Starting with a zoom out

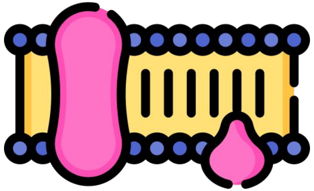
Antoine Taly
Laboratoire de Biochimie Théorique,
Institut de Biologie Physico Chimique

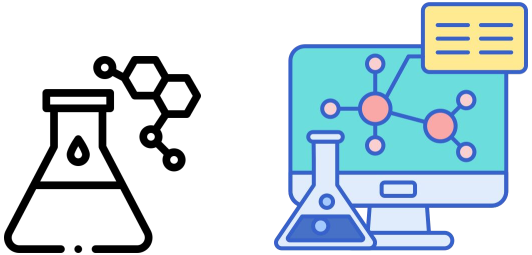
Where I am talking from

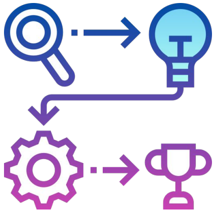
- Membrane Proteins
    - Signal transduction/Allostery
    - Conformational changes
    - Interactions with ligands

- Link experimental /modeling

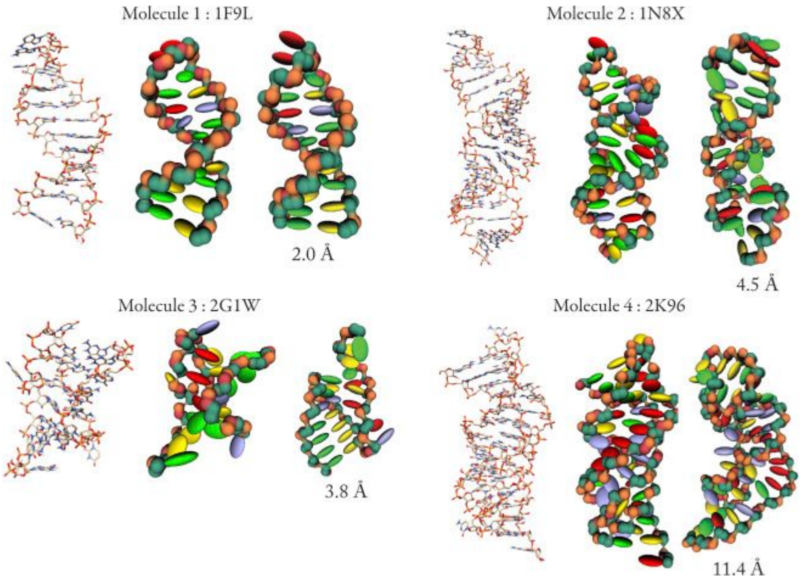
- Methods development
  →Drug Design

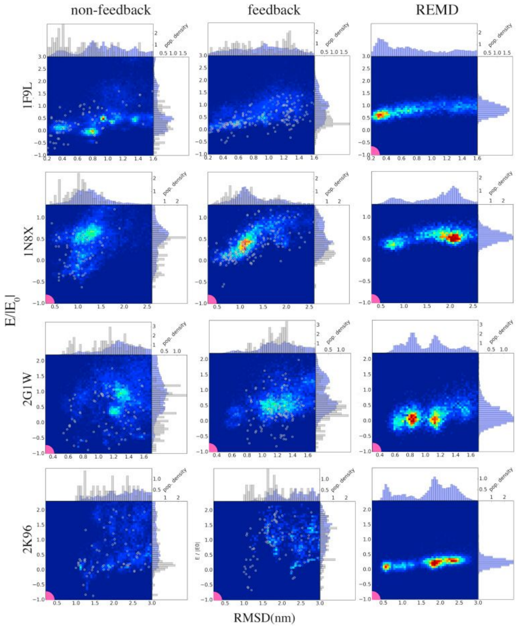






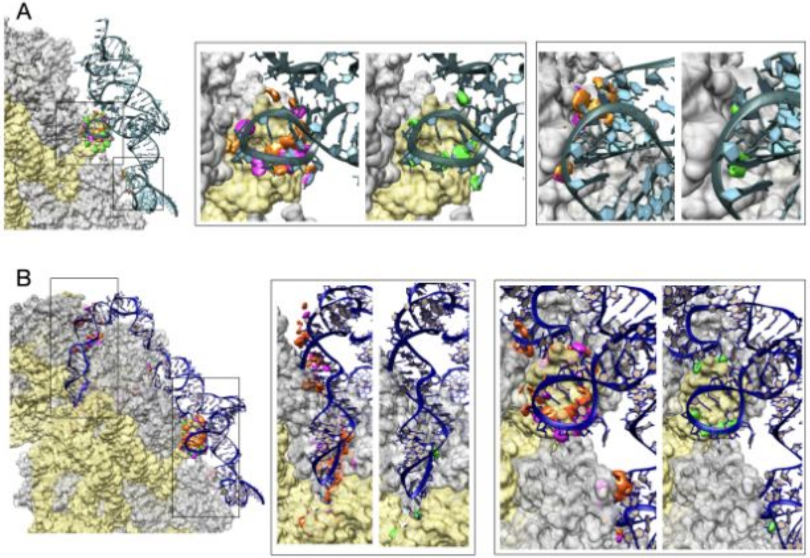
Citizen science





Mazzanti, L., et al. (2017). What can human-guided simulations bring to RNA folding?. *Biophysical journal*, *113*(2), 302-312.

protein/RNA interactions



Barquero Morera, D., Mattiotti, G., Kocev, A., Rousselot, A., Rouaud, L., Santuz, H., Baaden, M., Taly, A., & Pasquali, S. (2025). Statistical Molecular Interaction Fields: A Fast and Informative Tool for Characterizing RNA and Protein Binding Pockets. *JCTC*, 2025.

Zooming out

- Model and data (from a biochemist perspective)
- RNA beyond the central dogma/WC

What is a model?

- Description of a set of data
- Description of the properties of molecules based on a theoretical framework (quantum mechanics, classical mechanics, etc...)
- “There’s No Data Like More Data”

A model to describe a set of data

- A schematic representation can be useful to summarize a complicated set of data
- Meant to be pedagogical not an accurate description of the system
- Human made
- It can’t be used further to make predictions

A model to describe the properties of molecules

- Based on a **physically realistic** description of the properties of molecules
- **Operated by computers** using softwares that ground on those theoretical descriptions of molecules (molecular mechanics, quantum mechanics, etc...)
- A level of description has to be **chosen as a function of the question to be answered** (structure? Flexibility? Time scale? Chemical reaction? Size of molecule)
- **Allow predictions**: guide experiments, reduce the size of experiments (drug design, protein engineering), compare experiments

