Generative chemistry at scale: Practical lessons from Quantori and Nebius

Artem Elmuratov
Head of Life Sciences and Healthcare
Nebius

Yuriy Gankin, PhD
Chief Scientific Officer
Quantori

Denis Sapegin, PhD
Principal Cheminformatics Engineer
Quantori

Speakers

Artem Elmuratov

Head of Life Sciences and Healthcare

NEBIUS

Yuriy Gankin, PhD

Chief Scientific Officer

QUANTORI

Denis Sapegin, PhD

Principal Cheminformatics Engineer

QUANTORI

Let’s get to know each other

What will you learn?

Infrastructure strategy

Pipeline architecture

Performance insights

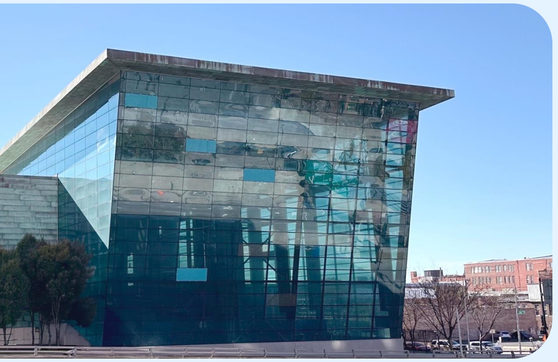
Lessons and practical tips for deploying similar systems in real-world biotech environments

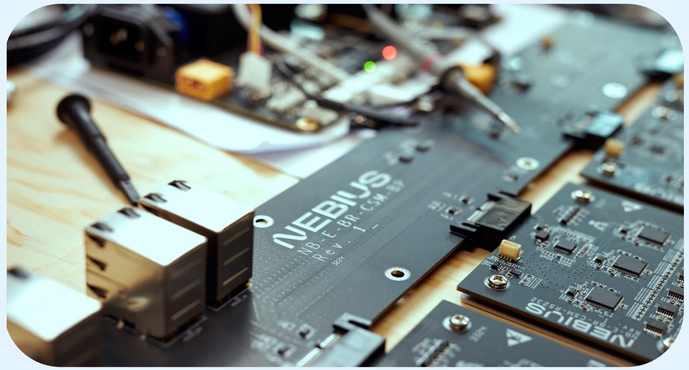
Agenda

1. How our collaboration started
2. Introduction of Quantori
3. Case: Generative chemistry at scale
4. How you can apply this in your work
5. Q&A

About Nebius

Nebius combines hyperscaler flexibility
with supercomputer performance.

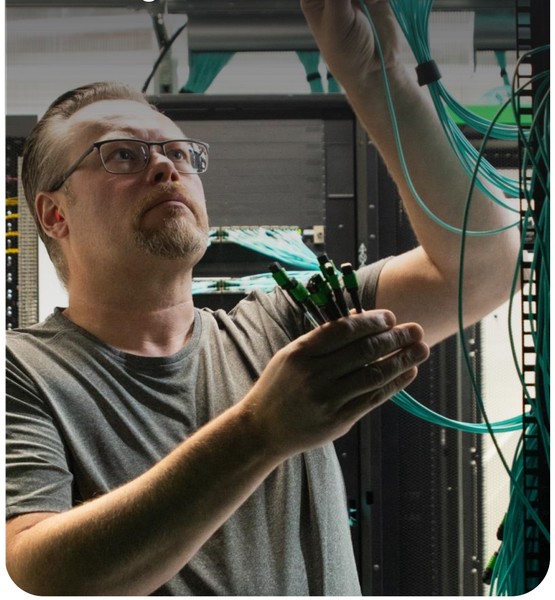




Key facts

Engineering-led by design

Decades of infrastructure experience, now fueling next-gen AI workloads.



NASDAQ-listed

Backed by deep capital and built for long-term scale.

NBIS

Strategic NVIDIA partner

Nebius is a Reference Platform NVIDIA Cloud Partner, with validated reference architectures and close collaboration with NVIDIA for enablement.

