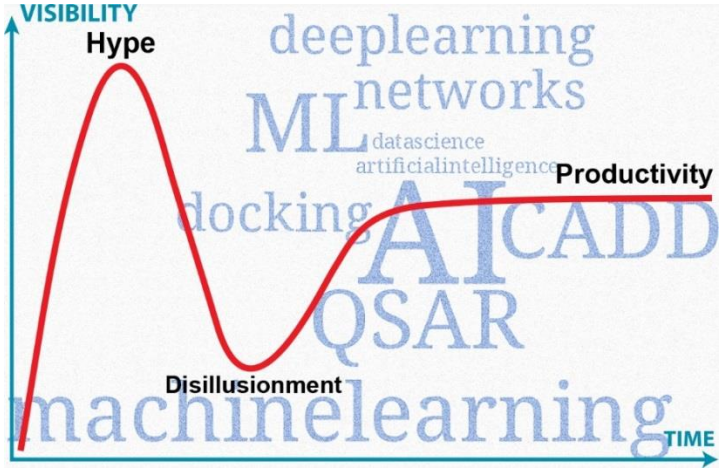
F1000Research F1000Research 2021, 10(Chem Inf Sci):397 Last updated: 21 SEP 2021

Check for updates

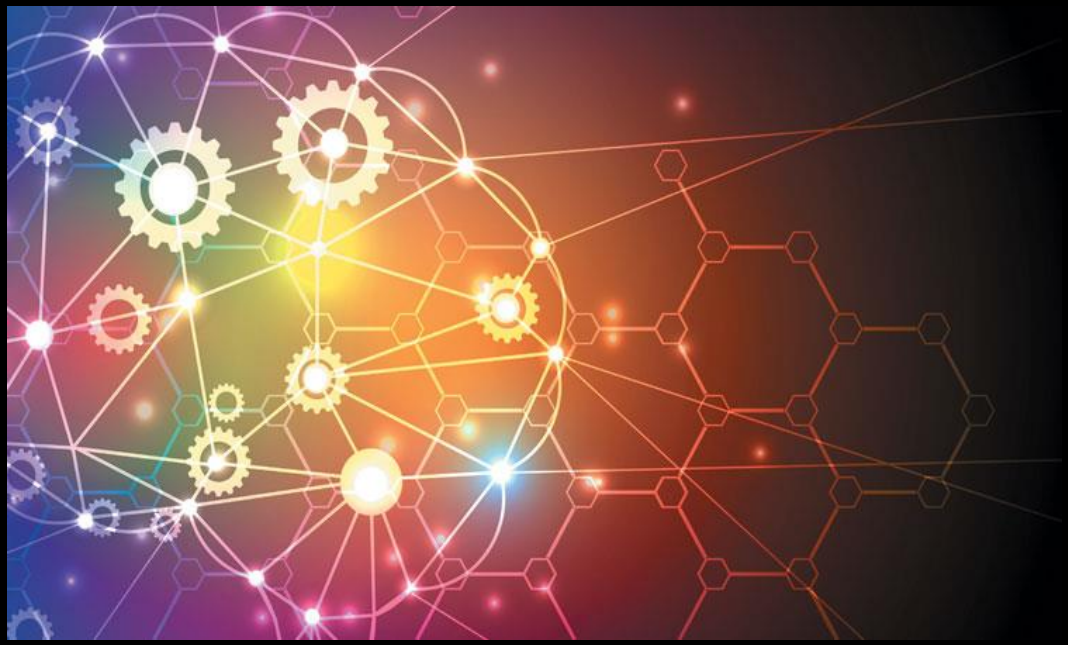
OPINION ARTICLE

Rationality over fashion and hype in drug design [version 1; peer review: 2 approved]

José L. Medina-Franco[1], Karina Martinez-Mayorga[2], Eli Fernández-de Gortari[3], Johannes Kirchmair[4], Jürgen Bajorath[5]



"Unrealistic expectations are a quick road to disappointments"



Summary

Take home messages

- Chemoinformatics: an independent discipline that impacts many areas of chemistry.
- Discovery of epi-drug candidates.
  - Development of *Epigenetic Target Profiler*.
  - Virtual screening identifies low micromolar and selective DNMT1 inhibitors.
- Challenges of chemoinformatics and CADD.
  - Revisit and expand chemical and biological spaces.
  - Several methodological challenges: data quality is a must.
  - Effective communication and education/training: beware of the "artificial intelligence extasy".