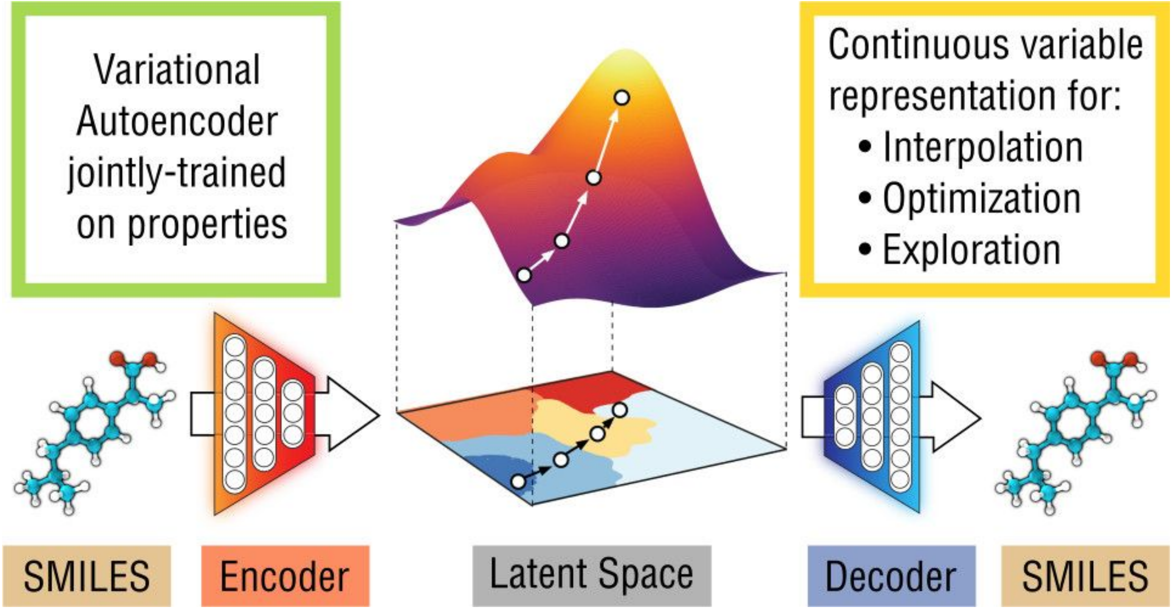
What makes something data?

- Emily Bender

  https://medium.com/@emilymenonbender/what-makes-something-data-f6d9f498f312

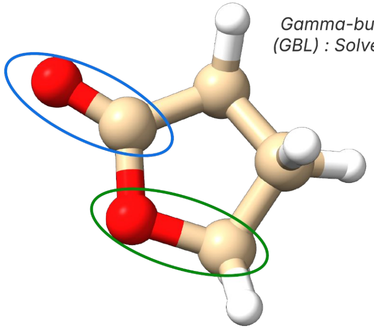
  - Andrews, Mel. 2025. The Immortal Science of ML: Machine Learning and the Theory-Free Ideal. *Erkenn.*
  - Li, Xiaochang. “There’s no data like more data” automatic speech recognition and the making of algorithmic culture. *Osiris* 38, no. 1 (2023): 165–182.
  - Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ‘21). Association for Computing Machinery*, New York, NY, USA, 610–623.

Generating molecules with a SMILES-based Variational Auto Encoder (VAE)



Anusha et al. Molecule Generation of Drugs Using VAE. In *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)* (Vol. 112, p. 170).

Representing molecules with SMILES 1



Gamma-butyrolactone (GBL) : Solvent industriel

🟦 *Branches*

🟩 *Cycle*

SMILES

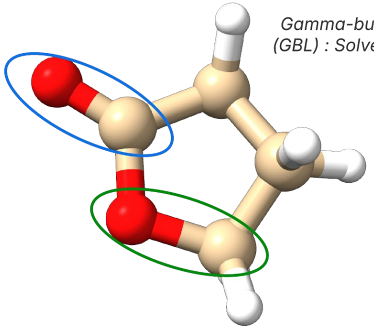
C1 C C C (= O) O1
1 2 3 4 5 6 7 8 9 10 11

