

Aprendizado de Máquina para o Estudo de Interações Intermoleculares



Dr. Walter F. de Azevedo, Jr.
azevedolab.net

walter@azevedolab.net

Departamento de Física
Instituto de Ciências Exatas
Universidade Federal de Alfenas-MG
Brasil



AULA MAGNA

Programa de Pós-Graduação em Física
UNIFAL-MG / UFSJ

Aprendizado de Máquina para o Estudo de Interações Intermoleculares



Dr. Walter Filgueira de Azevedo Jr.
Departamento de Física
Universidade Federal de Alfenas

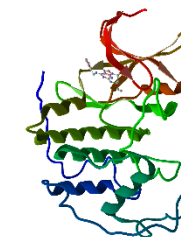
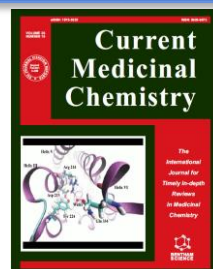
Data: 14 de setembro de 2023

Horário: 14h00

<https://meet.google.com/oyg-yfsv-dac>



- [Uma Abordagem de Sistemas](#)
- [Docagem Molecular](#)
- [Aprendizado de Máquina](#)
- [Espaço Químico](#)
- [Espaço de Proteínas](#)
- [Espaço de Funções Escores](#)
- [Estudo da Quinase Dependente de Ciclina 2](#)
- [Projetos em Andamento](#)
- [Rede Internacional de Colaboradores](#)
- [Referência Bibliográfica](#)
- [Agradecimentos](#)
- [Material Adicional](#)





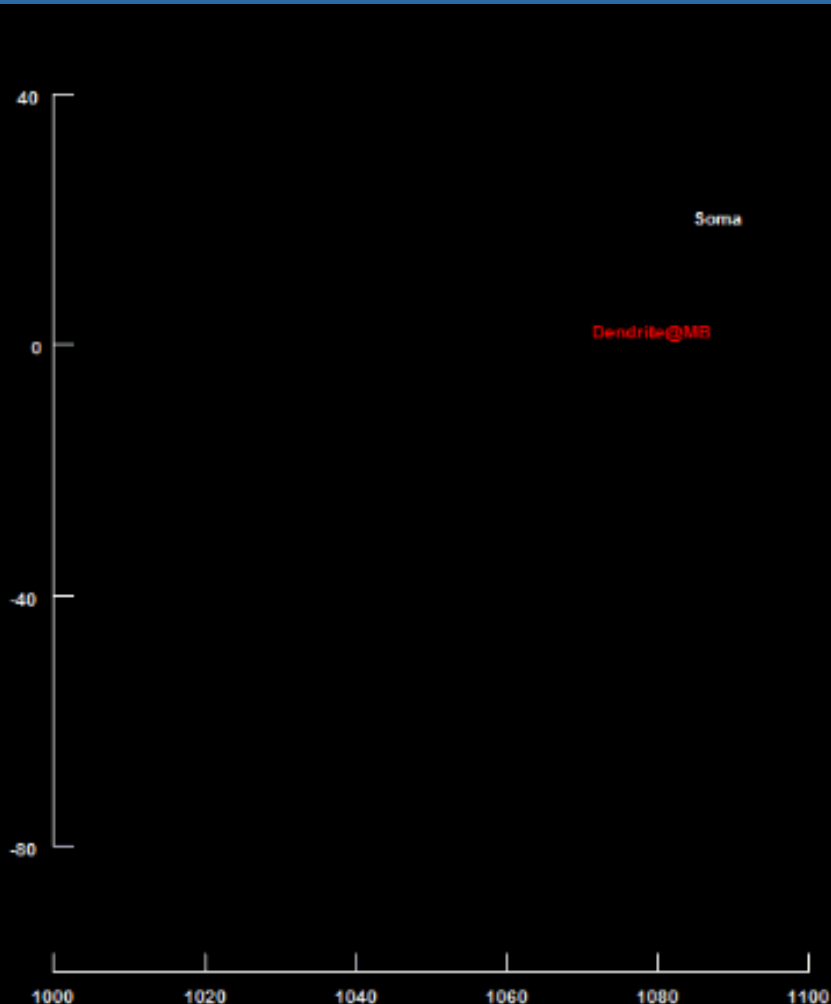
Modelagem Matemática de Sistemas

Modelagem Computacional

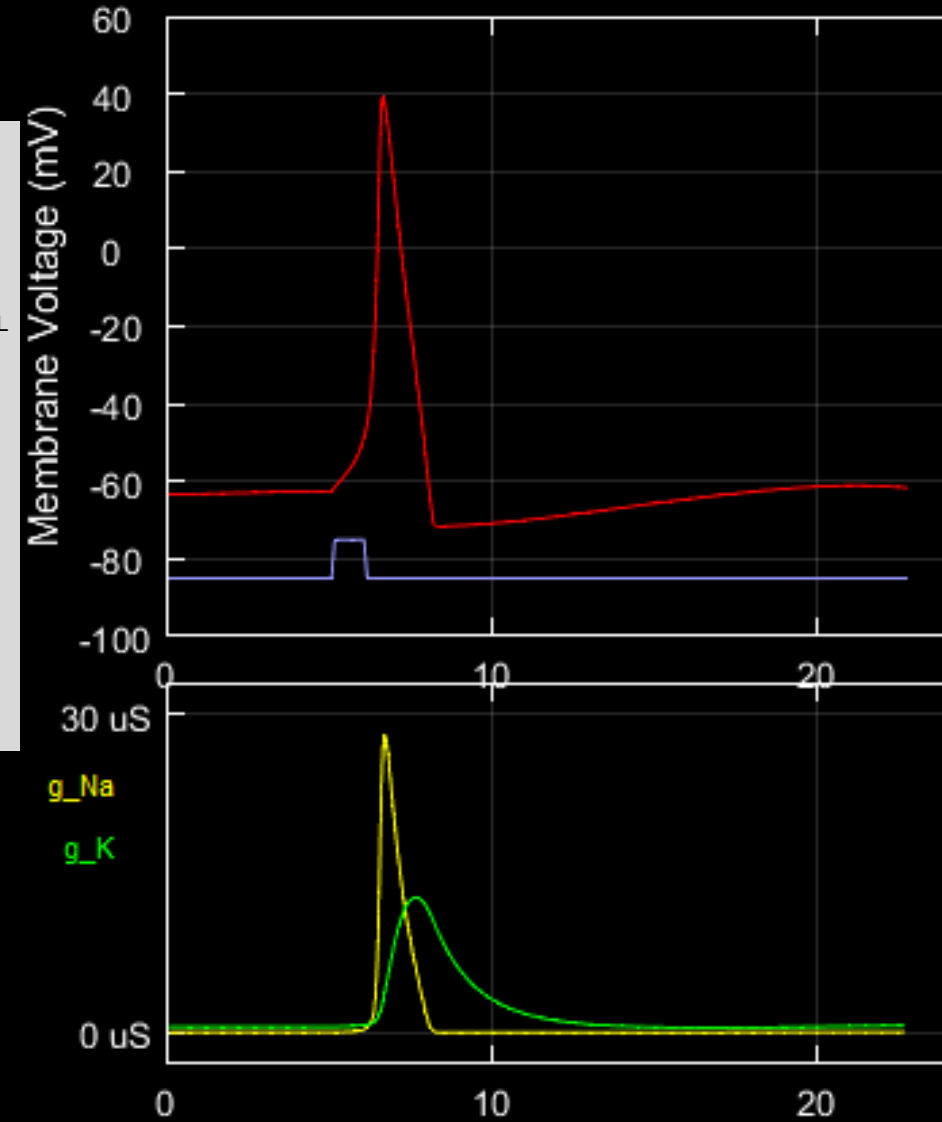
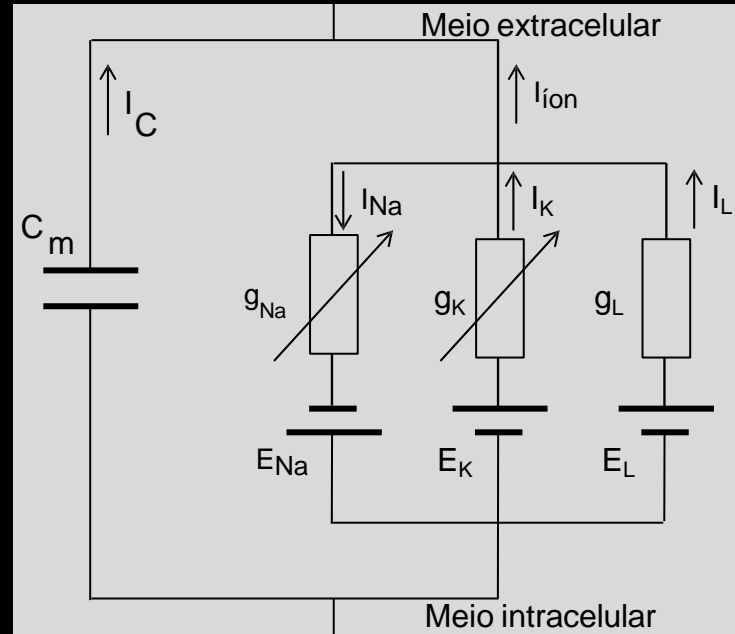
Simulação Computacional

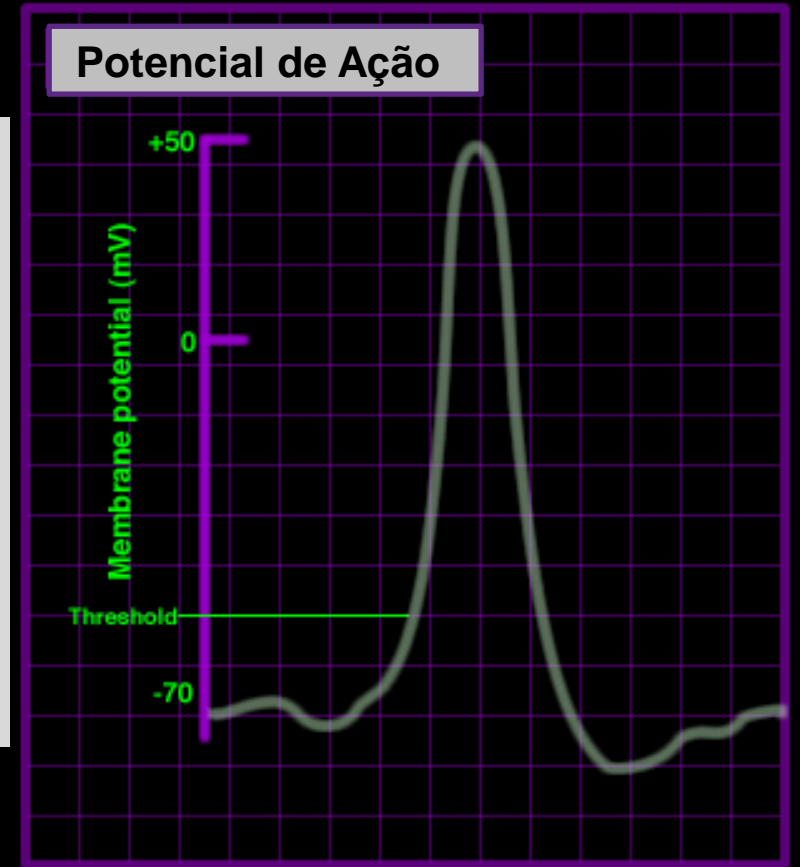
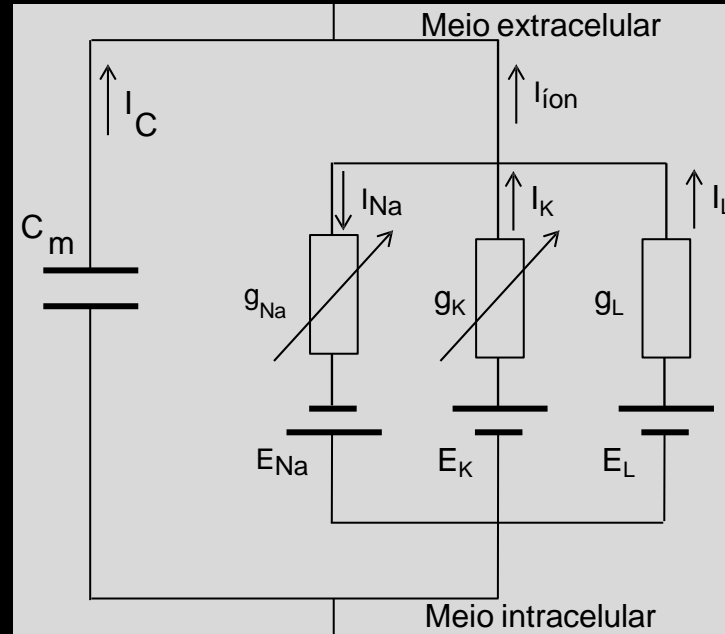
Visão de Sistemas

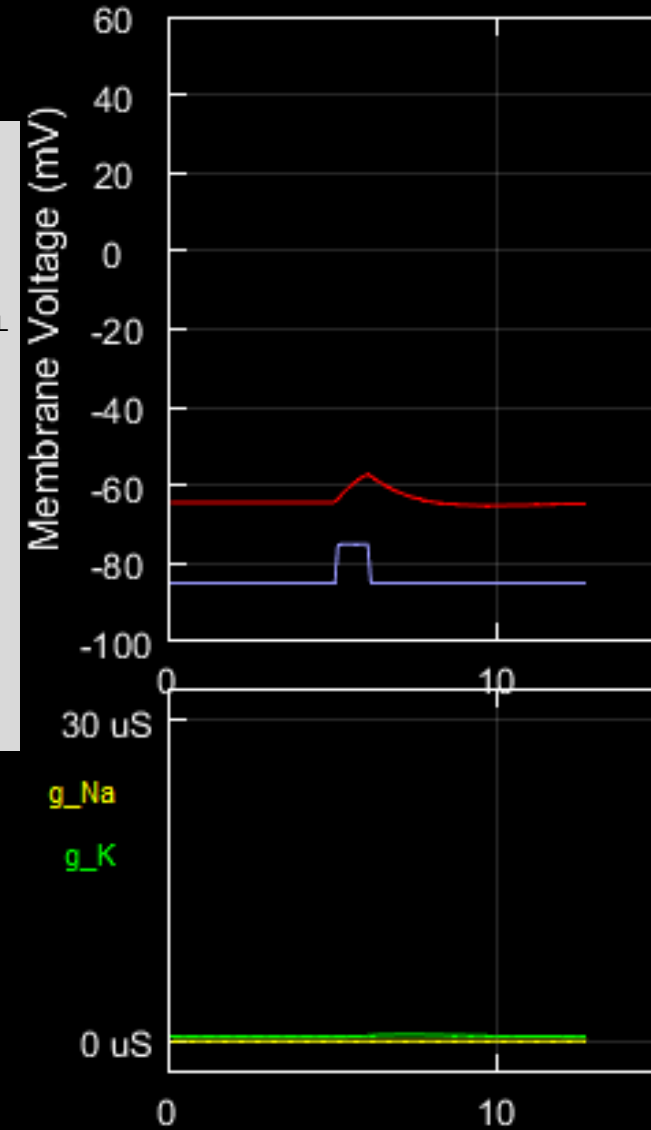
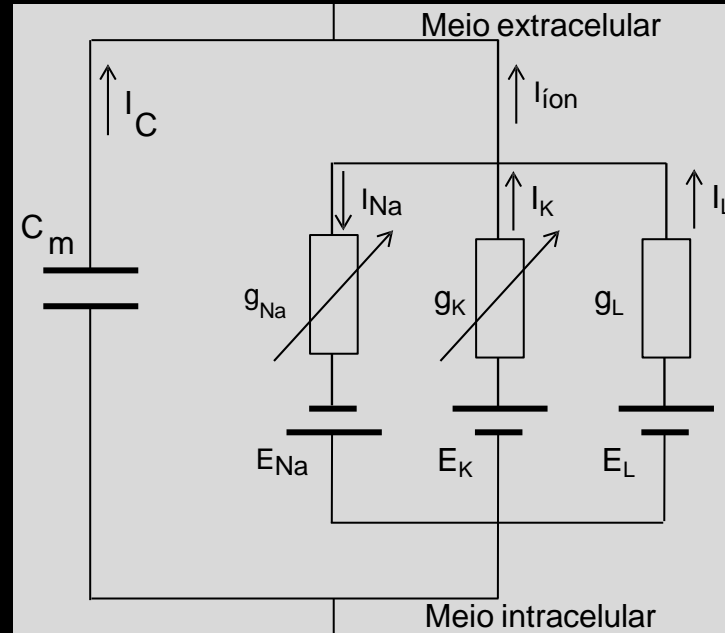
Previsão de Resultados

