LEARNING MATTER

High-throughput quantum chemistry, machine learning and atomistic materials simulations

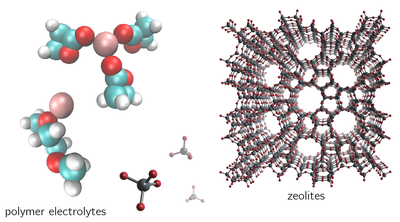
PI: Rafael Gómez-Bombarelli

Our expertise

AI-based softwares are beating humans in every task: driving cars, games, speech recognition etc. We are employing their power to improve materials discovery.

Materials

Our work emphasizes the role of molecular transformation in materials discovery. In particular, we are tackling problems in energy storage, catalysis, organic LEDs, biological sequences and quantum computing.

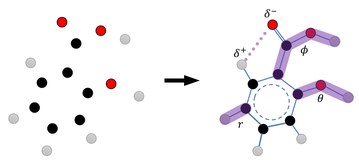


polymer electrolytes

zeolites

Science

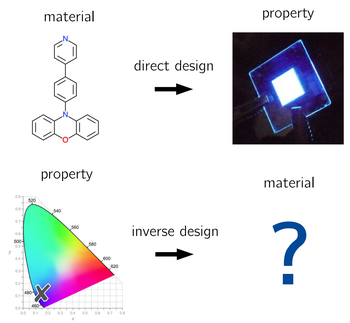
The combination of large experimental datasets and accurate theoretical simulations with statistical inference allows us to answer fundamental scientific questions like never before.



$$E = \sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} k_\phi[1 + \cos(n\phi - \gamma)] + \sum_{\text{nb}} \left(\frac{q_i q_j}{4\pi\varepsilon r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right)$$

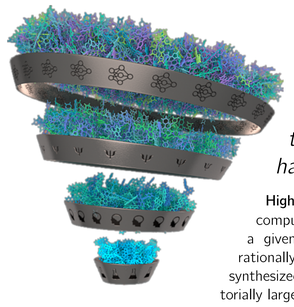
Engineering

Given a material, it is commonplace to use physics to predict its properties. We tackle the inverse problem: **given a target property, how to choose the optimal material for it?**



Our projects

High-throughput virtual screening



We use efficient exploration of the chemical space through machine-learning, first-principles calculations, and human expertise to find needles in materials haystacks.

High-throughput virtual screening combines computational investigation with experiments to optimize a given property within vast libraries of compounds. By rationally deciding which materials will be simulated or synthesized, best candidates can be found among combinatorially large options.

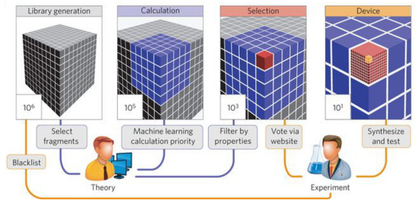
Organic Light Emitting Diodes

Search space: 1.6M molecules

Screening with deep learning and TD-DFT.

Incorporates high-throughput experimentation, molecular characterization, industrial expertise, and device fabrication and testing.

Thousands of molecules identified as promising. Candidates have shown external efficiency of up to 22%.

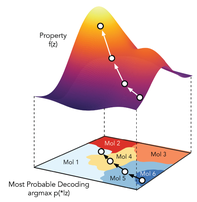


Generative models and inverse design

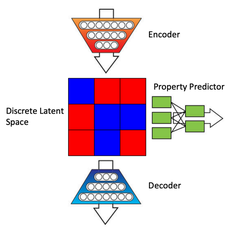
We teach statistical models how to design materials.

Variational autoencoders are widely used to perform statistical inference over a distribution of data points. In materials science, it can be used to convert molecular representations into a continuous, optimizable space.

We train autoencoders to learn the chemical similarity between molecules. Through neural network-based predictors and Gaussian processes, we allow for the inverse design of molecules based on their properties and latent representations.



Machine learning meets quantum computing



Quantum computing combined with machine learning models can use the power of quantum superposition to accelerate materials discovery.

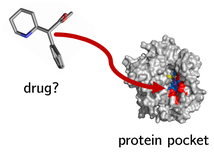
How can we use classical machine learning strategies more efficiently with quantum computers?

We are developing algorithms to discretize the latent space of variational autoencoders and make these models compatible with qubit technology.

Generation of molecular graphs

Given a protein pocket, how to create a molecule that binds to it?

We are conditioning the generation of molecular graphs to properties such as the geometry of protein pockets. By sampling molecules which fit to the pocket, we can easily create new drugs.



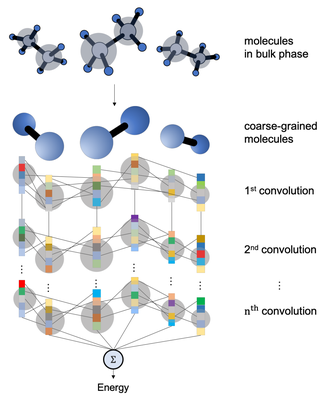
Machine learning and atomistic simulations

We blur the boundaries of materials physics, chemistry and computer science.