Encodage 1 parmi n →

x

	C	1	(	=	O	)
1	■					
2		■				
3	■					
4	■					
5	■					
6			■			
7				■		
8					■	
9						■
10					■	
11		■				

Representing molecules with SMILES 1



Gamma-butyrolactone (GBL) : Solvent industriel

🟦 *Branches*

🟩 *Cycle*

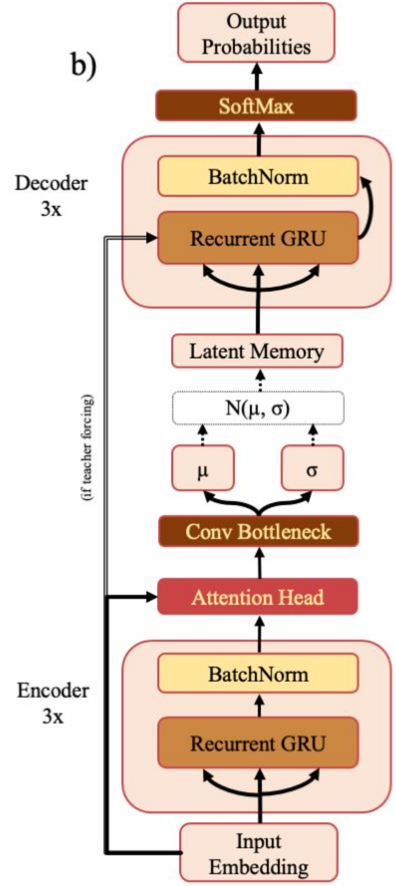
SMILES

C 1 C C C **(= O) O 1**
1 2 3 4 5 6 7 8 9 10 11

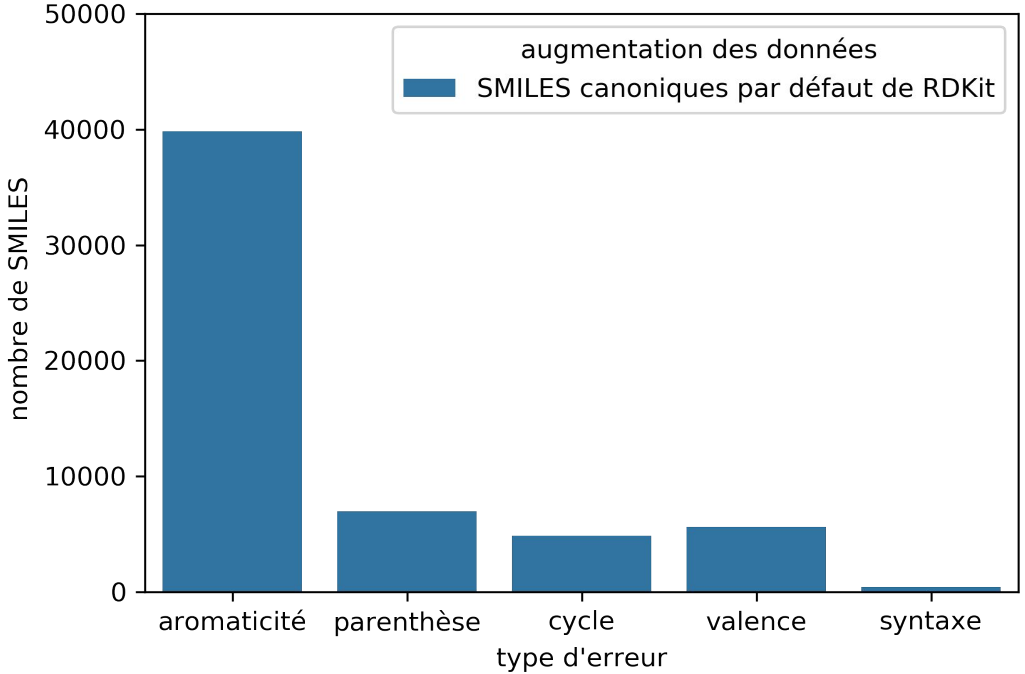
Encodage 1 parmi n

x

	C	1	(	=	O	)
1	■					
2		■				
3	■					
4	■					
5	■					
6			■			
7				■		
8					■	
9						■
10					■	
11		■				

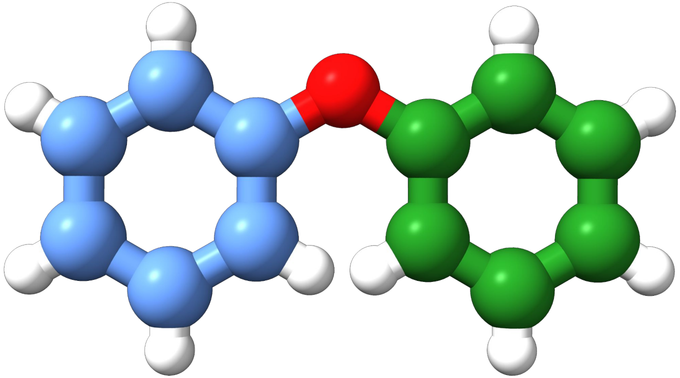


Representing molecules with SMILES 2



⇒ There are often errors in the generated molecules

Representing molecules with SMILES 3

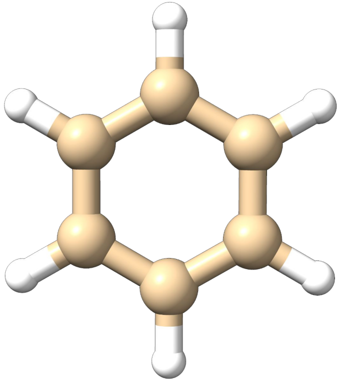


c1(Oc2ccccc2)ccccc1

c1ccccc1Oc2ccccc2

⇒ we can choose a simpler solution

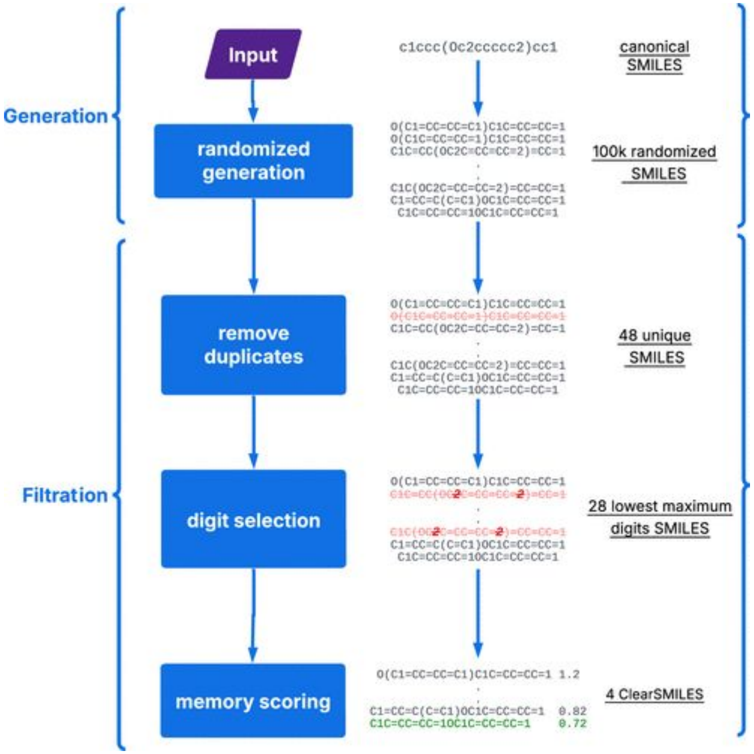
Representing molecules with SMILES 3

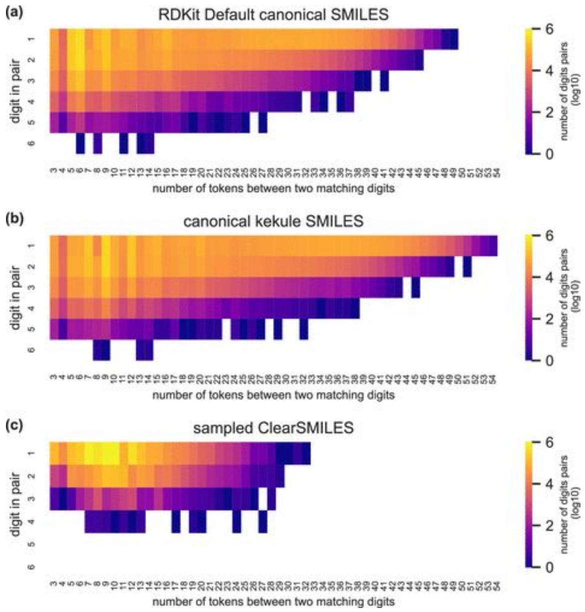


- Compact : c1ccccc1
- **Kekule** : C1=CC=CC=C1
- Aromatiques links C1:C:C:C:C:C1

⇒ we can vary the aromaticity representation

ClearSMILES





Reboul, E., Wefers, Z., Prabakaran, H., Waldispuhl, J., & Taly, A. (2025). Improving the reliability of molecular string representations for generative chemistry. *Journal of Chemical Information and Modeling*, *65*(19), 10221-10238.

Can a model learn from the data alone?