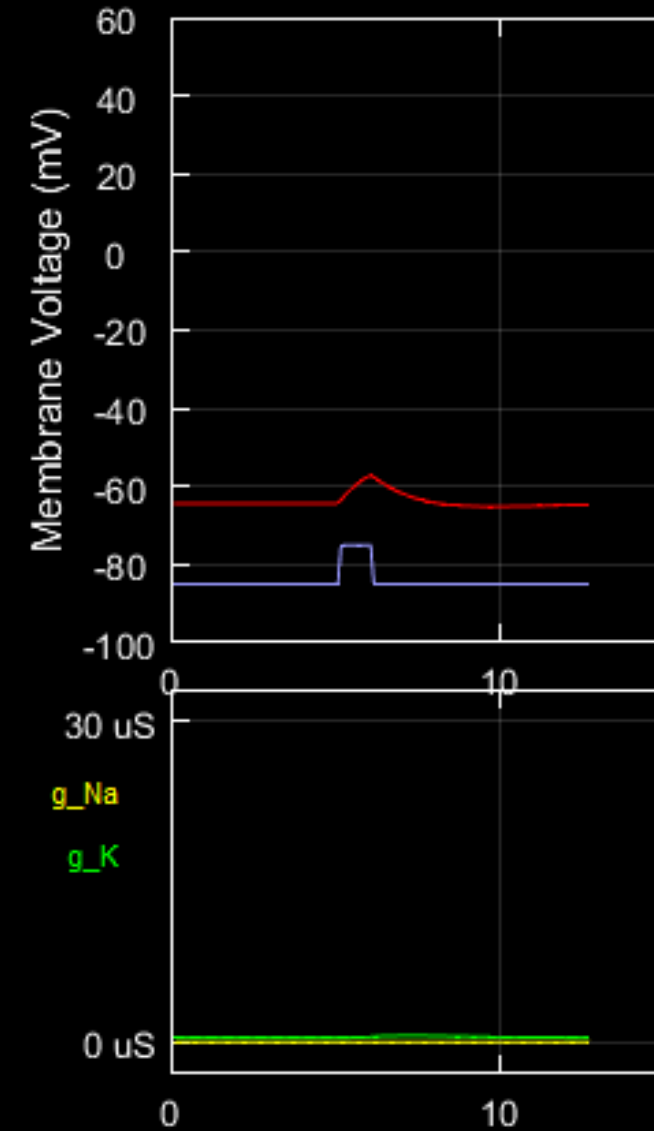
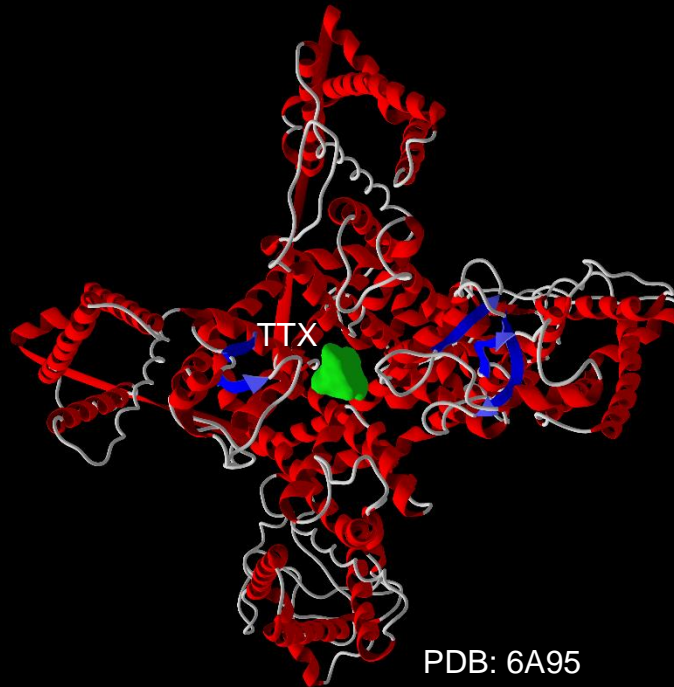


O primeiro modelo desenvolvido em biologia de sistemas computacional foi o potencial de ação do axônio de sépia (Levchenko, 2001) proposto por Hodgkin-Huxley em 1952.

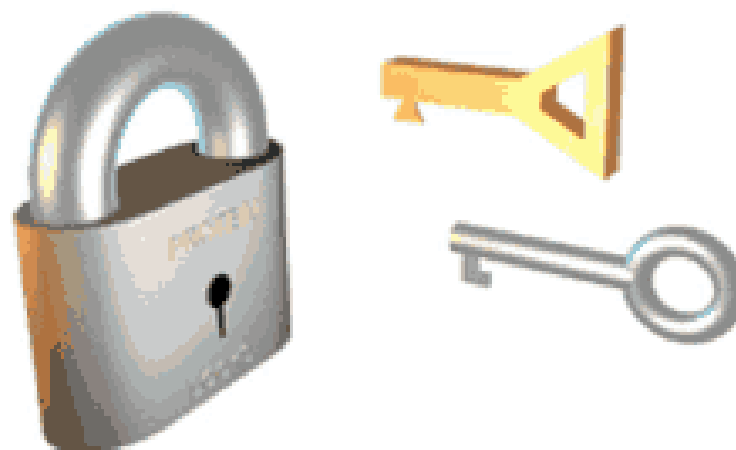


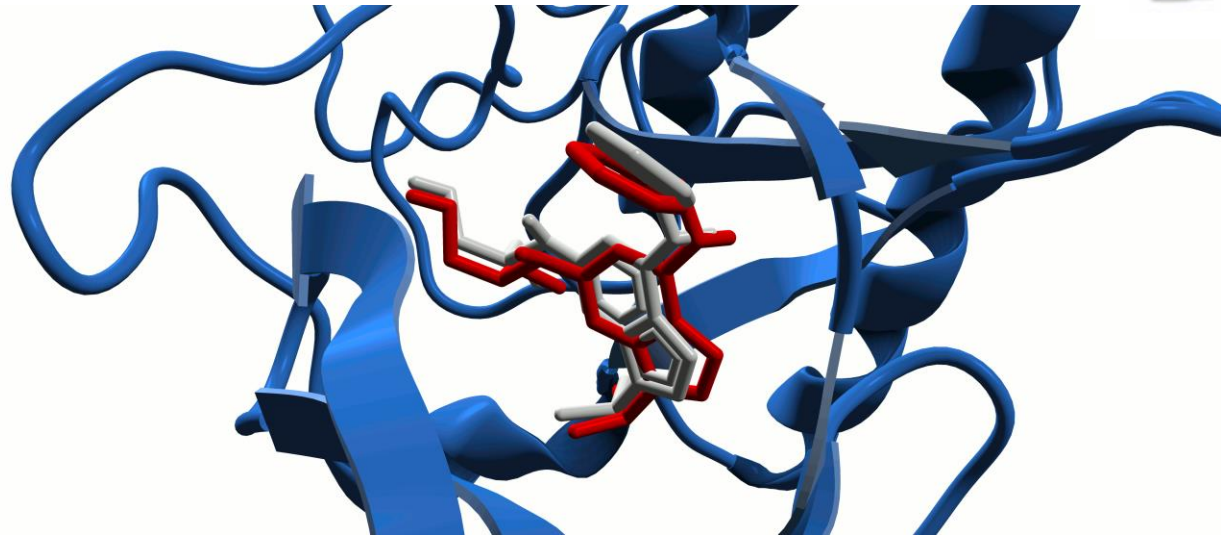
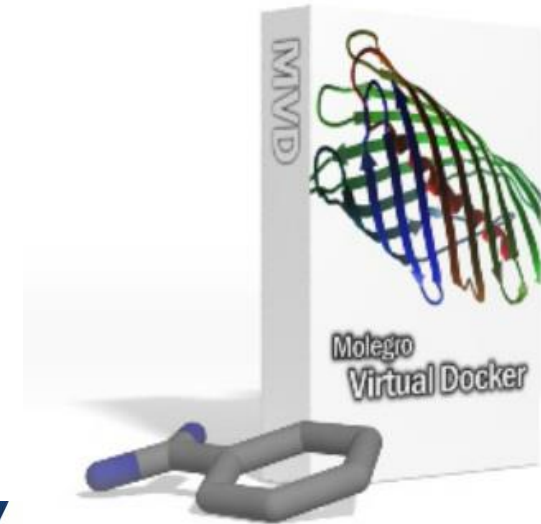










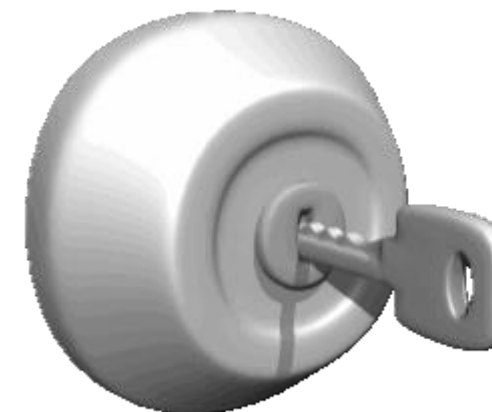


$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) +$$

$$W_{HBond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right)$$

$$W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} +$$

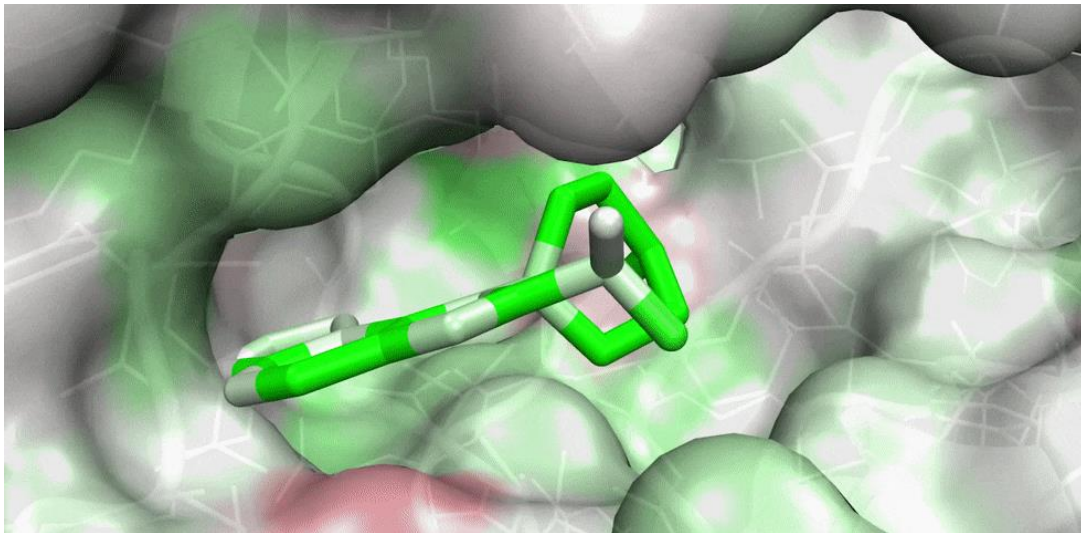
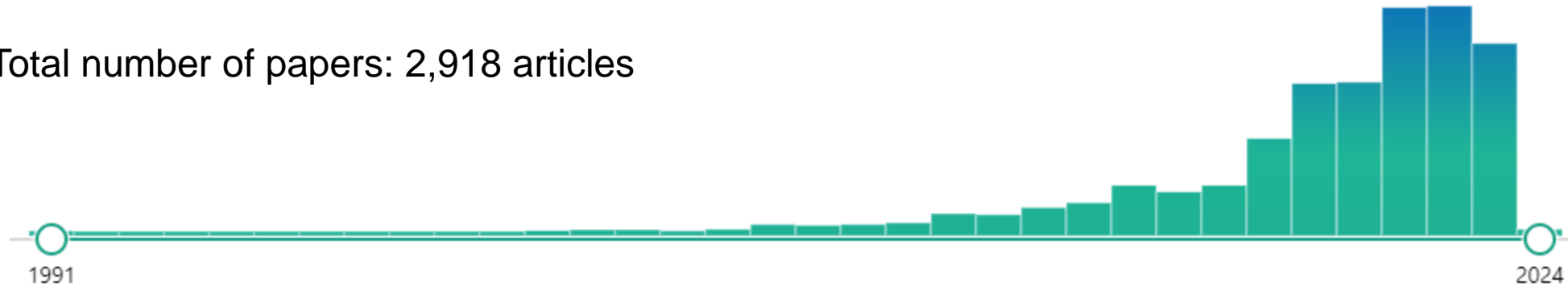
$$W_{sol} \sum_{i,j} (S_i v_j + S_j v_i) e^{\left(-r_{ij}^2 / 2\sigma^2 \right)} + W_{tor} N_{Tor}$$





Evolution of the Number of Papers by Year: Machine Learning for Drug Discovery

Total number of papers: 2,918 articles

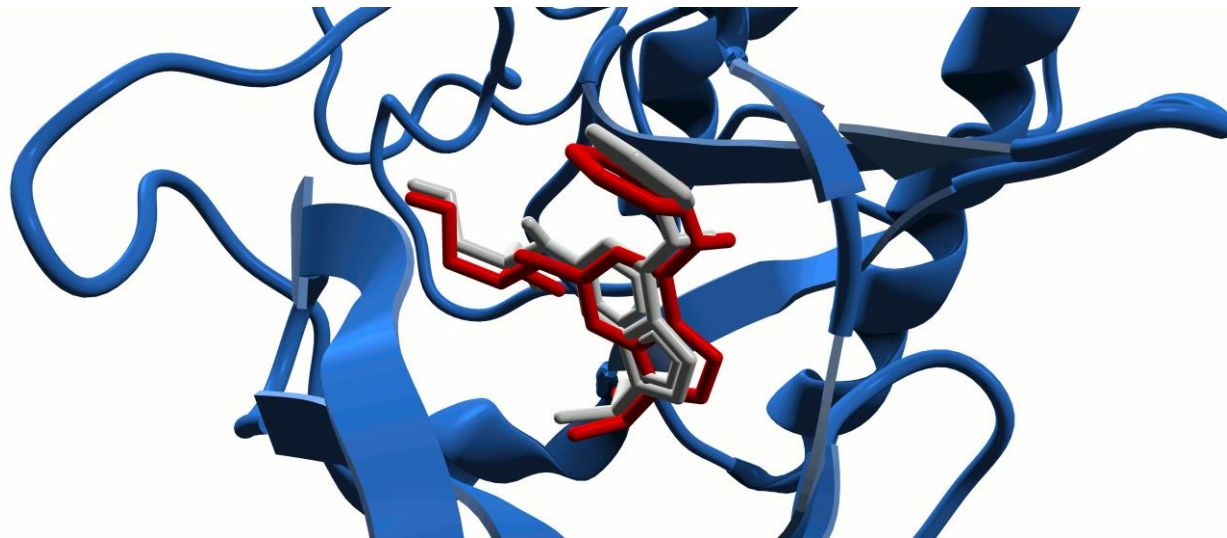
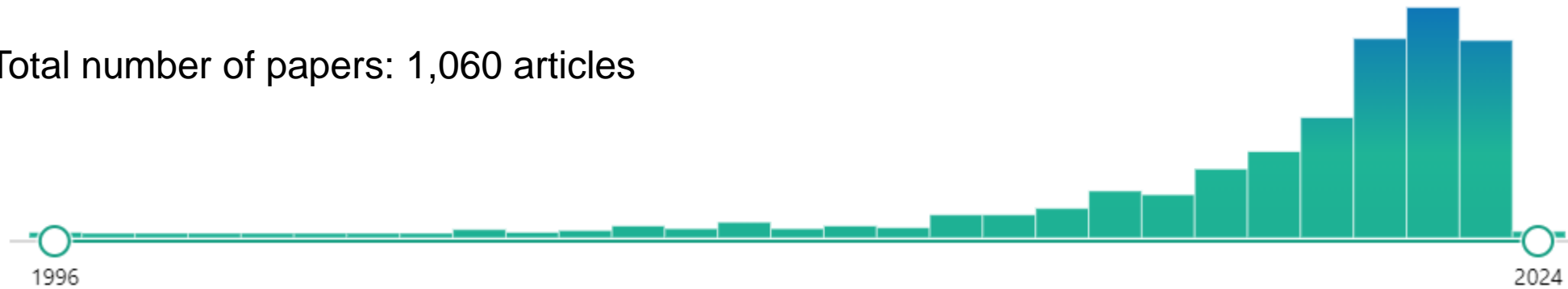


Source: Pubmed on September 12, 2023



Evolution of the Number of Papers by Year: Machine Learning for Docking

Total number of papers: 1,060 articles

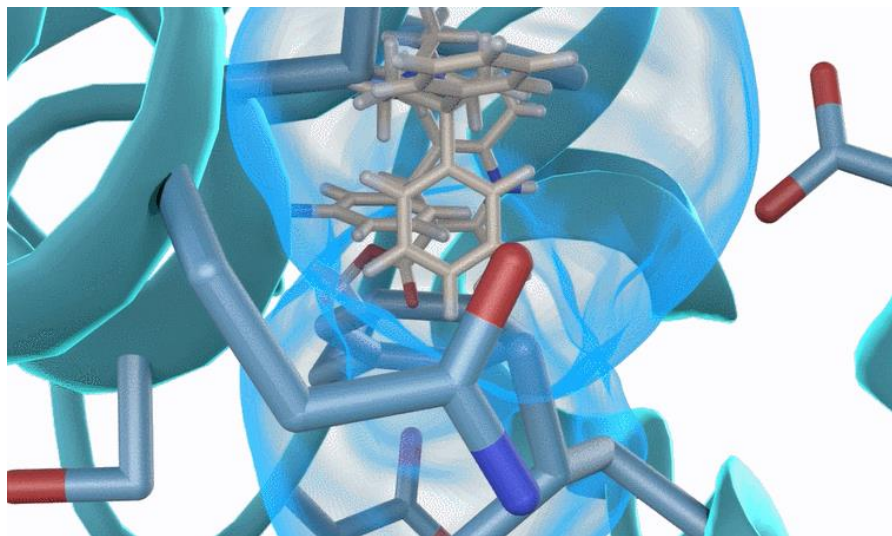
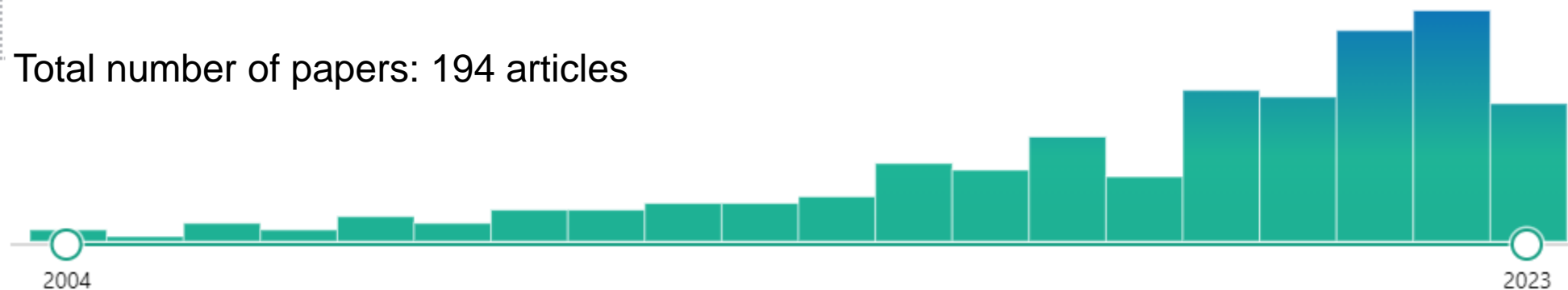