We generate in-house datasets based on DFT calculations for molecules.

Expensive DFT and molecular dynamics simulations are accelerated using graphs and machine learning predictors.

Many-body effects are captured by several interactions between local chemical environments, described by graph ensembles and convolutions.



Polymer electrolyte for organic batteries

Polymer electrolytes can be cheaper and environmental-friendly solutions for batteries.

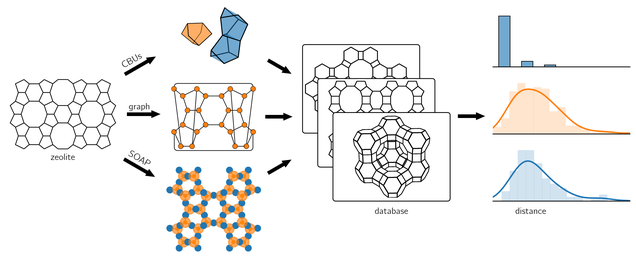
We work towards the fast computation of ion-polymer interactions, charge transport, and thermodynamic properties of the compounds.

We develop coarse-grained molecular simulations for probing longer time scales and bigger systems.

Data-driven materials discovery

We use data as a powerful tool to discover and optimize materials.

Discovery of zeolites with graph theory and DFT calculations

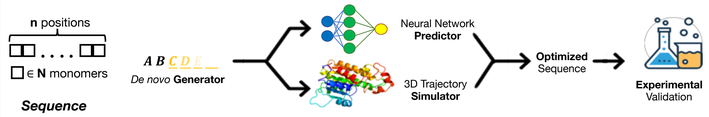


Zeolites are extremely important catalysts for chemical industry, despite being tricky to synthesize.

How can we optimize their catalytic activity and make them in the lab?

We are studying how to design zeolites tailored for certain chemical reactions based on unsupervised learning over graphs and crystal structures. Based on their parameters, we design reaction pathways to synthesize new frameworks with improved catalytic activity.

Optimization of biological sequences for improved experimental performance



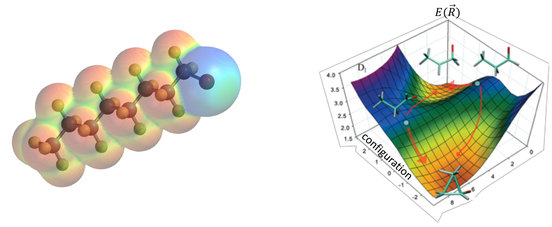
Biological sequences (DNA/RNA, peptides/proteins, sugars) are ubiquitous in their presence and applications.

How can we select the best out of the combinatorial chemical space of more than 10^{150} sequences?

We are using deep learning and Bayesian optimization techniques over topological representations of monomers. The objective is to understand sequence-property relationships, and predict novel optimized sequences which are validated *in vitro*.

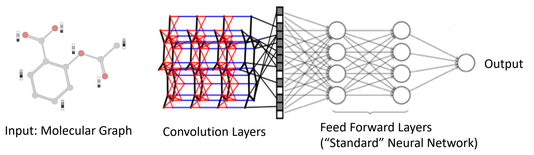
Our tools

Density functional theory and molecular dynamics



High-throughput DFT and MD calculations to predict geometries and properties of molecules and solids.

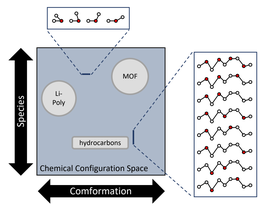
State-of-the-art statistical learning models

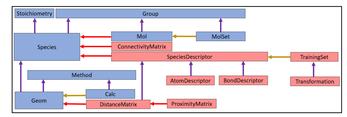


State-of-the-art models for chemistry, graphs, statistical inference, and machine learning.

In-house development of neural force-fields, generative models, representation learning and graph-theoretical approaches.

Database creation and management





Combinatorial generation of libraries of materials.

Database management for proteins, molecules and solids.

References

1. Gómez-Bombarelli, R. et al. *Nat. Mater.* **15**, 1120 (2016).
2. Gómez-Bombarelli, R. et al. *ACS Cent. Sci.* **4**, 268 (2018).
3. Schwalbe-Koda, D. and Gómez-Bombarelli, R. arXiv:1812.02685.
4. Wang, W. and Gómez-Bombarelli, R. arXiv:1812.02706.
5. Duvenaud, D. et al. *NeurIPS* 2015, 2224

Massachusetts Institute of Technology

Department of Materials Science and Engineering
gomezbombarelli.mit.edu, rafagb@mit.edu