Physics/Chemistry does not appear magically

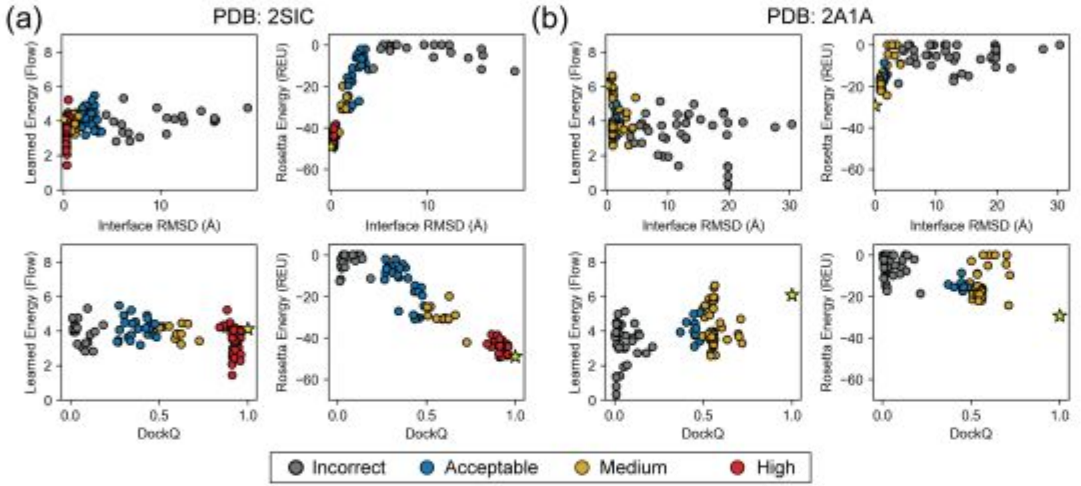
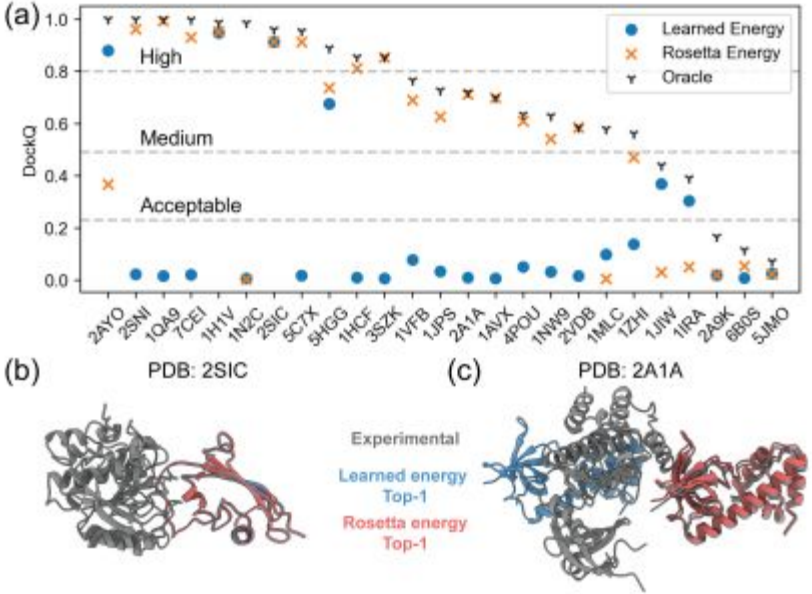


FIG. 5. Learned energy (recovered from integrating over flow trajectories) and Rosetta energy of 120 generated docking poses plotted against interface RMSD (top) or DockQ (bottom) for (a) PDB ID: 2SIC and (b) PDB ID: 2A1A. Individual points are colored by the quality of the corresponding docking pose based on the CAPRI classification [43]: incorrect, gray; acceptable, blue; medium quality, gold; high quality, red. Energies computed for the ground truth structures are shown as yellow stars.

Can We Extract Physics-like Energies from Generative Protein Diffusion Models? Sudeep Sarma, Harrison Truscott, Da Xu, Kendall Reid, Lee-Shin Chu, Jacky Chen, Jeffrey J. Gray bioRxiv 2025.11.28.690021; doi: https://doi.org/10.1101/2025.11.28.690021

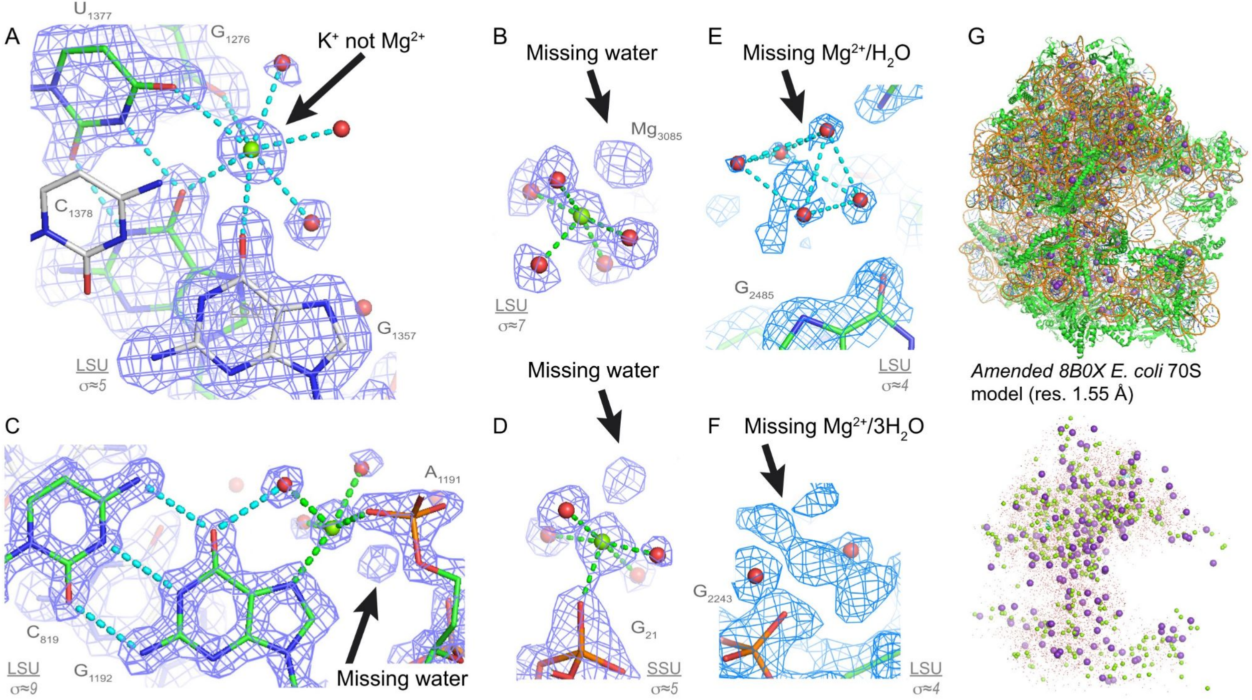
Physics/Chemistry does not appear magically



Can We Extract Physics-like Energies from Generative Protein Diffusion Models? Sudeep Sarma, Harrison Truscott, Da Xu, Kendall Reid, Lee-Shin Chu, Jacky Chen, Jeffrey J. Gray bioRxiv 2025.11.28.690021; doi: https://doi.org/10.1101/2025.11.28.690021