Acknowledgments

DIFACQUIM´s students & alumni

- Norberto Sánchez
- Jesús Naveja
- Fernando Prieto
- Edgar López
- Euridice Juárez
- Marisa Santibáñez
- Claudia Oviedo
- Alexis Flores
- Jocelyn Salazar
- Diana Prado

Collaborators

Eli Fernádez
INL, Portugal

Paola Arimondo
CNRS, France

Hyang-Min Byun
New Castle, UK

Massimo Bertinaria
Università degli Studi di Torino

Laura Alvarez
UAEM, Mexico

Mark Muller
Topogen, Colorado, USA

Alexandre Gagnon
Université du Québec à Montréal

Mayra Antúnez
UAEM, Mexico

Alfonso Dueñas
INCAN, Mexico

Carlos Velazquez
University of Alberta, Canada

Keith Robertson
Mayo Clinic, USA

Federica Catti
Arkansas State University

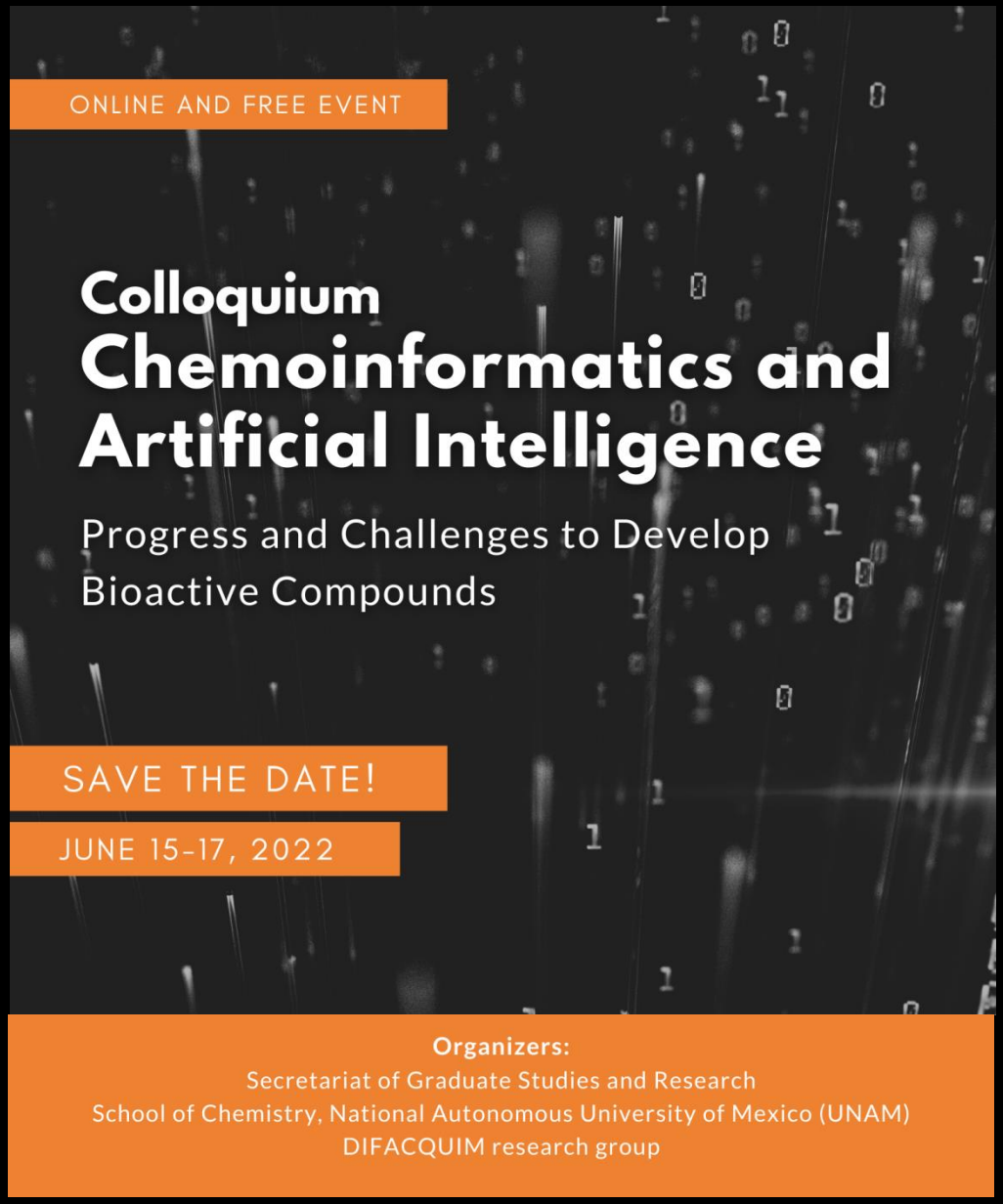
Funding

CONACyT 282785

PAPIIT IA203718, IN201321

PAIP 5000-9163

Miztli Supercomputer (UNAM)



16 speakers.

Progress and challenges to develop **peptides, natural products, drug candidates**.

Speakers from:
Academia, industry and other research institutions.

Registration

QR registration code

www.difacquim.com
@difacquim