Source: Pubmed on September 12, 2023

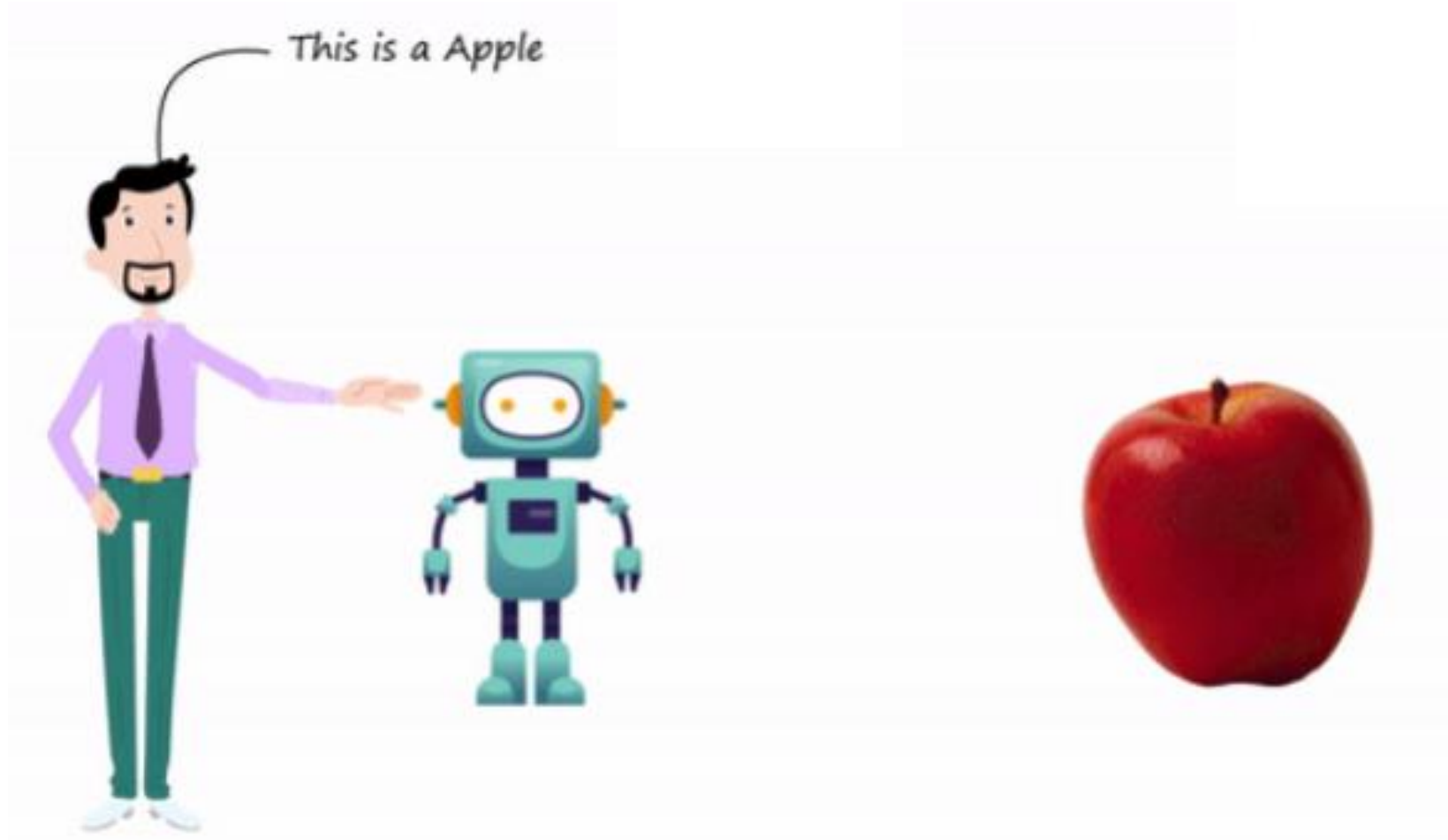


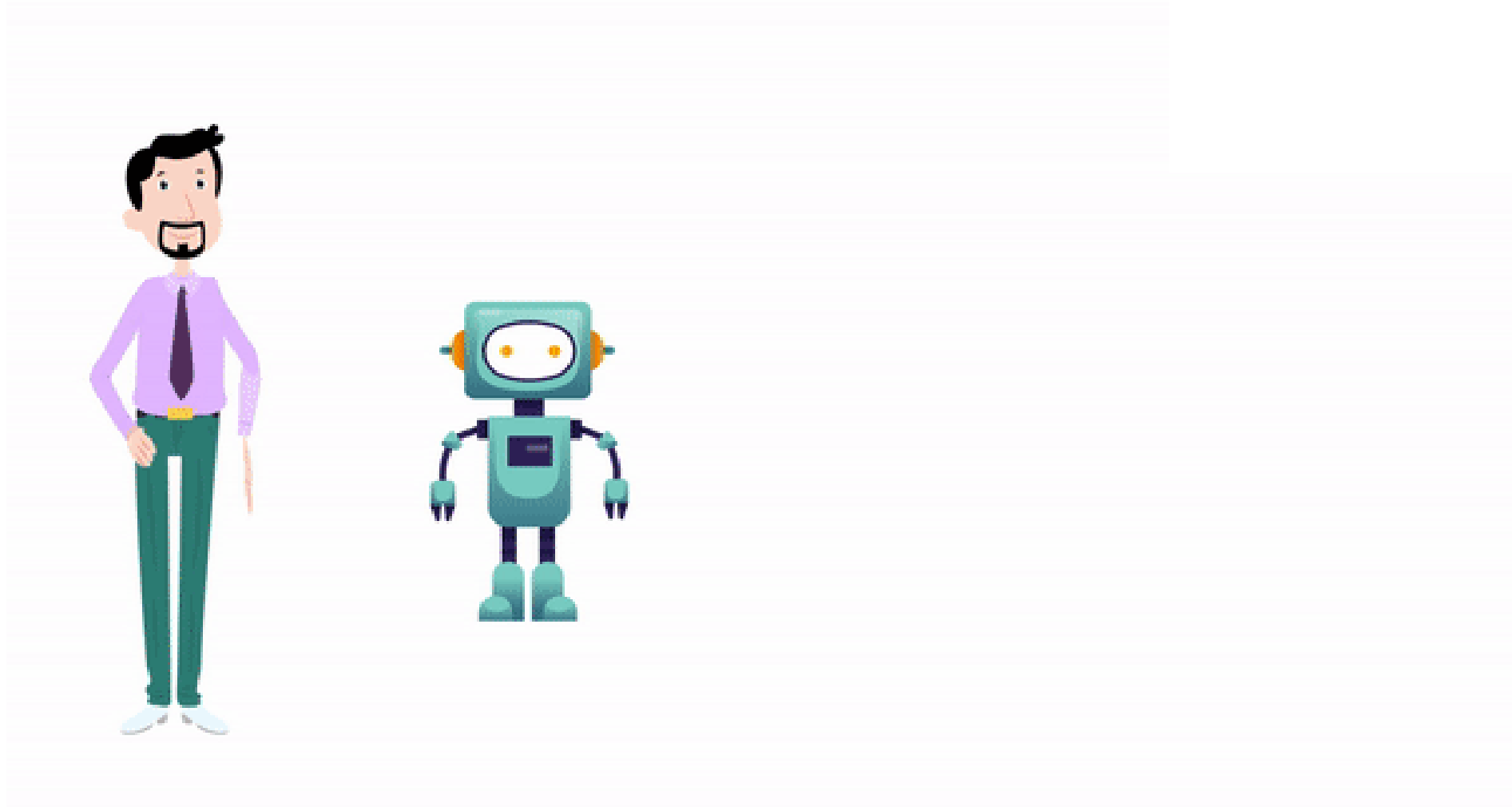
Evolution of the Number of Papers by Year: Machine Learning for Scoring Function

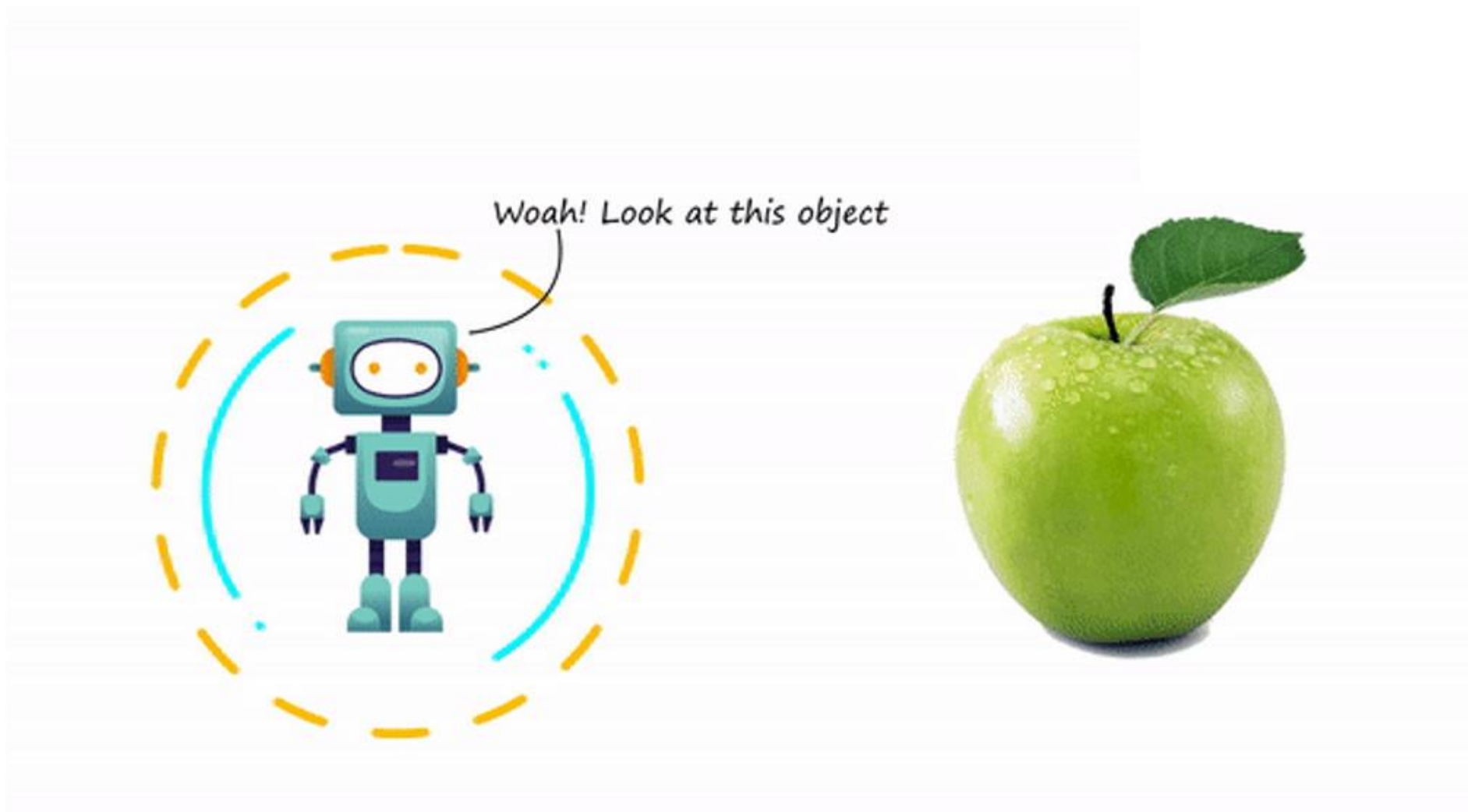
Total number of papers: 194 articles

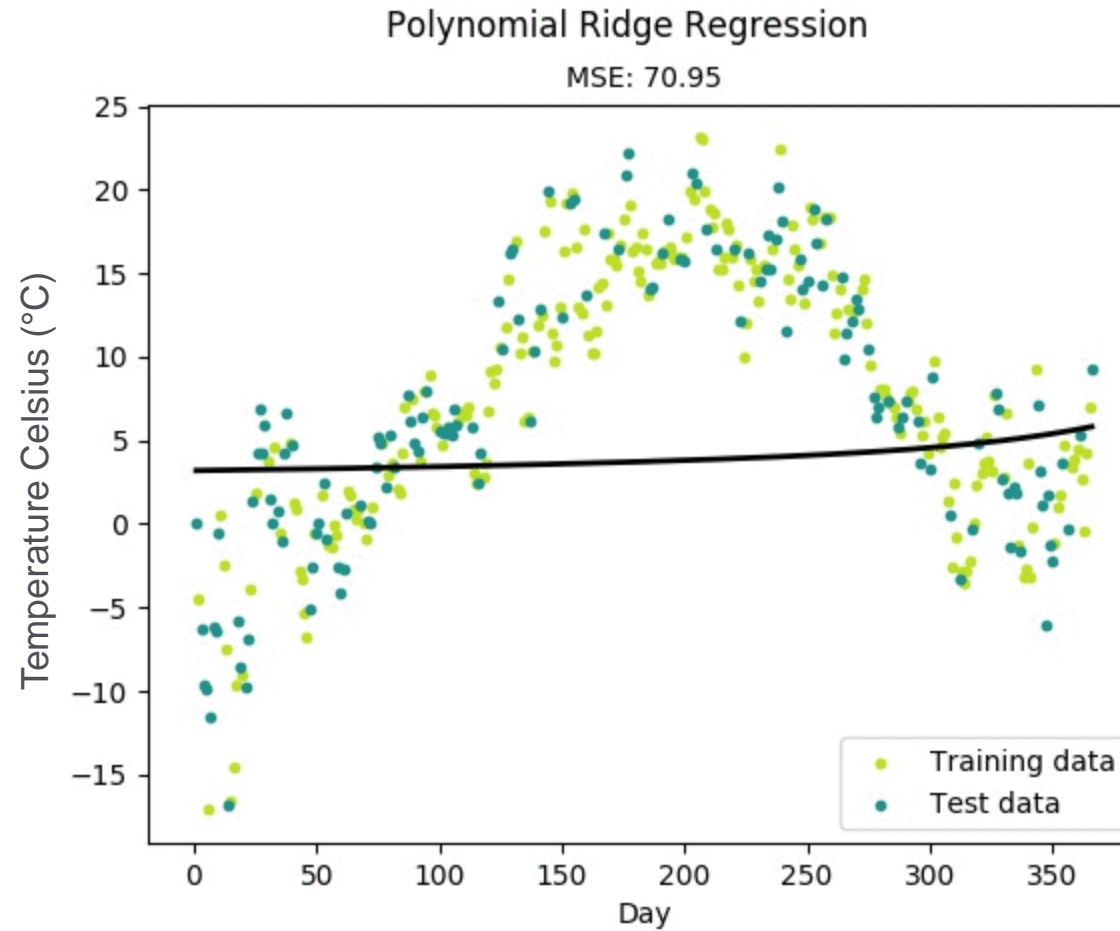


Source: Pubmed on September 12, 2023











Modelo de regressão

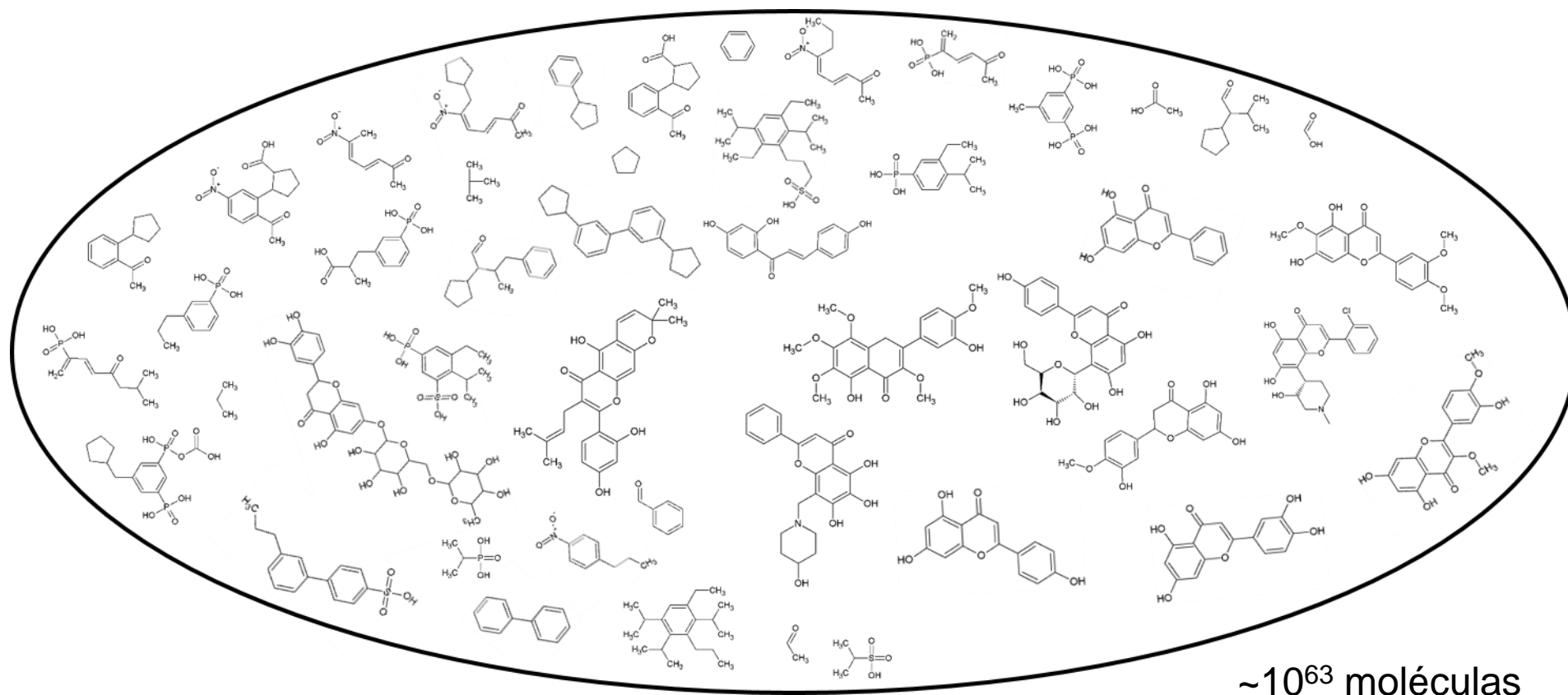
$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