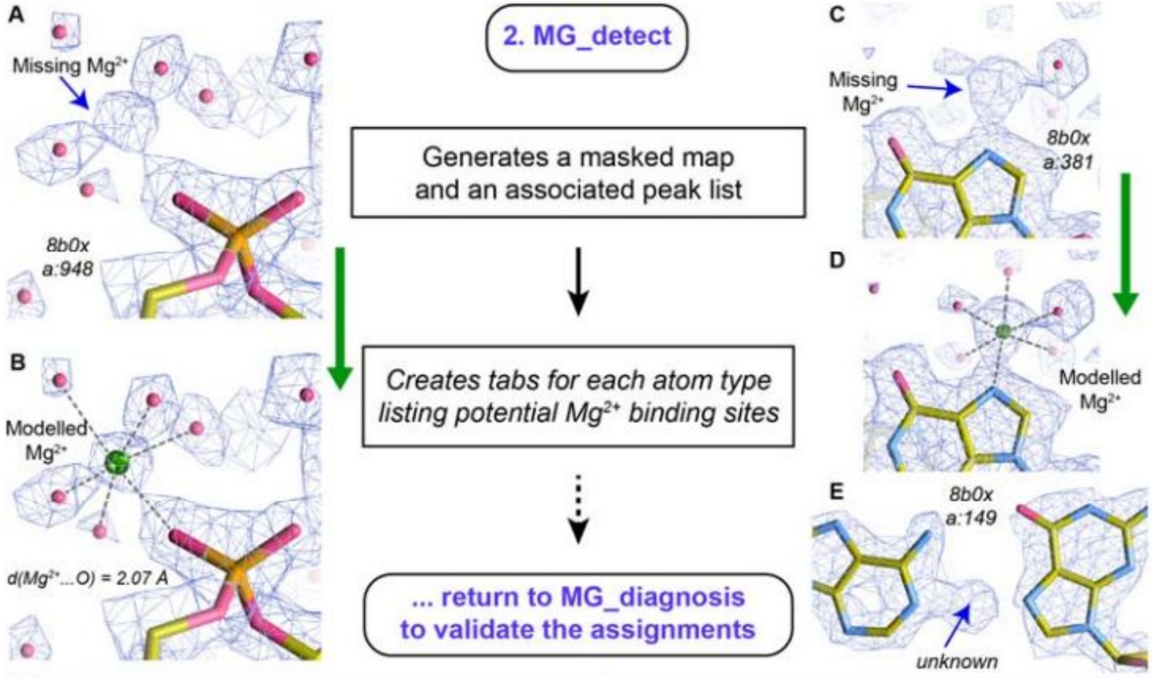
Errors in data associated with RNA

How well are ions seen in structures ?



Filip Leonarski, Anja Henning-Knechtel, Serdal Kirmizialtin, Eric Ennifar, Pascal Auffinger, Principles of ion binding to RNA inferred from the analysis of a 1.55 Å resolution bacterial ribosome structure – Part I: Mg2+, *Nucleic Acids Research*, Volume 53, Issue 1, 13 January 2025, gkae1148, https://doi.org/10.1093/nar/gkae1148

Missing ions



2. MG_detect

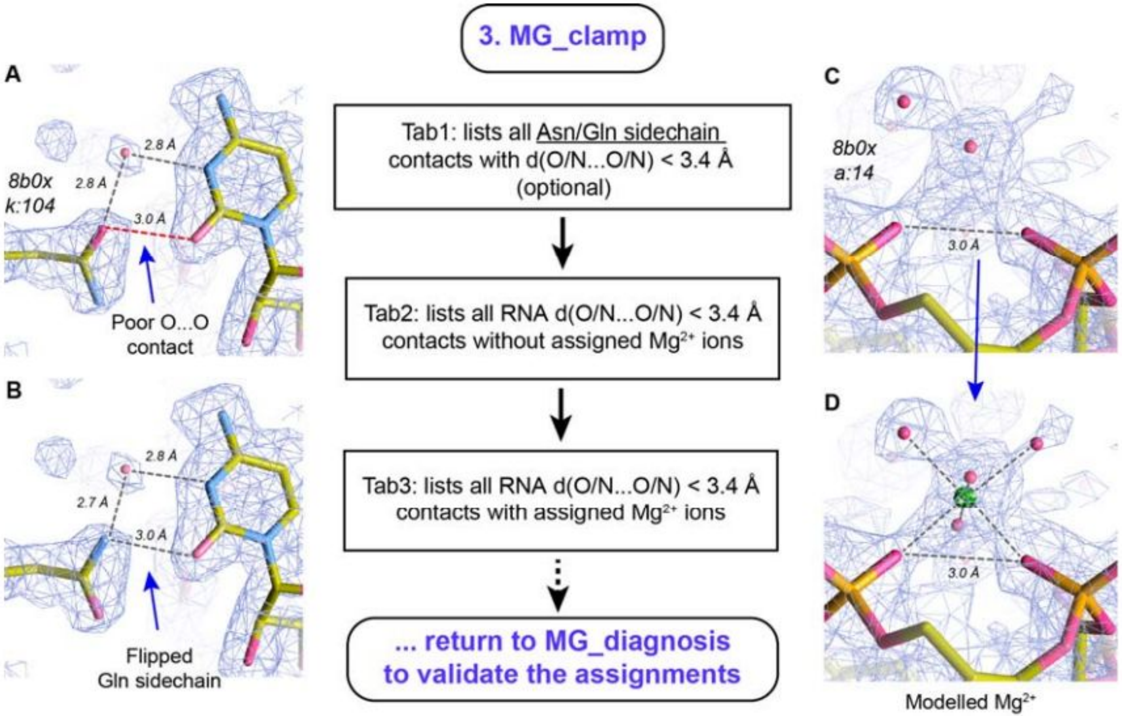
Generates a masked map and an associated peak list

Creates tabs for each atom type listing potential Mg^{2+} binding sites

... return to MG_diagnosis to validate the assignments

Naleem, N., Henning-Knechtel, A., Kirmizialtin, S., & Auffinger, P. (2025). Cat_Wiz: A stereochemistry-guided toolkit for locating, diagnosing and annotating Mg2+ ions in RNA structures. *bioRxiv*, 2025-10.

Wrong bonding atom



3. MG_clamp

Tab1: lists all Asn/Gln sidechain contacts with d(O/N...O/N) < 3.4 Å (optional)

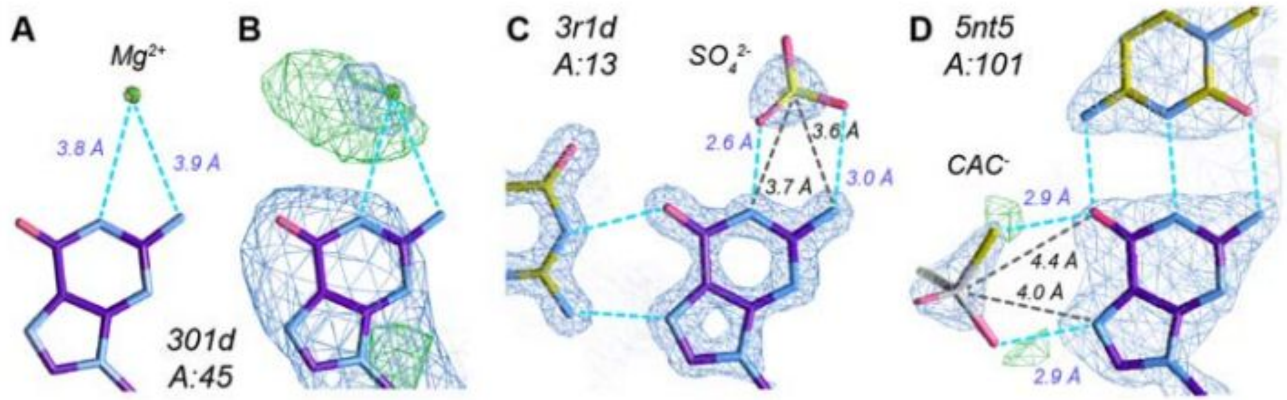
Tab2: lists all RNA d(O/N...O/N) < 3.4 Å contacts without assigned Mg^{2+} ions

Tab3: lists all RNA d(O/N...O/N) < 3.4 Å contacts with assigned Mg^{2+} ions

... return to MG_diagnosis to validate the assignments

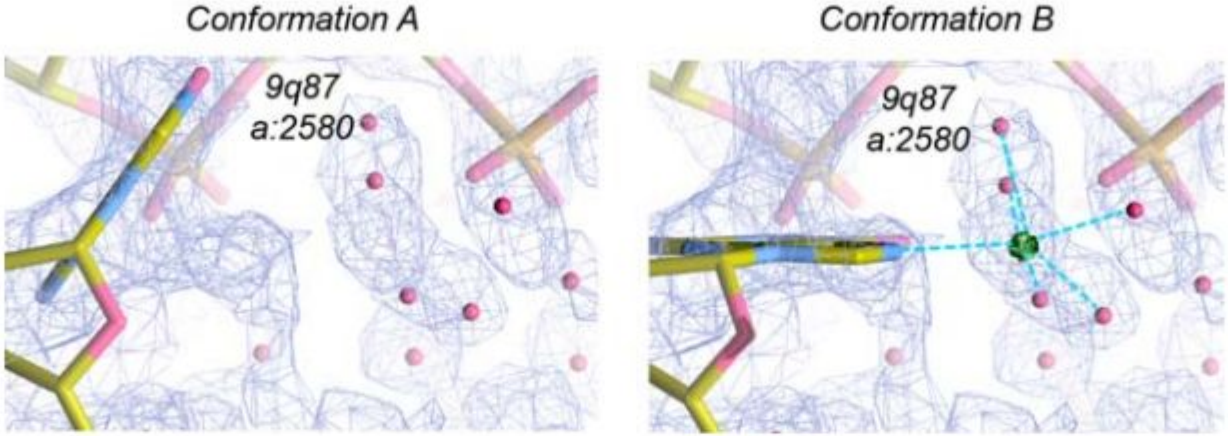
Naleem, N., Henning-Knechtel, A., Kirmizialtin, S., & Auffinger, P. (2025). Cat_Wiz: A stereochemistry-guided toolkit for locating, diagnosing and annotating Mg2+ ions in RNA structures. *bioRxiv*, 2025-10.

Wrong ion



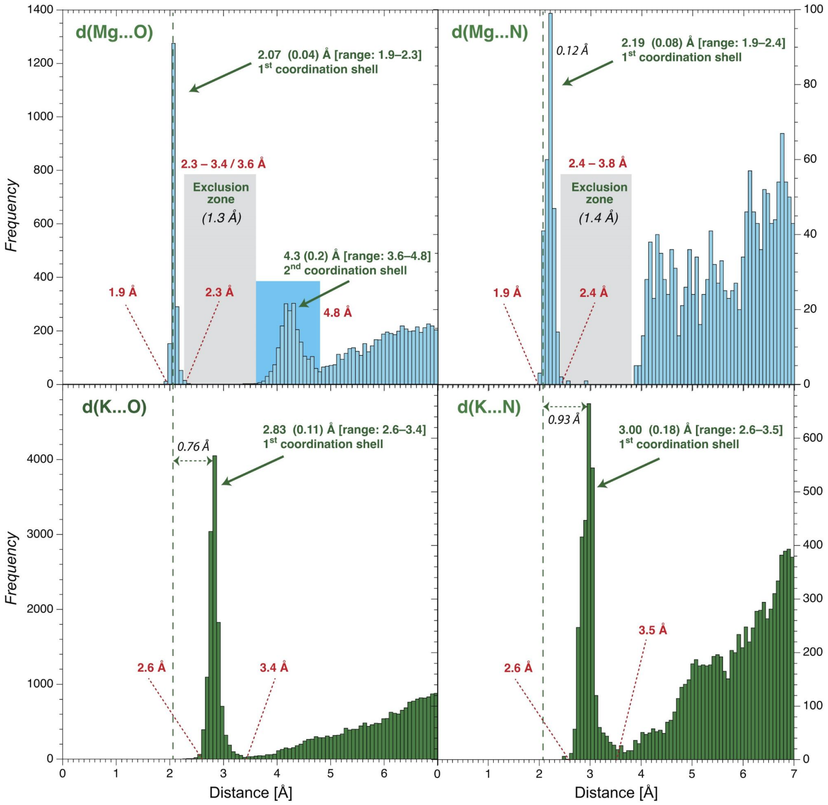
Naleem, N., Henning-Knechtel, A., Kirmizialtin, S., & Auffinger, P. (2025). Cat_Wiz: A stereochemistry-guided toolkit for locating, diagnosing and annotating Mg2+ ions in RNA structures. *bioRxiv*, 2025-10.

Alternative conformation (+ ion)



Naleem, N., Henning-Knechtel, A., Kirmizialtin, S., & Auffinger, P. (2025). Cat_Wiz: A stereochemistry-guided toolkit for locating, diagnosing and annotating Mg2+ ions in RNA structures. *bioRxiv*, 2025-10.

How well are ions seen in structures ?



⇒ Chemistry can help

Filip Leonarski, Anja Henning-Knechtel, Serdal Kirmizialtin, Eric Ennifar, Pascal Auffinger, Principles of ion binding to RNA inferred from the analysis of a 1.55 Å resolution bacterial ribosome structure – Part I: Mg2+, *Nucleic Acids Research*, Volume 53, Issue 1, 13 January 2025, gkae1148, https://doi.org/10.1093/nar/gkae1148

Beyond the central dogma/WC

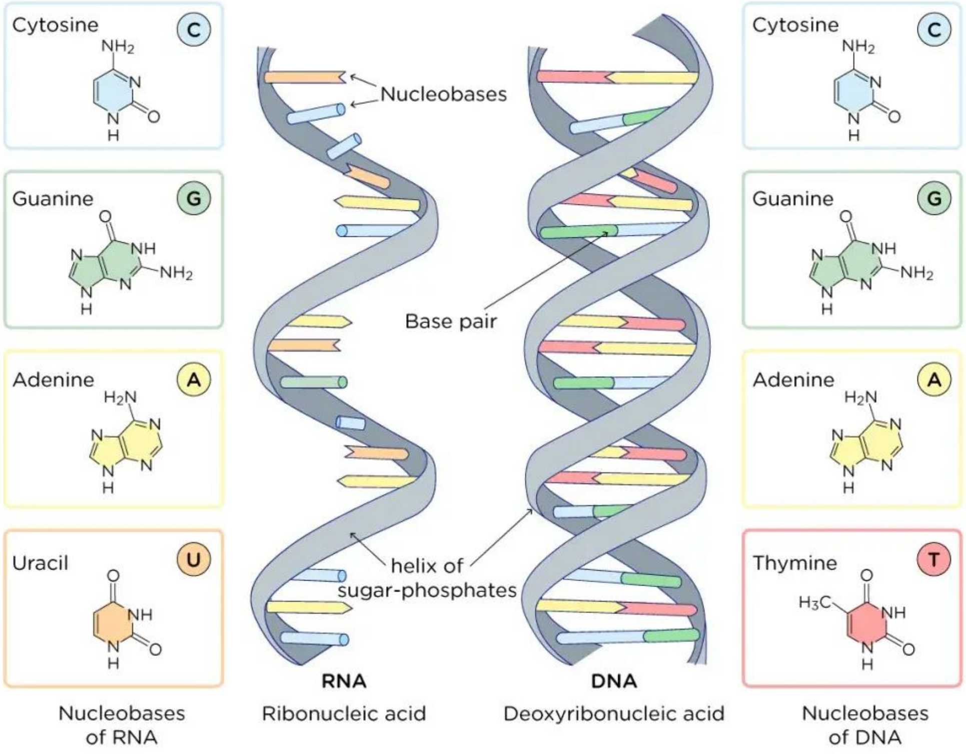
RNA beyond the central dogma

- "fonction" -->

  property/activity

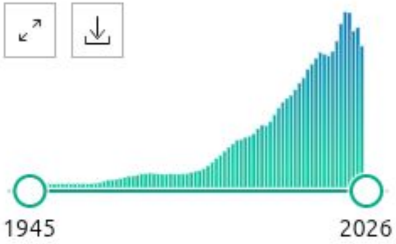
  [why--> How]