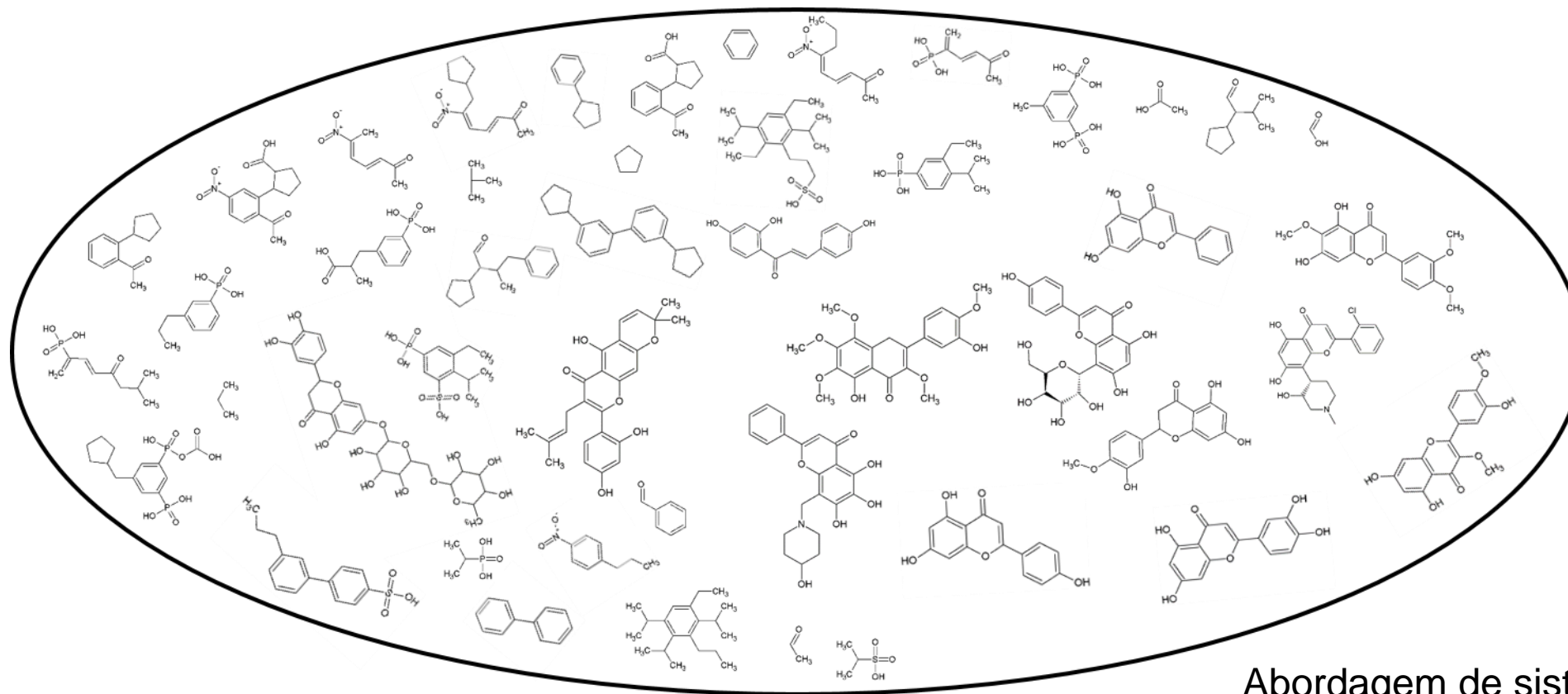
\hat{y} valor previsto

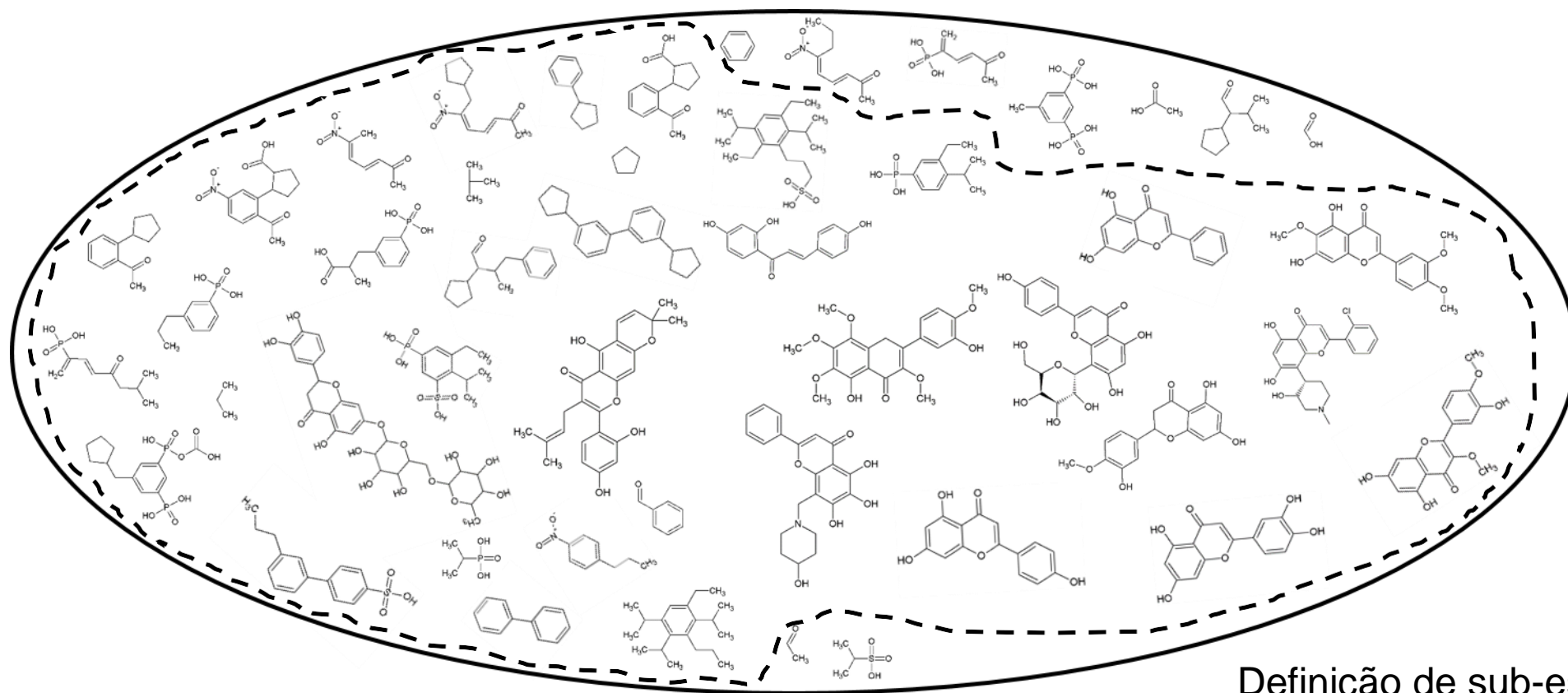
θ_j é j^{th} parâmetro do modelo

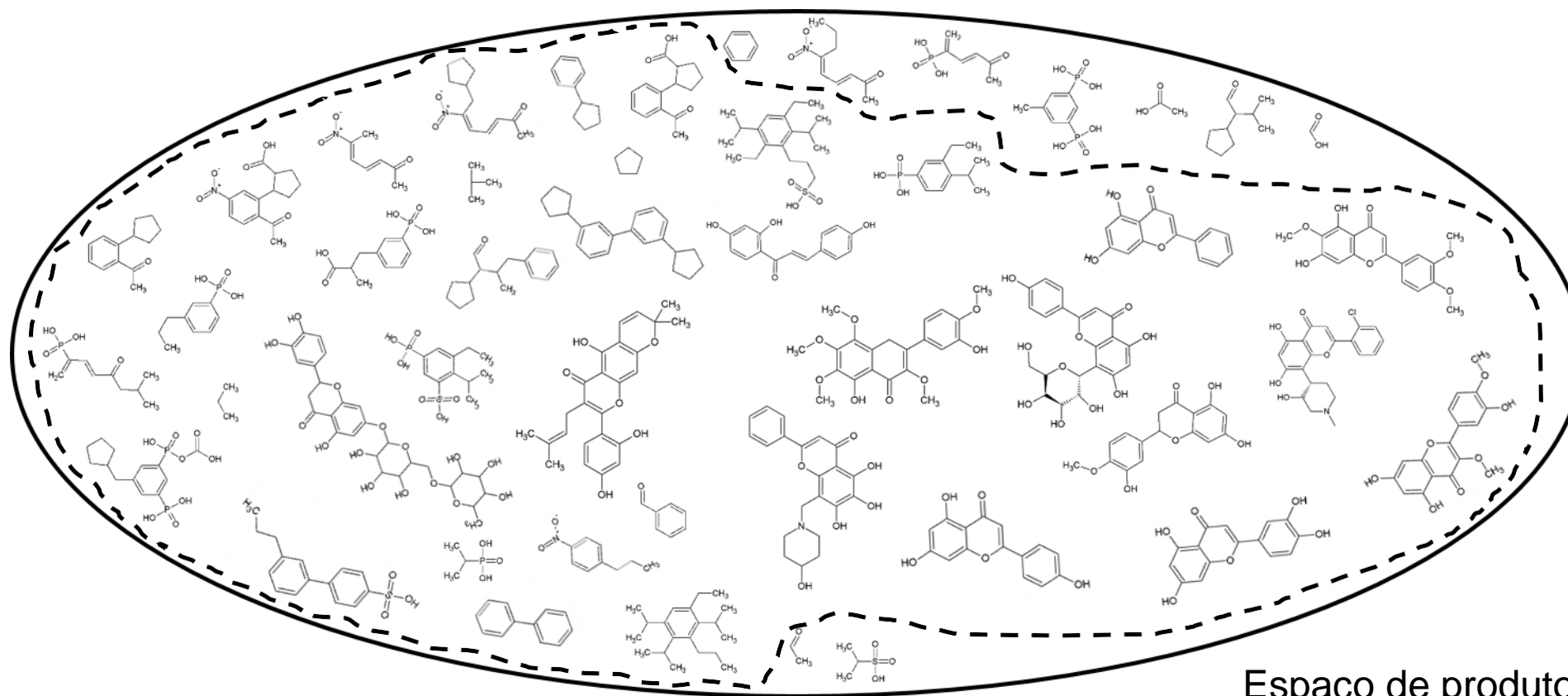
x_i é i^{th} valor da variável independente (*feature*)

n é o número de variáveis independentes (*features*)