- interactions beyond WC

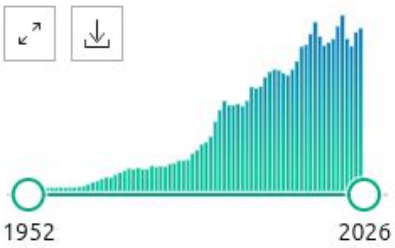


Jack Westin

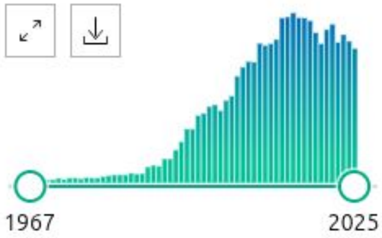
- RNA



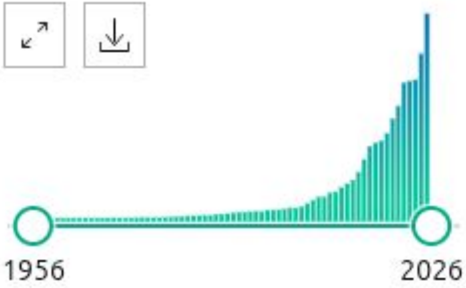
RNA + translation



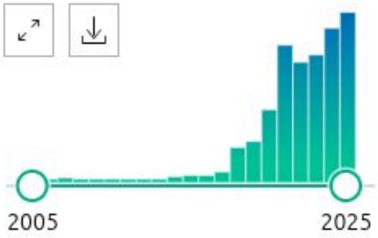
- RNA+folding



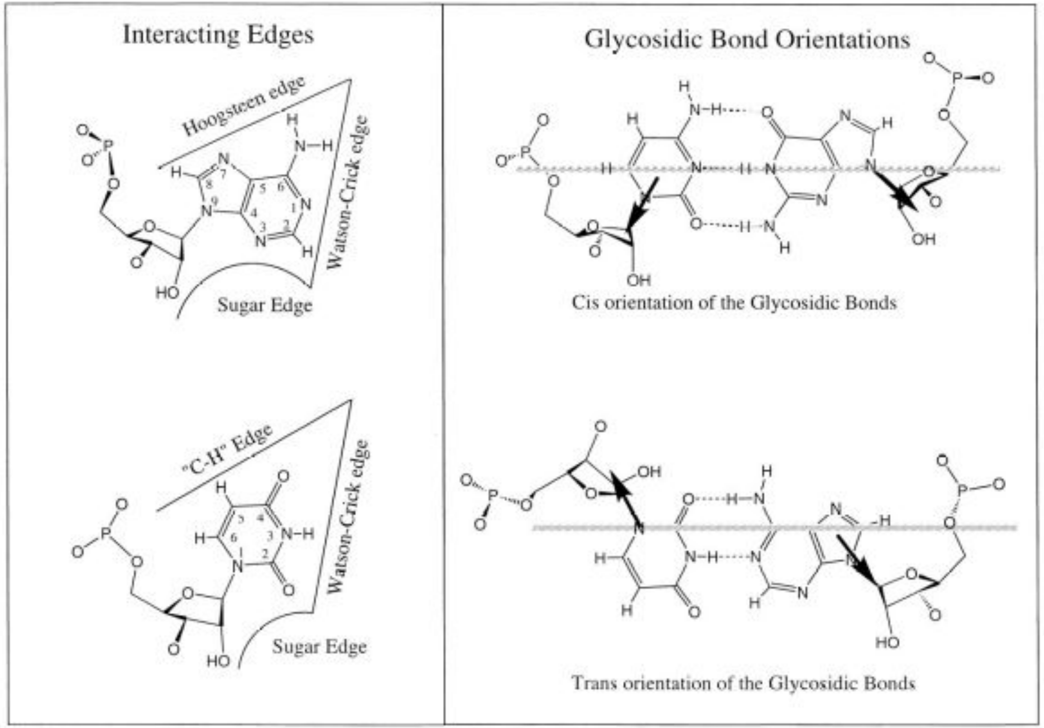
- RNA+integrative



RNA + LLPS



Beyond WC



Nucleic Acids Res, Volume 30, Issue 16, 1 August 2002, Pages 3497–3531, https://doi.org/10.1093/nar/gkf481

Baise pairings beyond WC (structure)

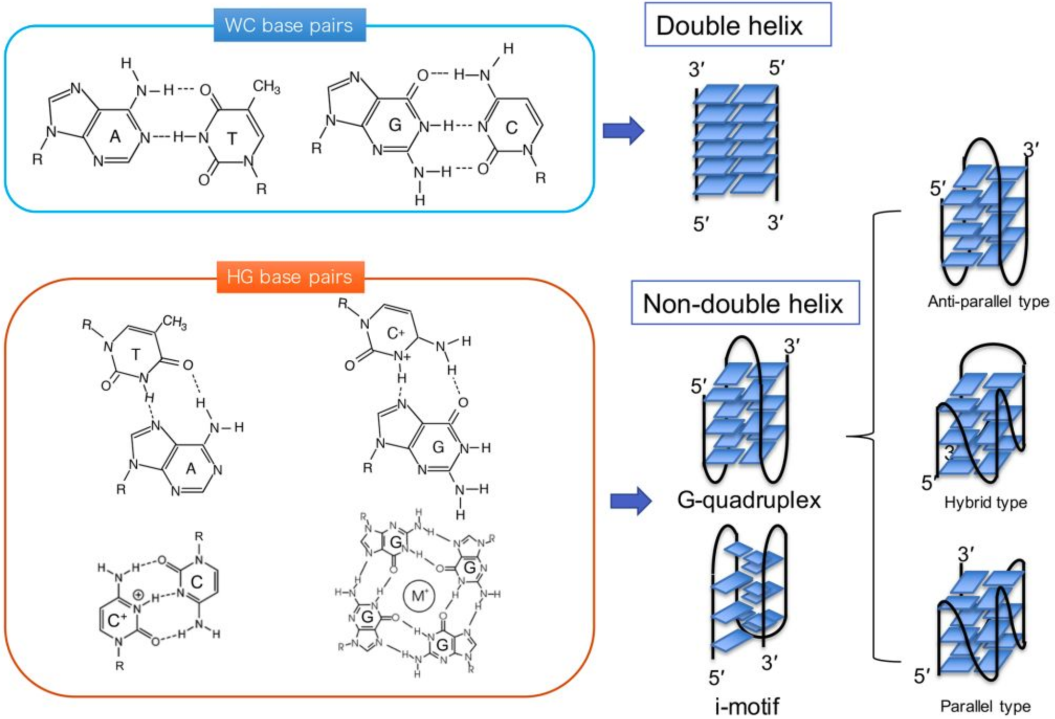


Figure 1. Various NA structures constructed via WC and HG bps.

Takahashi, S., & Sugimoto, N. (2021). Watson–Crick versus Hoogsteen base pairs: chemical strategy to encode and express genetic information in life. *Accounts of chemical research*, *54*(9), 2110-2120.

Beyond the central dogma

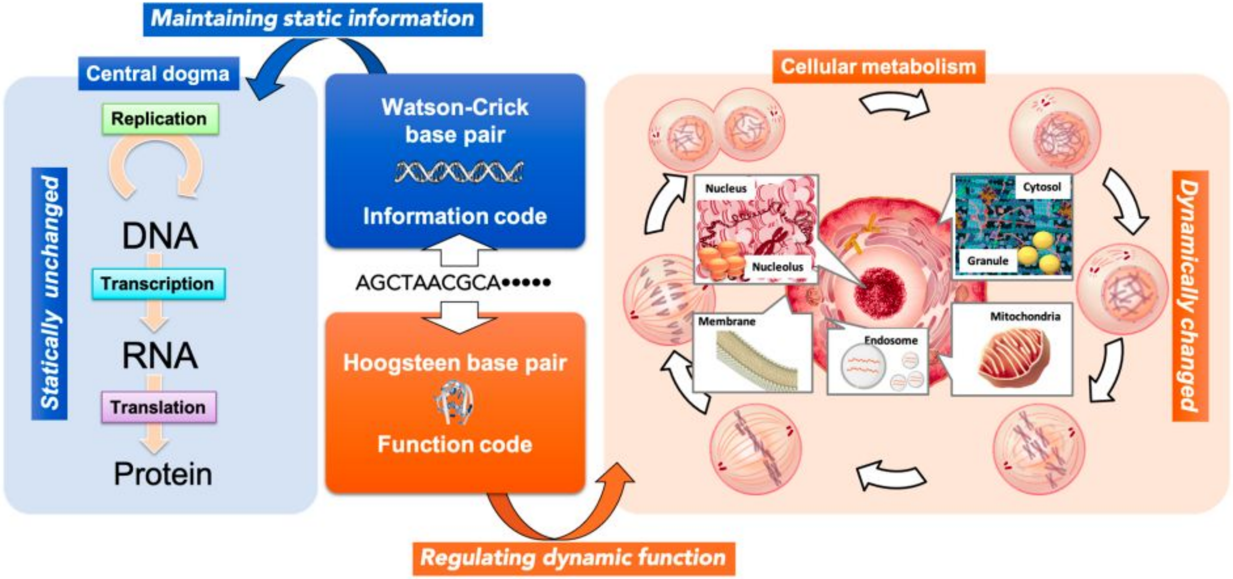
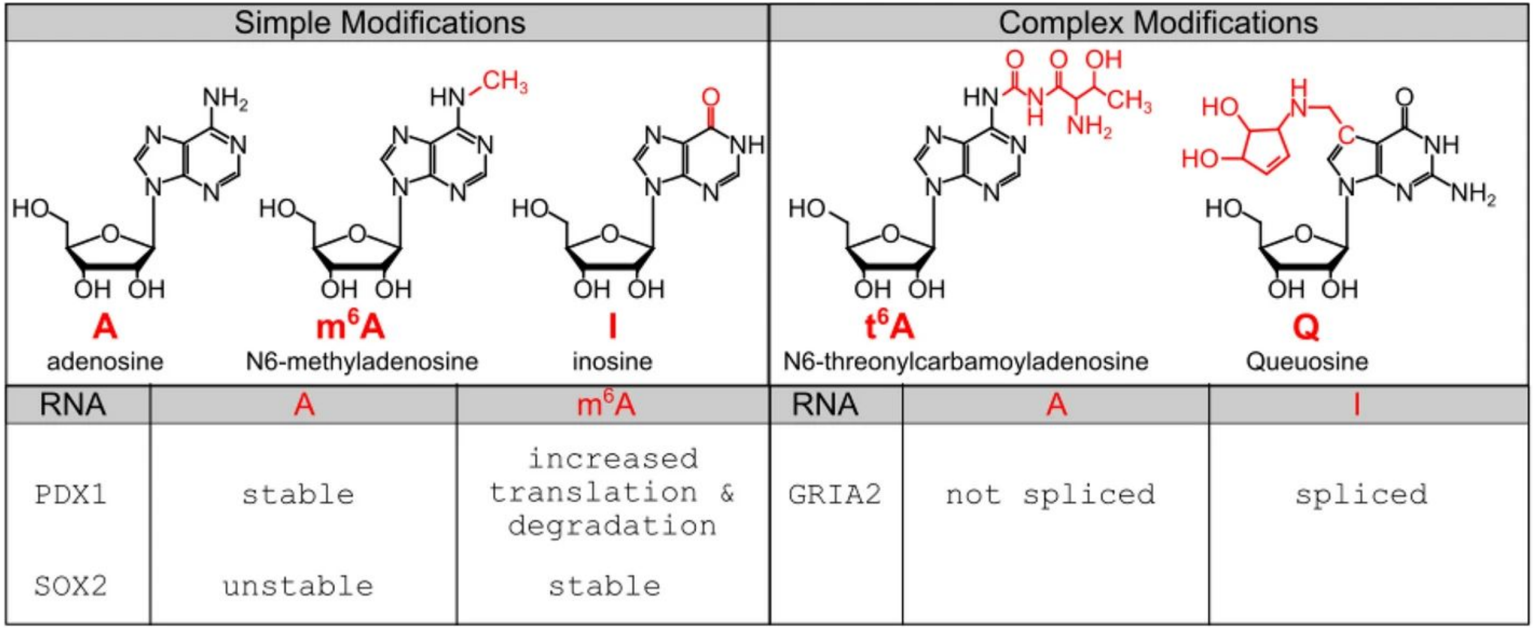


Figure 7. Role of WC bp vs HG bp in cells. WC bps encode genetic information for the central dogma of life, which should be statically unchanged, whereas HG bps express genetic function for the cellular metabolism, which is dynamically changed by the environment.

Takahashi, S., & Sugimoto, N. (2021). Watson–Crick versus Hoogsteen base pairs: chemical strategy to encode and express genetic information in life. *Accounts of chemical research*, *54*(9), 2110-2120.

Beyond AUGC



Simple Modifications			Complex Modifications		
RNA	A	m^6A	RNA	A	I
PDX1	stable	increased translation & degradation	GRIA2	not spliced	spliced
SOX2	unstable	stable			

The International Human RNome Project Consortium. Unlocking the regulatory code of RNA: launching the Human RNome Project. *Genome Biol* 26, 367 (2025). https://doi.org/10.1186/s13059-025-03824-y