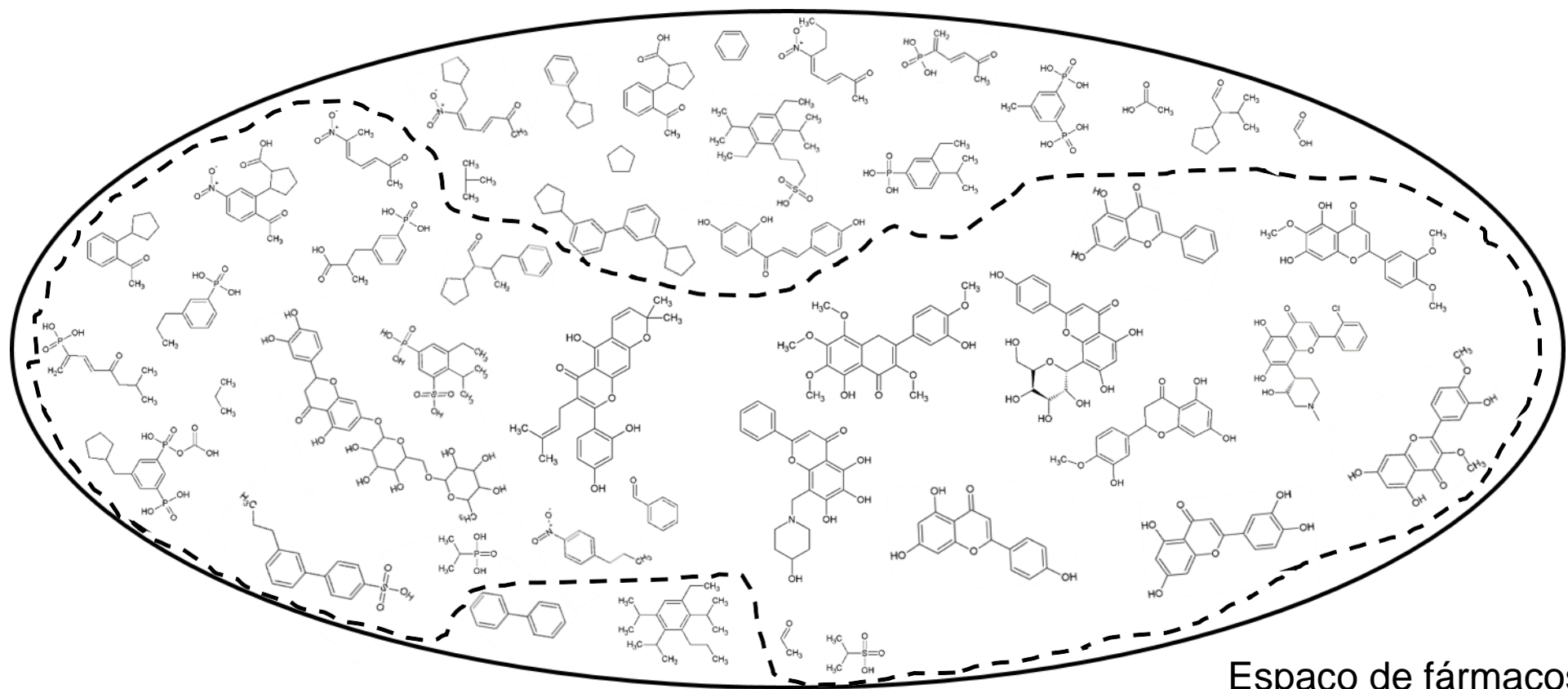


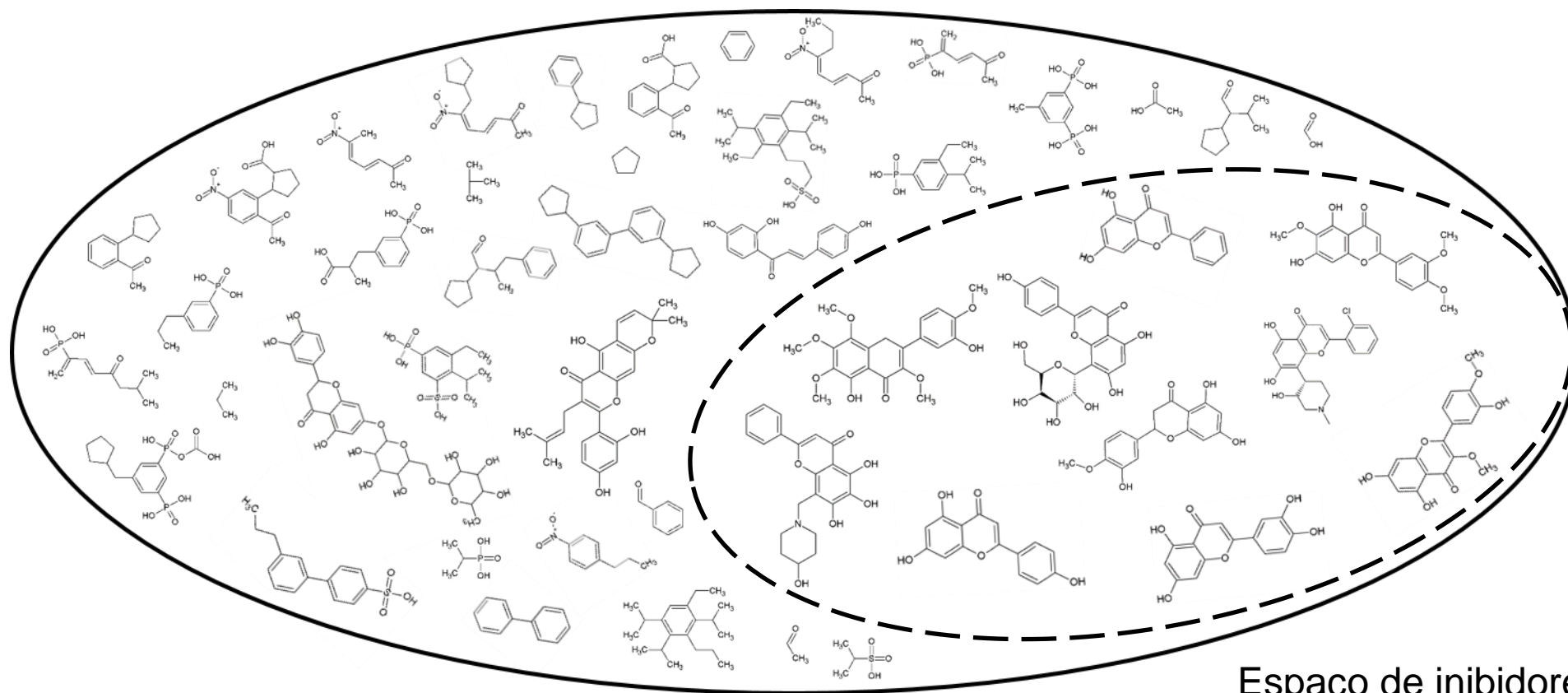




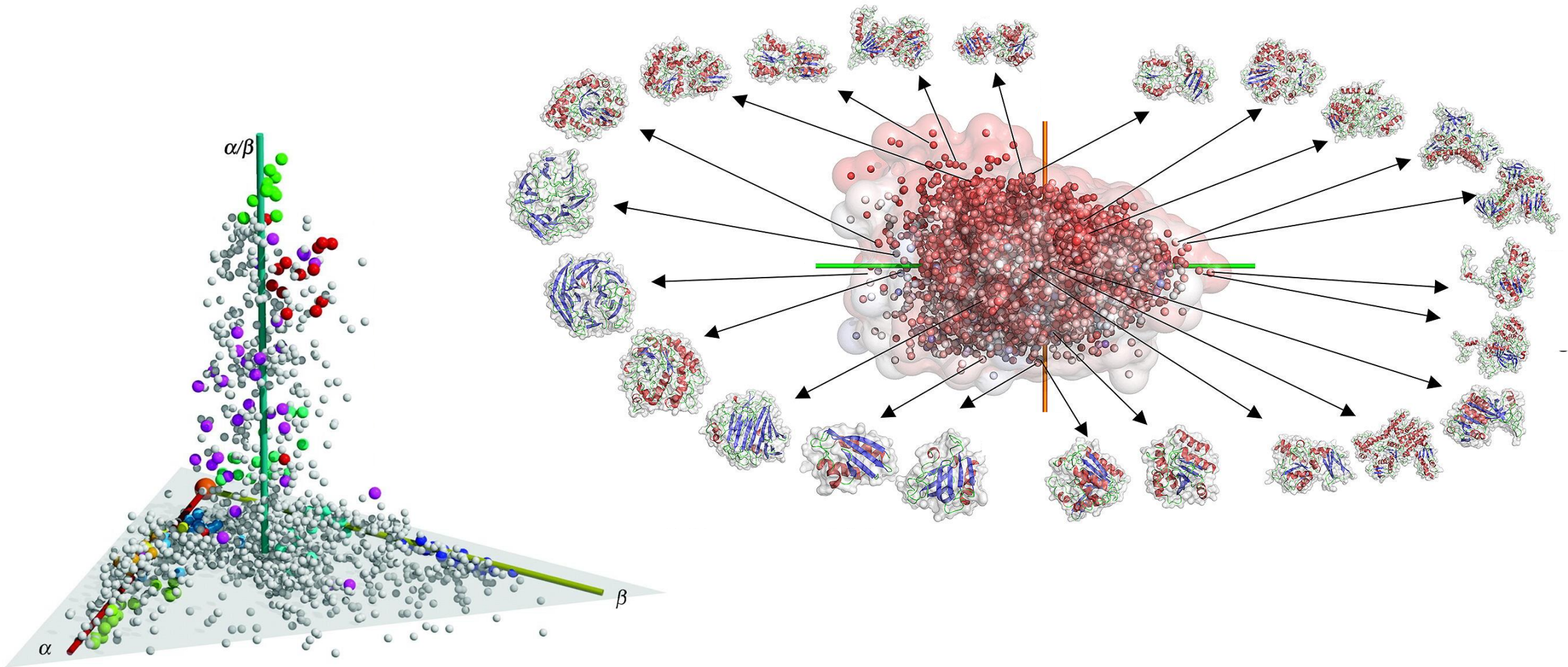


Espaço de produtos naturais





Espaço de inibidores de uma enzima específica





$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{x-iy}{x^2+y^2}$$

$$Q = A = \frac{m}{M} RT \ln \frac{V_2}{V_1}$$

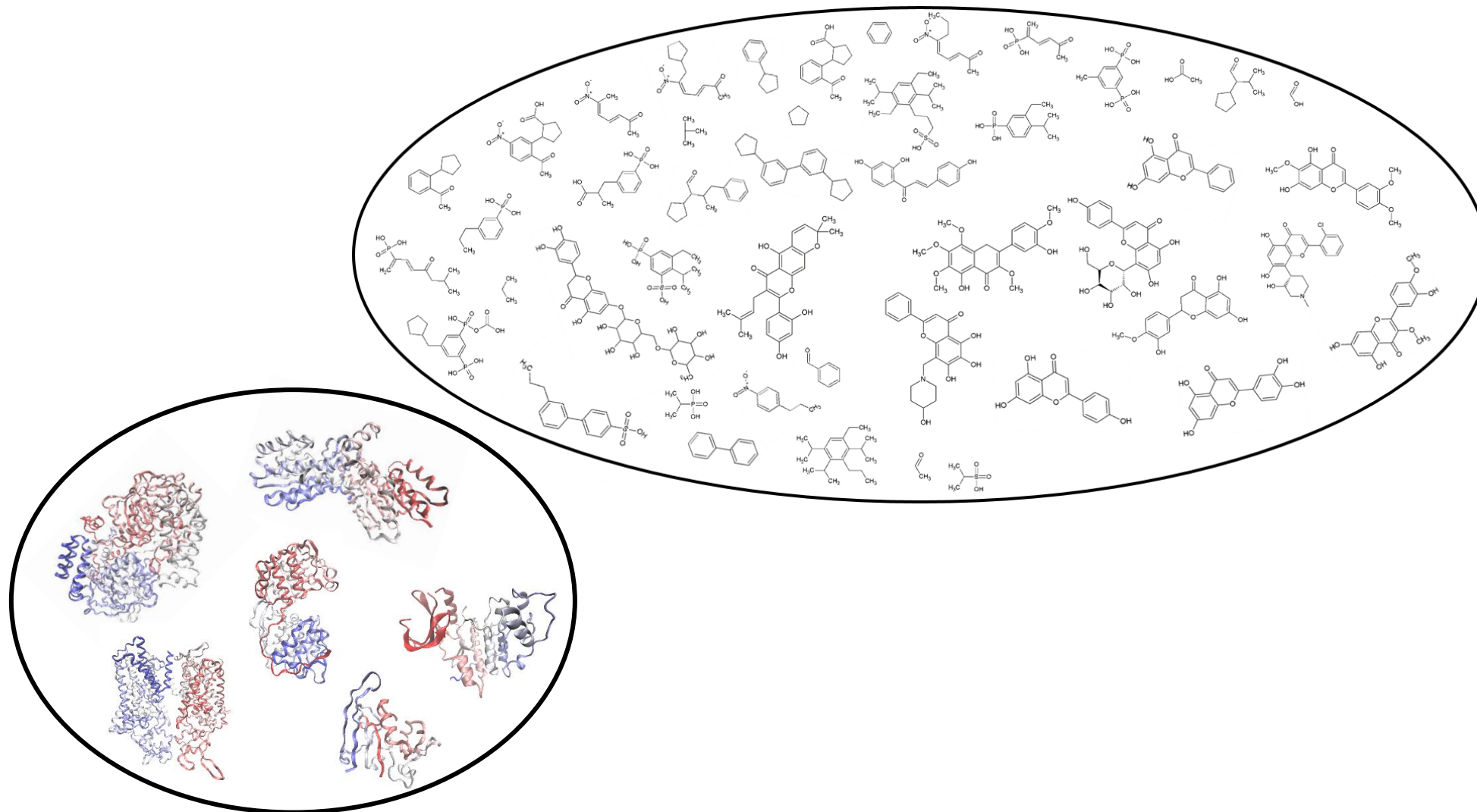
The image displays a complex, overlapping pattern of mathematical equations. The primary equation visible is the complex conjugate identity:

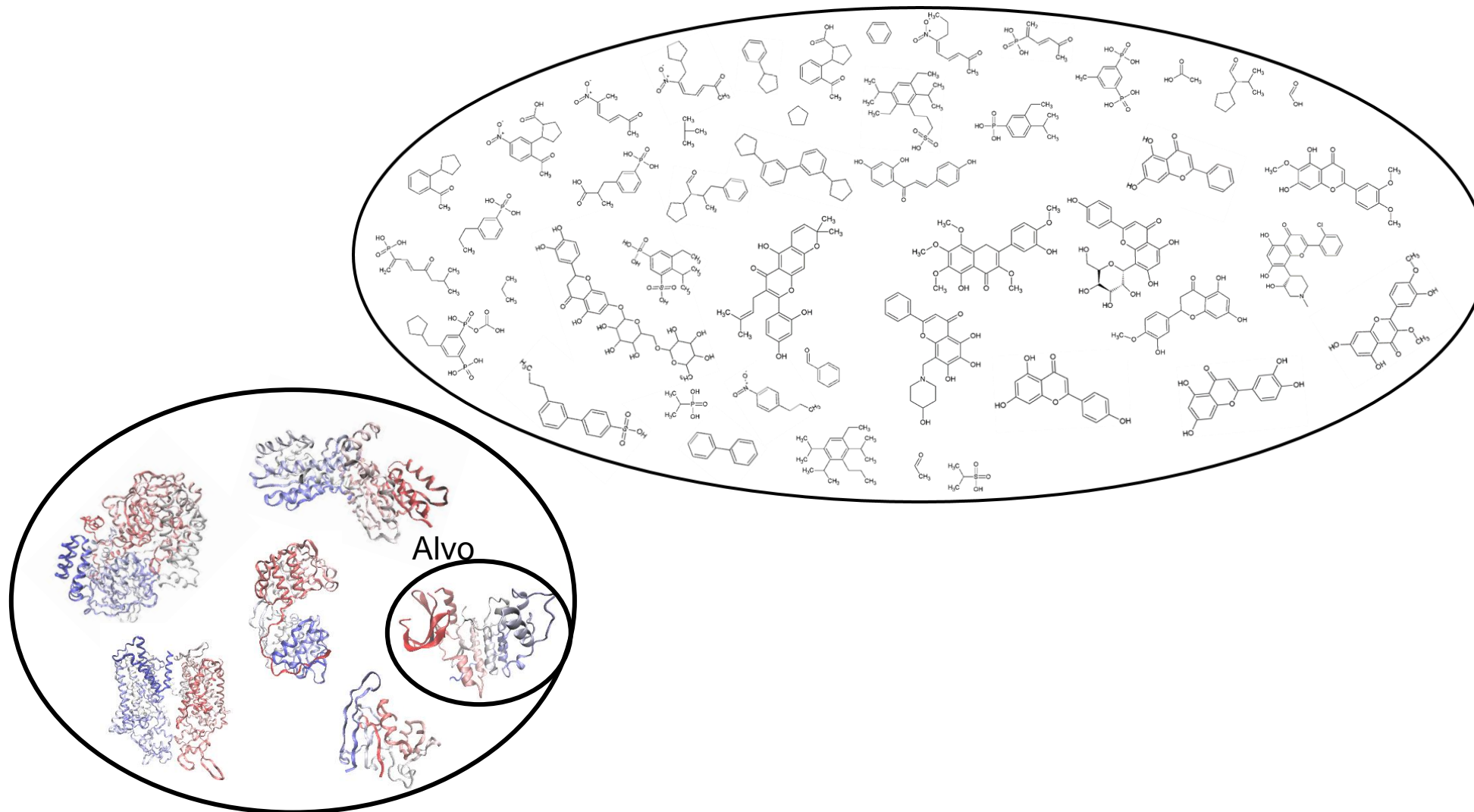
$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{x-iy}{x^2+y^2}$$
 where $z = x + iy$ and $\bar{z} = x - iy$.

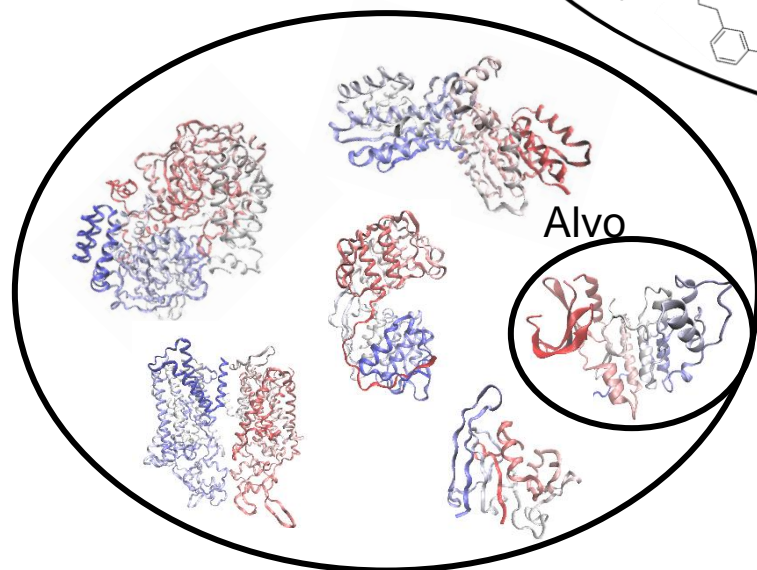
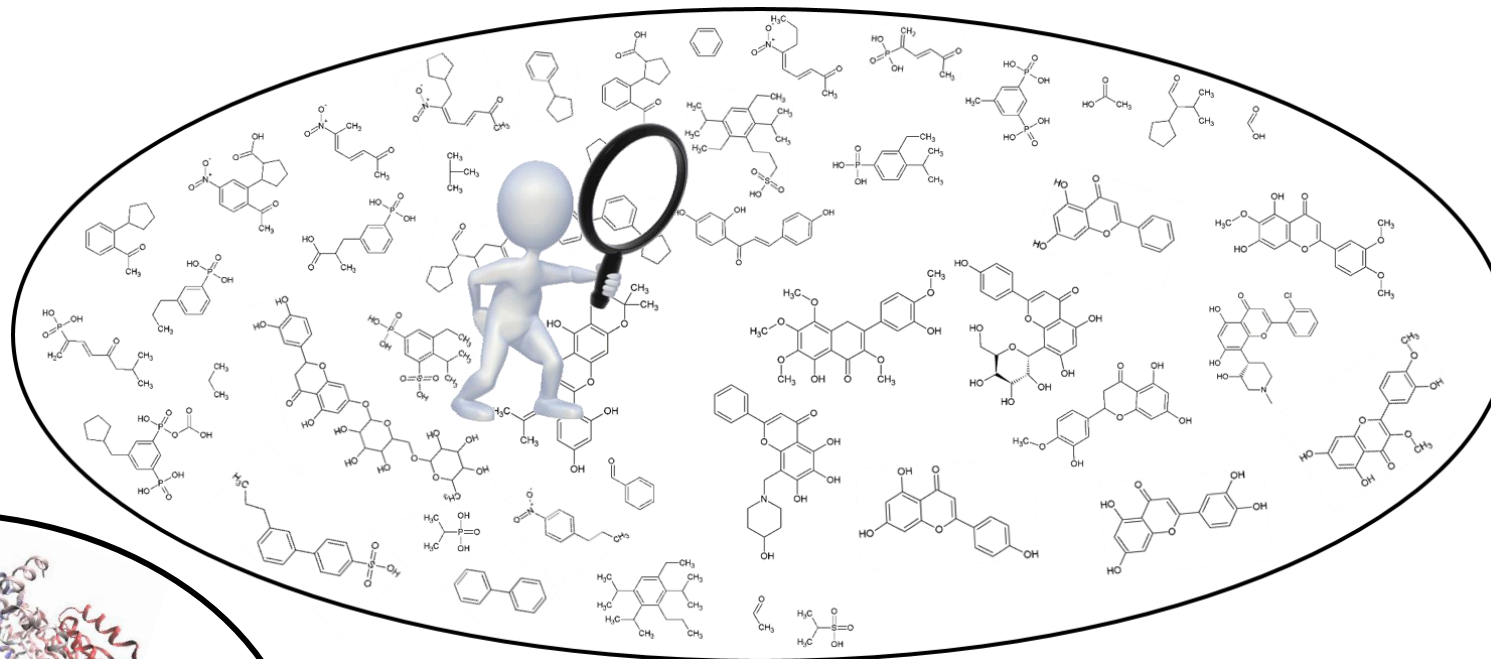
Another prominent equation is the thermodynamic expression for heat Q and work A in an isothermal expansion of an ideal gas:

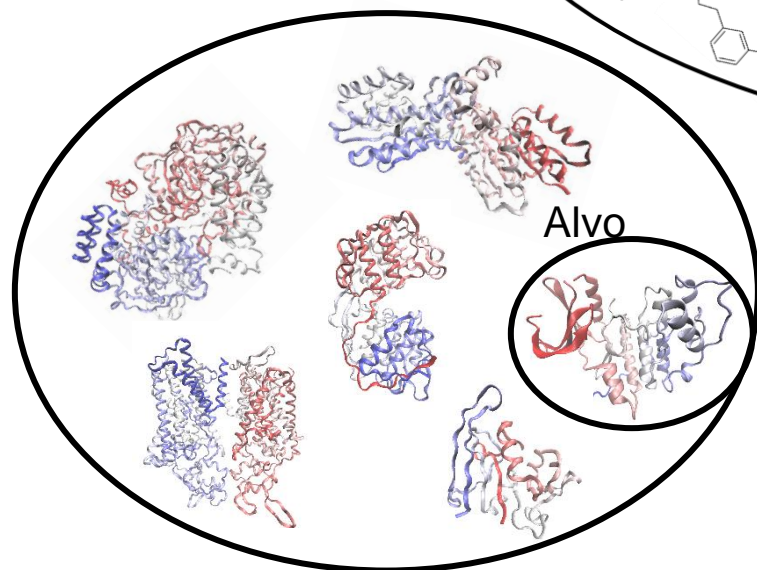
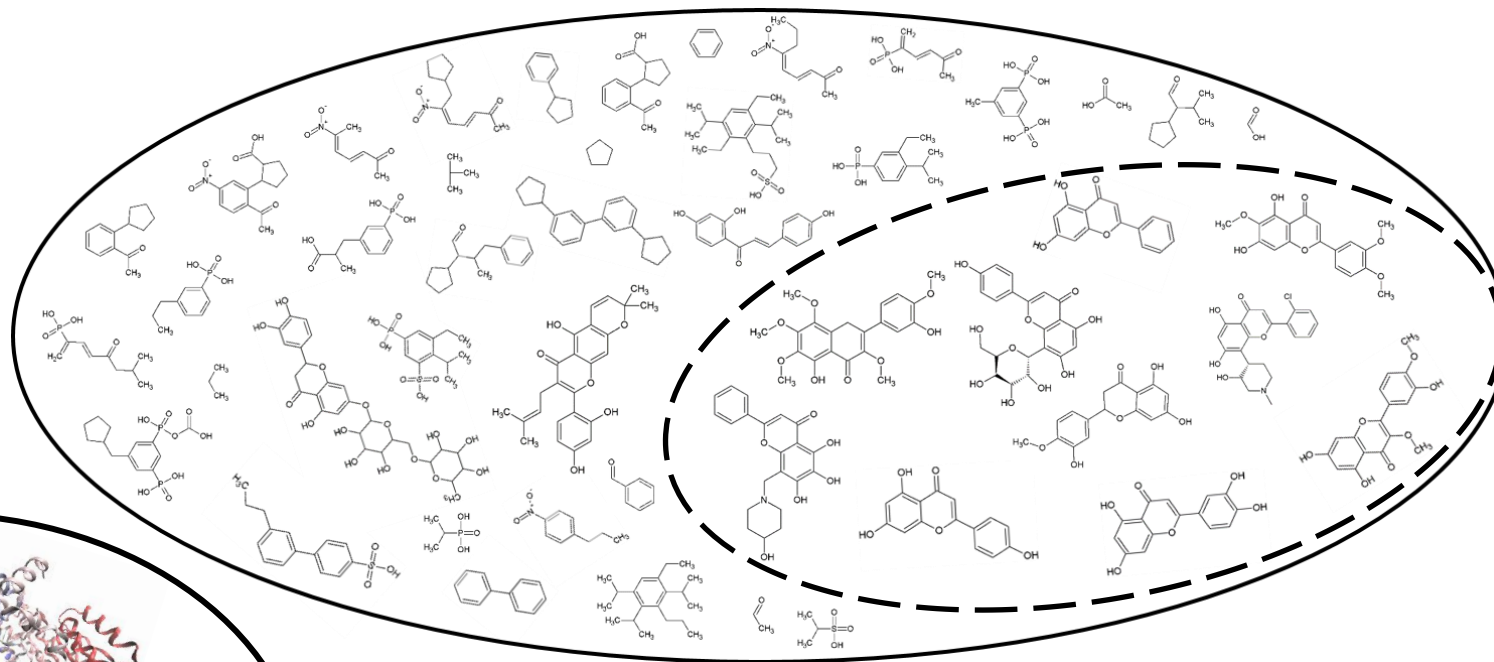
$$Q = A = \frac{m}{M} RT \ln \frac{V_2}{V_1}$$

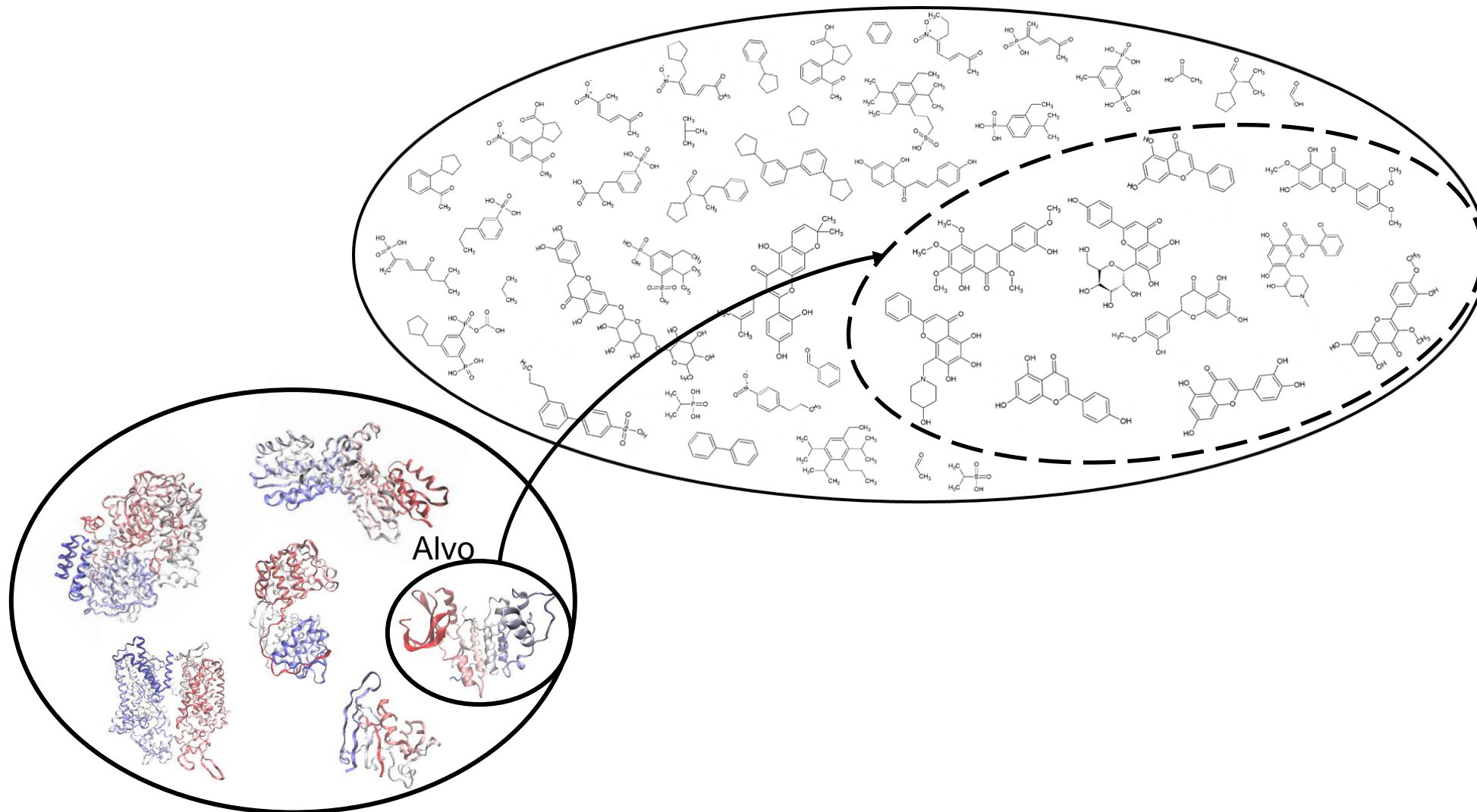
The background is filled with numerous smaller, semi-transparent versions of these equations, creating a dense, layered visual effect.

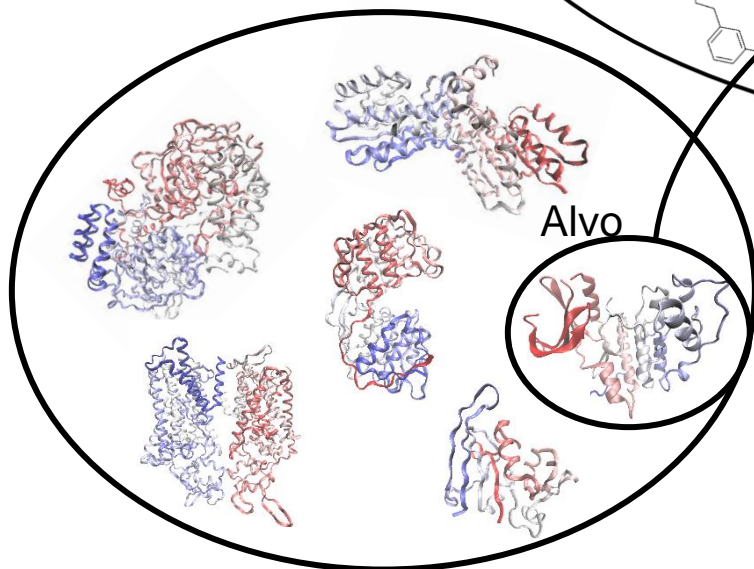
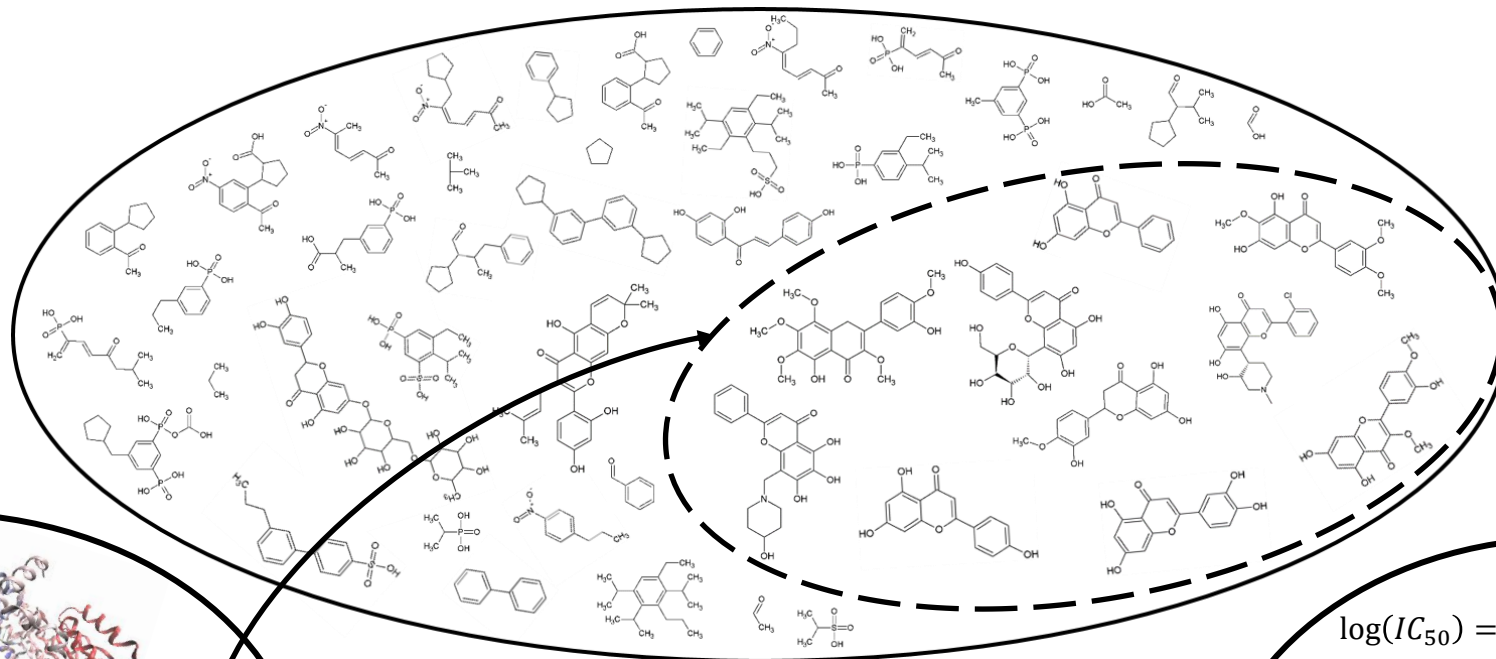












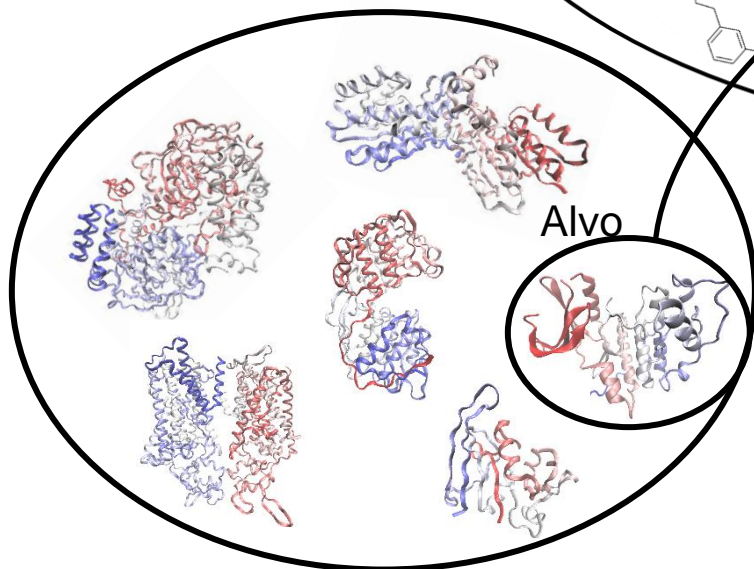
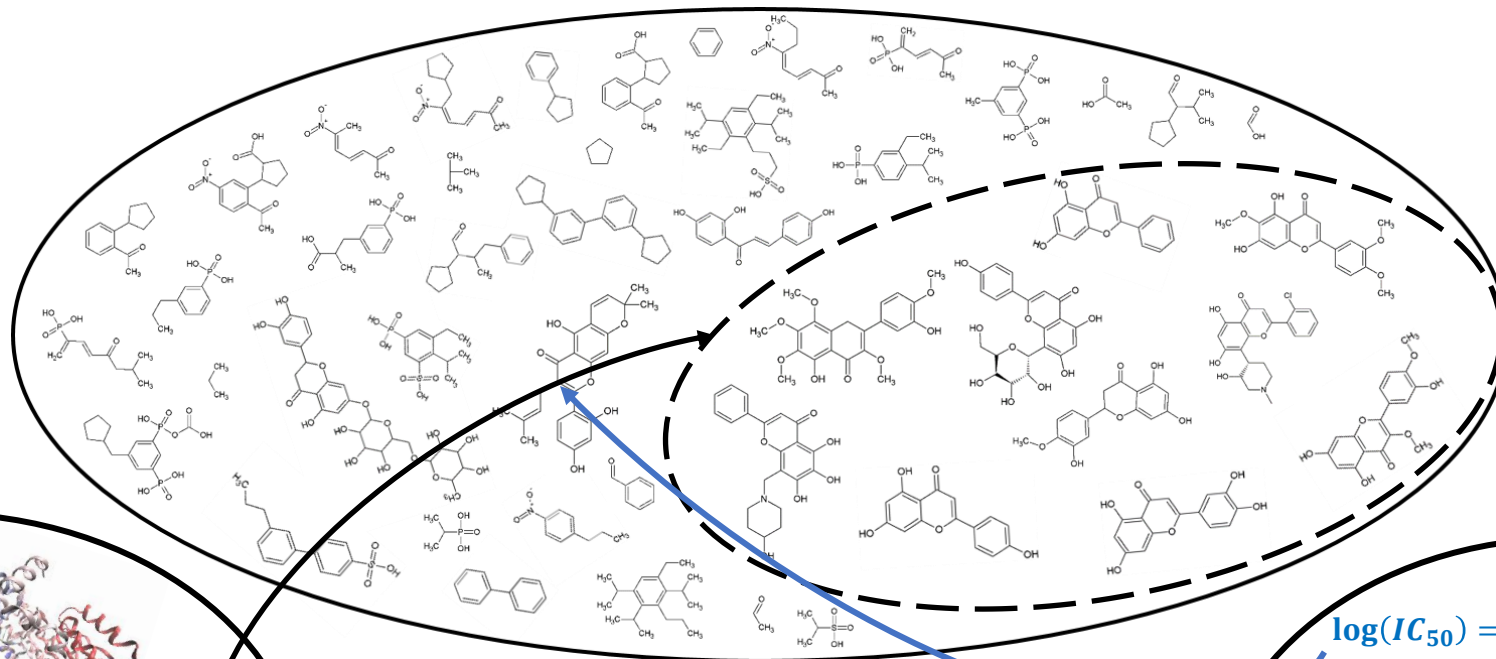
TABA

$$\log(IC_{50}) = \sum_{i=0}^N \omega_i x_i + \sum_{j=0}^N \alpha_j x_j^i$$

$$\Delta G = \sum_{i=0}^N \omega_i x_i \quad f = \sum_{i=1}^N \alpha_i x_i - x_j^{-3} + \sum_{j=1}^M x$$

$$\Delta S = \alpha_j - x_i \sum_{i=1}^N x_i y_j \quad f = \alpha_j \beta_i + x$$

$$\log(K_I) = \sum_{i=0}^N \omega_i x_i + \sum_{j=1}^M \sum_{l=1}^N \lambda$$



Alvo



TABA

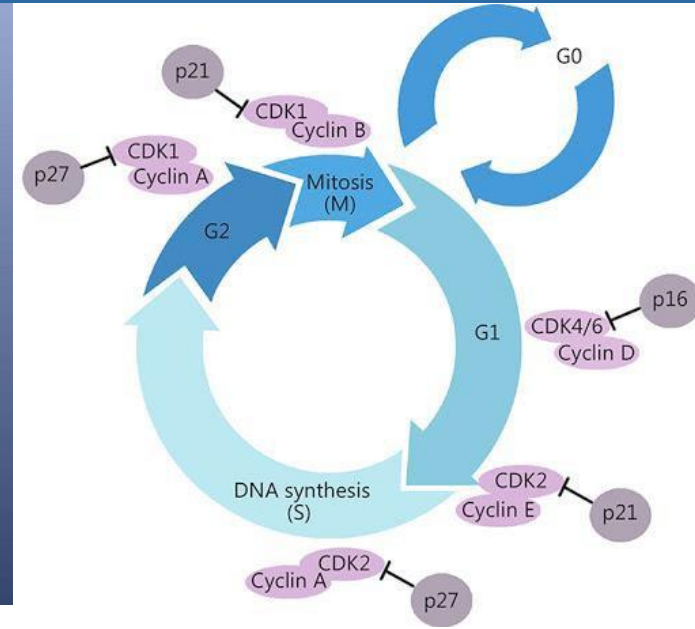
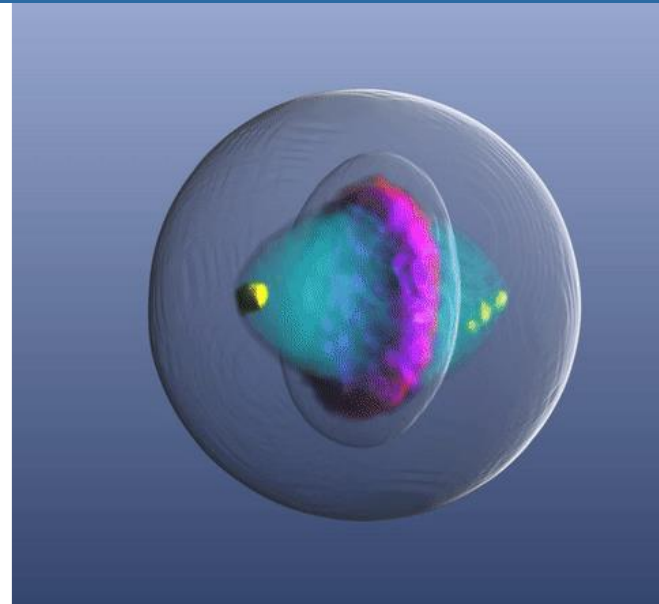
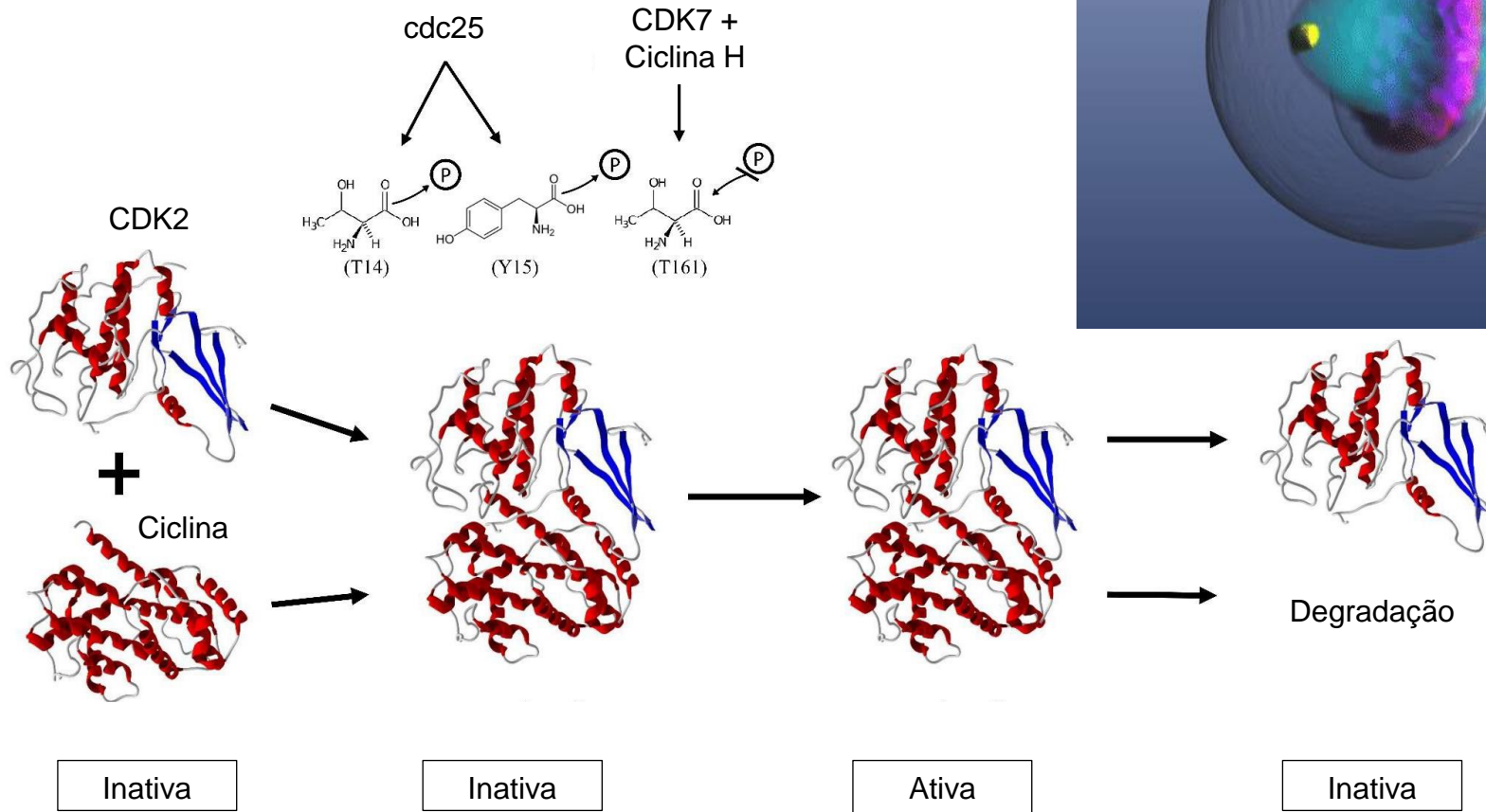
$$\log(IC_{50}) = \sum_{i=0}^N \omega_i x_i + \sum_{j=0}^N \alpha_j x_j^i$$

$$\Delta G = \sum_{i=0}^N \omega_i x_i \quad f = \sum_{i=1}^N \alpha_i x_i - x_j^{-3} + \sum_{j=1}^M x$$

$$\Delta S = \alpha_j - x_i \sum_{i=1}^N x_i y_j \quad f = \alpha_j \beta_i + x$$

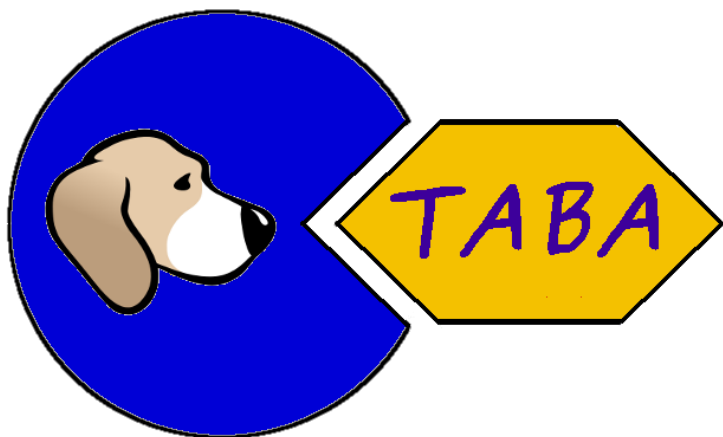
$$\log(K_I) = \sum_{i=0}^N \omega_i x_i + \sum_{j=1}^M \sum_{i=1}^N \lambda$$



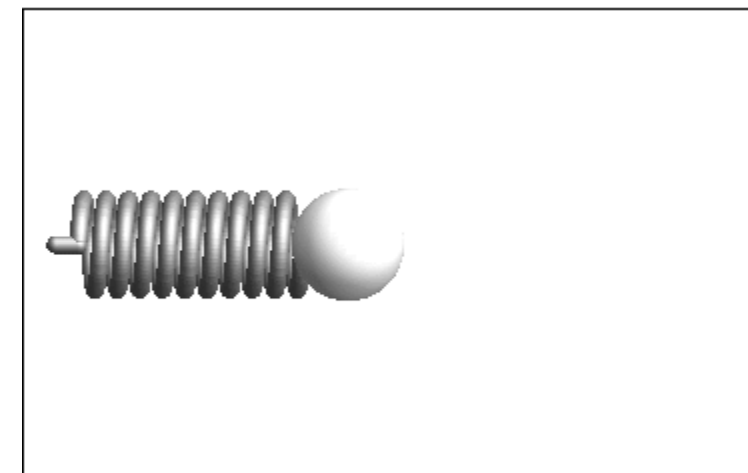
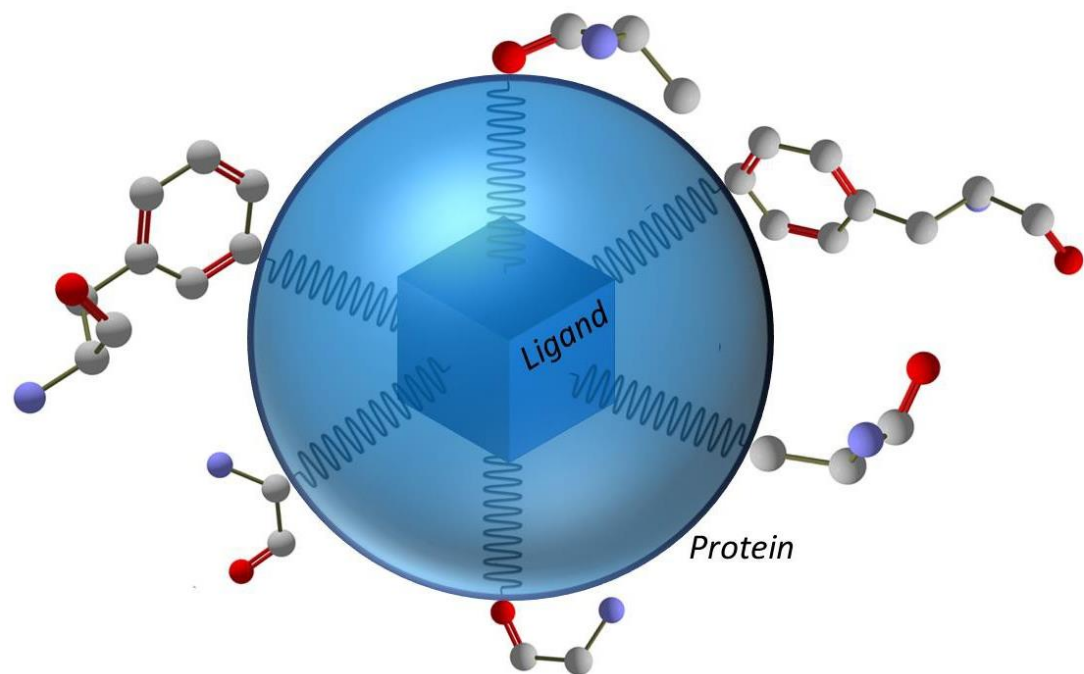




Taba-Tool to Analyze the Binding Affinity



Disponível em: <https://github.com/azevedolab>



$$PBA = \theta_0 + \sum_{k=1}^p \sum_{l=1}^q \left[\sum_{i=1}^n \sum_{j=1}^m \theta_{k,l} (d_{k,l,i,j} - d_{0,k,l})^2 \right]$$

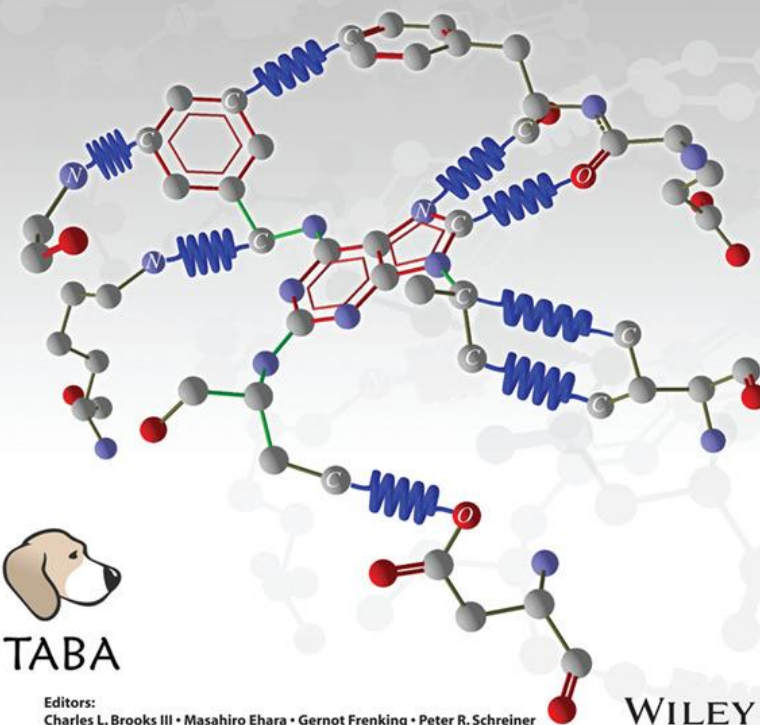


Scoring Functions	ρ	p-value1	R ²	p-value2
Free Energy ^a	-0.133	0.7324	0.204	0.2227
Final Intermolecular Energy ^a	0.133	0.7324	0.204	0.2228
vdW+Hbond+desolv Energy ^a	0.133	0.7324	0.204	0.2228
Electrostatic Energy ^a	0.533	0.1392	0.376	0.0789
Final Total Internal Energy ^a	-0.133	0.7324	0.089	0.4365
Torsional Free Energy ^a	0.068	0.8630	0.000	0.9792
Plants Score ^b	0.183	0.6368	0.001	0.9348
MolDock Score ^b	0.217	0.5755	0.010	0.7950
Rerank Score ^b	0.333	0.3807	0.007	0.8336
Interaction Score ^b	0.367	0.3317	0.013	0.7698
Protein Score ^b	0.367	0.3317	0.025	0.6839
Water Score ^b	-0.569	0.1098	0.395	0.0699
Internal Score ^b	0.033	0.9322	0.001	0.9369
Electrostatic Score ^b	0.548	0.1269	0.204	0.2218
Electrostatic Long Score ^b	-0.548	0.1269	0.204	0.2218
H-Bond Score ^b	0.650	0.0581	0.512	0.0301
Ligand Efficiency 1 Score ^b	0.150	0.7001	0.024	0.6935
Ligand Efficiency 3 Score ^b	0.283	0.4600	0.023	0.6968
Affinity Score ^c	-0.067	0.8647	0.117	0.3669
Gauss1 Score ^c	-0.367	0.3317	0.120	0.3603
Gauss2 Score ^c	-0.283	0.4600	0.018	0.7297
Repulsion Score ^c	-0.700	0.0358	0.240	0.1804
Hydrophobic Score ^c	0.100	0.7980	0.002	0.9157
Hydrogen Score ^c	-0.583	0.0992	0.340	0.0993
Taba (3 variables, d ≤ 4.5 Å)	0.783	0.01252	0.794	0.0107

Predictive performance of scoring functions (test set). ^aAutoDock 4, ^bMolegro Virtual Docker (MVD), ^cAutoDock Vi respectively.



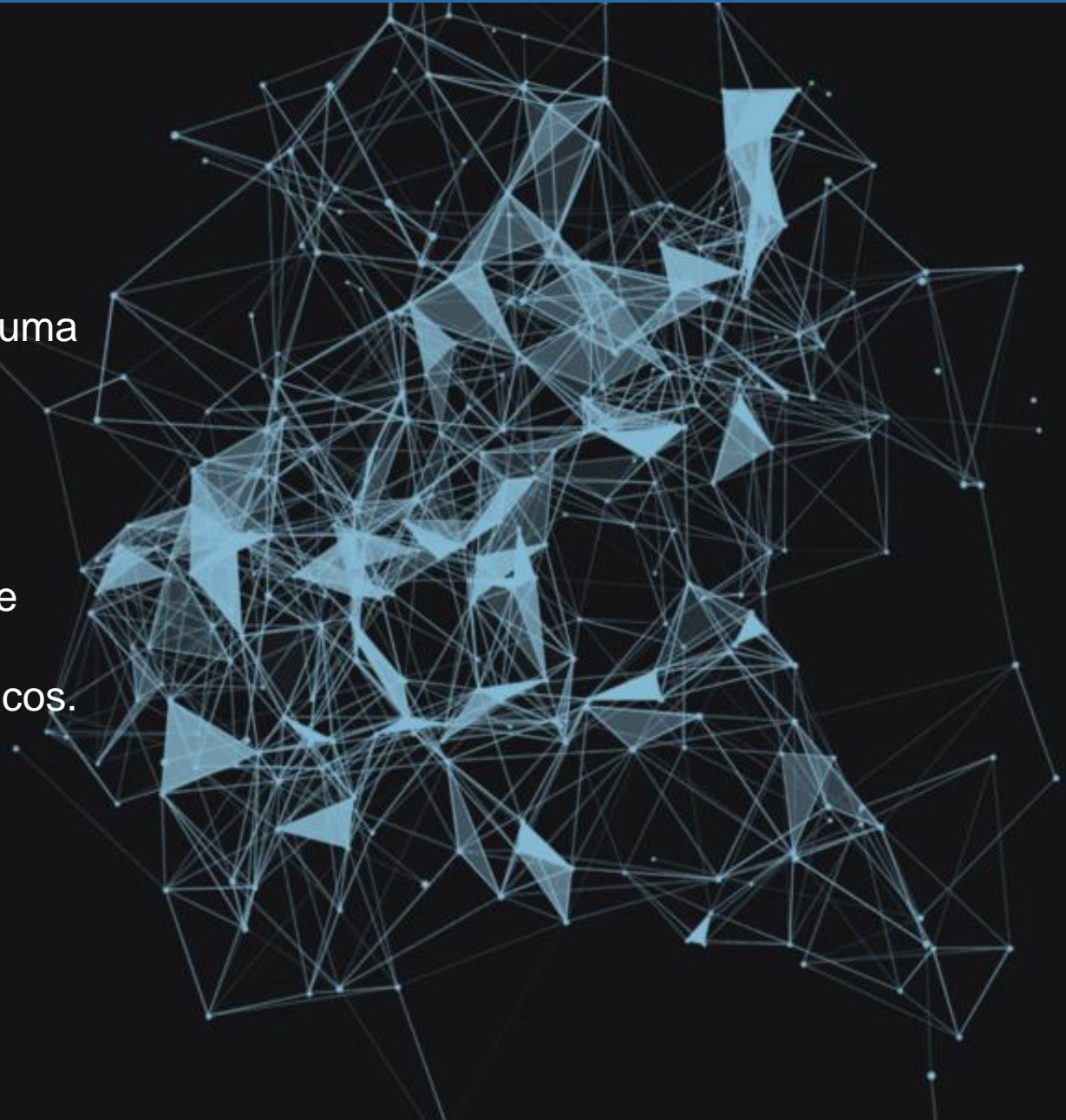
www.c-chem.org


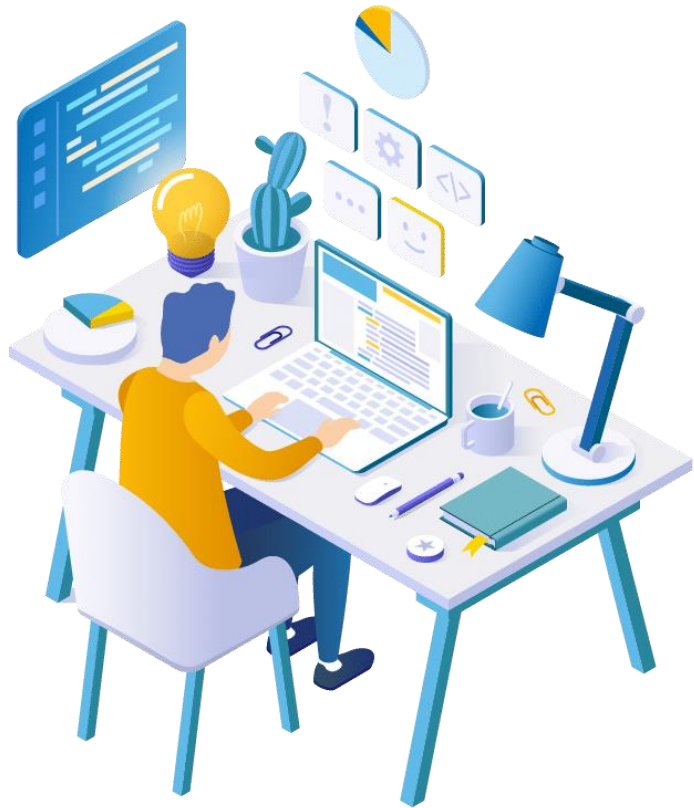




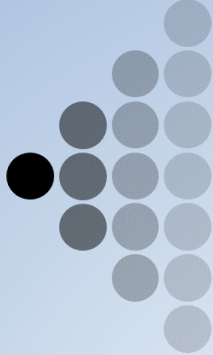
Conclusão

- Taba permite a exploração do Espaço de Funções Escores com eficiência;
- Podemos usar o programa Taba para gerar uma base de dados de funções escore (SF database);
- Acreditamos que o conceito de Espaço de Funções Escores pode ser uma mudança de paradigma no uso de abordagens computacionais para a descoberta de fármacos.

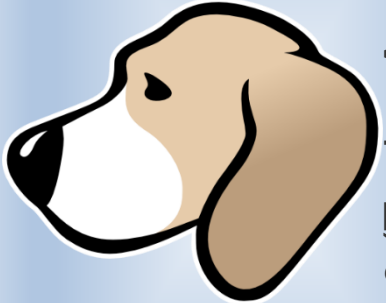




SAnDReS 2.0 (Statistical Analysis of Docking Results and Scoring functions)



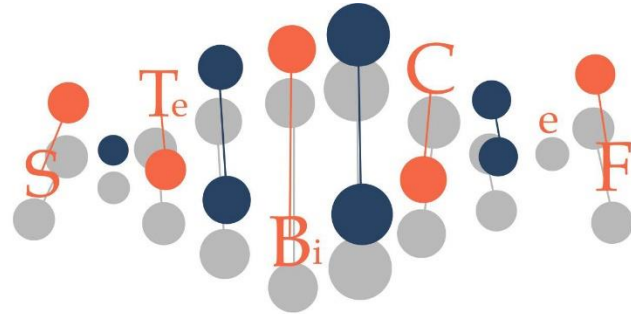
SFSXplorer (Scoring Function Space Explorer)



Taba 2.0 (Tool to Analyze the Binding Affinity)

Funding

The Brazilian National Council for Scientific and Technological Development (CNPq) (Process 306298/2022-8) supports this research project.



DIPARTIMENTO DI SCIENZE E TECNOLOGIE
BIOLOGICHE CHIMICHE E FARMACEUTICHE (STEBICEF)



I N F I Q C



UNIVERSITÉ
DE REIMS
CHAMPAGNE-ARDENNE





[Prof. Dra. Fernanda Canduri](#)



[Dra. Olga Tarasova](#)



[Profa. Dra. Stéphanie Baud](#)



[Prof. Dr. Enrique González-Vergara](#)



[Prof. Dr. Marco Tutone](#)



[Prof. Dr. I. Camps](#)



Universidad Nacional
de Córdoba

[Prof. Dr. Marcos A. Villarreal](#)



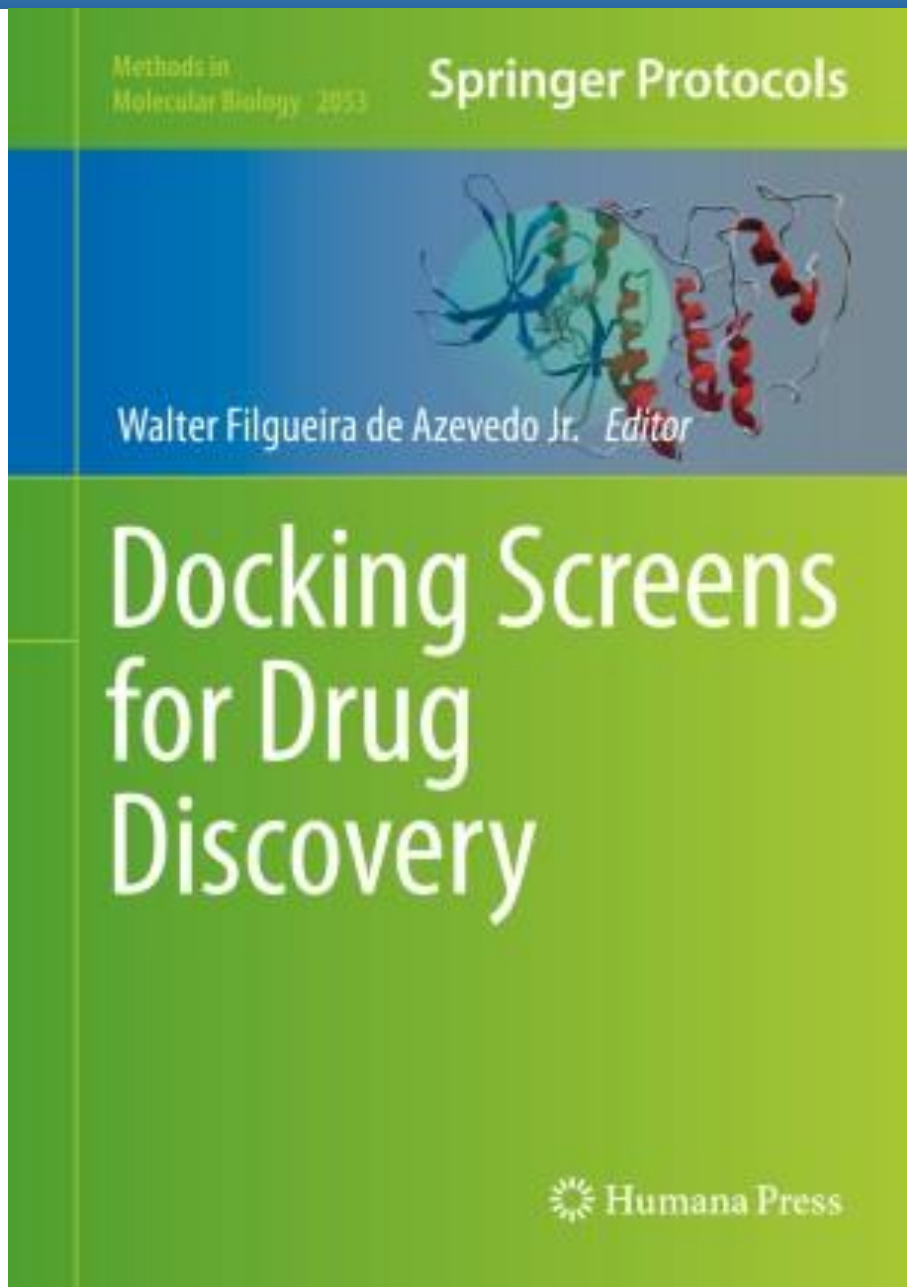
[Prof. Dr. Nelson José Freitas da Silveira](#)



[Prof. Dr. Rodrigo Quiroga](#)



[Prof. Dr. Vladimir V. Poroikov](#)



Book | © 2019

Docking Screens for Drug Discovery

[Home](#) > [Book](#)**Editors:** [Walter Filgueira de Azevedo Jr.](#)

Includes cutting-edge techniques

Provides step-by-step detail essential for reproducible results

Contains key implementation advice from the experts

Part of the book series: [Methods in Molecular Biology](#) (MIMB, volume 2053)**39k** [Accesses](#) | **339** [Citations](#) | **6** [Altmetric](#)

Source: Springer on September 12, 2023



Ph.D Students



Maurício Boff de Ávila



Silvana da Cunha Russo



Gabriela Bitencourt-Ferreira

Master Student



Camila Rizzotto

Specialization Student



Amauri Duarte da Silva

Undergraduate Student

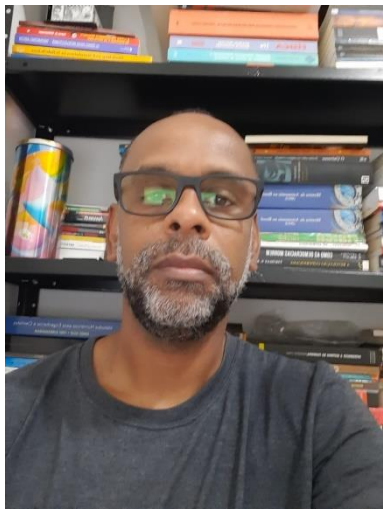


Martina Veit Acost

Thank
you!



Thank
you!



[Prof. Dr. Artur Justiniano Roberto Jr](#)



[Prof. Dr. I. Camps](#)



[Prof. Dr. Nelson José Freitas da Silveira](#)

Thank
you!



Thank
you!



EACH



Escola de Artes, Ciências e Humanidades
Universidade de São Paulo



CNPq

Capes

PPG-Física-Unifal-MG

Departamento de Física-ICEX-Unifal-MG



Thank
you!

[PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA](#)

Universidades Federais de Alfenas e São João del-Rei



Bitencourt-Ferreira G, de Azevedo WF Jr. How Docking Programs Work. *Methods Mol Biol.* 2019; 2053: 35–50.

Bitencourt-Ferreira G, de Azevedo WF Jr. Molecular Dynamics Simulations with NAMD2. *Methods Mol Biol.* 2019; 2053: 109–124.

Bitencourt-Ferreira G, de Azevedo WF Jr. Homology Modeling of Protein Targets with MODELLER. *Methods Mol Biol.* 2019; 2053: 231–249.

Compiani M, Capriotti E. Computational and theoretical methods for protein folding. *Biochemistry.* 2013;52(48):8601–8624.

Da Silva AD, Bitencourt-Ferreira G, de Azevedo WF Jr. Taba: A Tool to Analyze the Binding Affinity. *J Comput Chem.* 2020; 41(1): 69–73.

De Azevedo WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH. Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human cdk2 complexed with roscovitine. *Eur J Biochem.* 1997 Jan 15;243(1-2):518 – 526.

De Azevedo WF Jr. Molecular dynamics simulations of protein targets identified in *Mycobacterium tuberculosis*. *Curr Med Chem.* 2011; 18(9):1353–1366.

De Azevedo Junior WF, Bitencourt-Ferreira G, Godoy JR, Adriano HMA, Dos Santos Bezerra WA, Dos Santos Soares AM. Protein-Ligand Docking Simulations with AutoDock4 Focused on the Main Protease of SARS-CoV-2. *Curr Med Chem.* 2021; 28(37): 7614–7633.

Douangamath A, Fearon D, Gehrtz P, Krojer T, Lukacik P, Owen CD, Resnick E, Strain-Damerell C, Aimon A, Ábrányi-Balogh P, Brandão-Neto J, Carbery A, Davison G, Dias A, Downes TD, Dunnett L, Fairhead M, Firth JD, Jones SP, Keeley A, Keserü GM, Klein HF, Martin MP, Noble MEM, O'Brien P, Powell A, Reddi RN, Skyner R, Snee M, Waring MJ, Wild C, London N, von Delft F, Walsh MA. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat Commun.* 2020 Oct 7;11(1):5047.

Fourches D, Ash J. 4D-quantitative structure-activity relationship modeling: making a comeback. *Expert Opin Drug Discov.* 2019; 14(12): 1227–1235.

Heberlé G, de Azevedo WF Jr. Bio-inspired algorithms applied to molecular docking simulations. *Curr Med Chem.* 2011;18(9):1339 – 1352.

Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33-8,27–28.



Markowitz F. All biology is computational biology. PLoS Biol. 2017;15(3):e2002050.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009; 30(16): 2785–2791.

Nelson P. Physical Models of Living Systems. 2nd Ed. Chiliagon Science (Verlag) 978-1-7375402-4-3 (ISBN), 2022.

Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781–1802.

Thomsen R, Christensen MH. MolDock: a new technique for high-accuracy molecular docking. J Med Chem. 2006; 49(11): 3315–3321.

Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 1995 Feb;8(2):127–134.

Zhao J, Cao Y, Zhang L. Exploring the computational methods for protein-ligand binding site prediction. Comput Struct Biotechnol J. 2020;18:417-426.

*Que a luz da ciência acabe com
as trevas do negacionismo*

Aprendizado de Máquina para o Estudo de Interações Intermoleculares



Dr. Walter F. de Azevedo, Jr.
azevedolab.net

walter@azevedolab.net

Departamento de Física
Instituto de Ciências Exatas
Universidade Federal de Alfenas-MG
Brasil