NVIDIA | NEBIUS

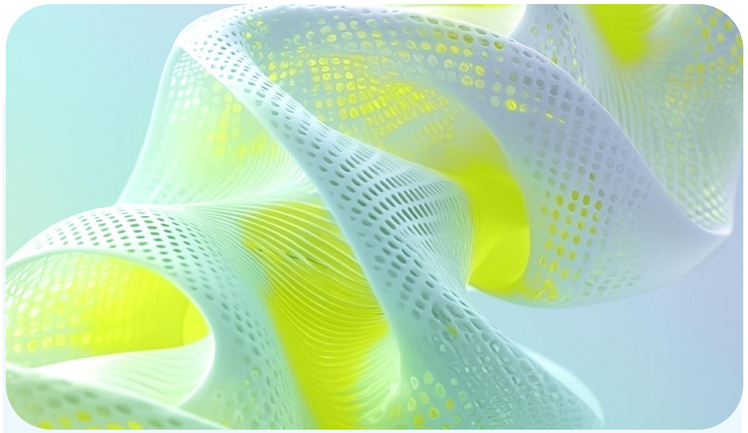
Vertically integrated

Full-stack control, from hardware manufacturing to platform software.

Accelerate your life-saving breakthroughs with our affordable, reliable and easy-to-use cloud infrastructure

Life sciences and drug discovery

With accelerated computing, researchers can virtually model millions of molecules and screen hundreds of potential drugs simultaneously, reducing costs and speeding time-to-solution.



Genomics and multiomics

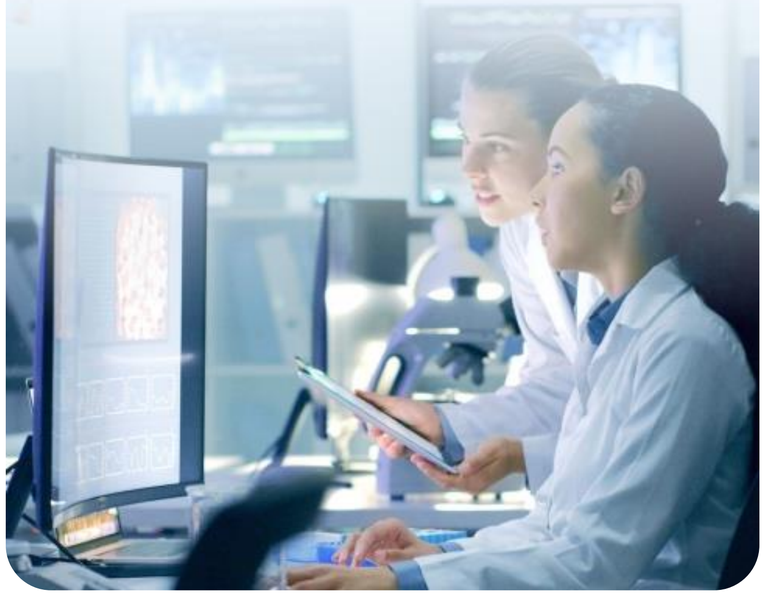
Using HPC to accelerate genome analysis in population and cancer genomic studies can help identify rare diseases and bring tailored therapeutics to market faster, advancing the journey to precision medicine.

HealthTech

AI-powered tools can be an extra set of eyes, helping to read images quickly, calculate measurements, monitor changes and identify urgent findings to optimize workflows and enhance patient care.

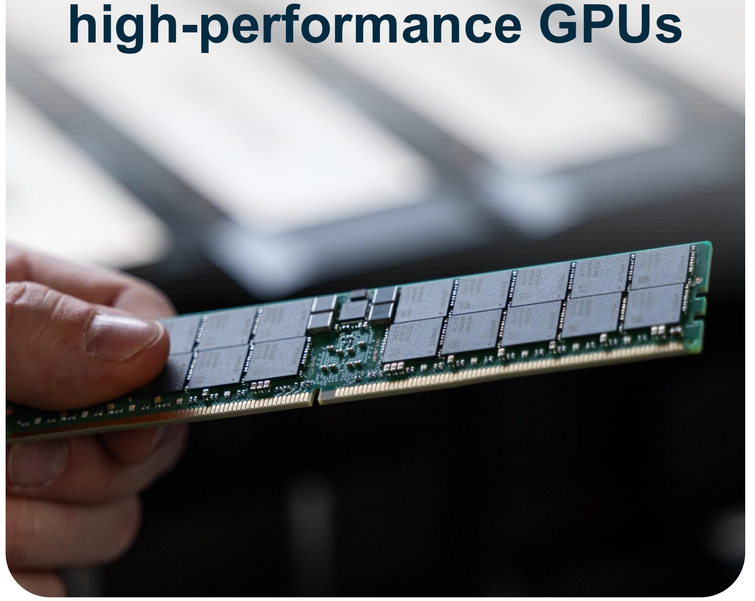
Life sciences use cases

Train foundational AI models for biological research



Run inference on state-of-the-art biological AI models

Accelerate your workflows by using your preferred tools on high-performance GPUs



QUANTORI × NEBIUS

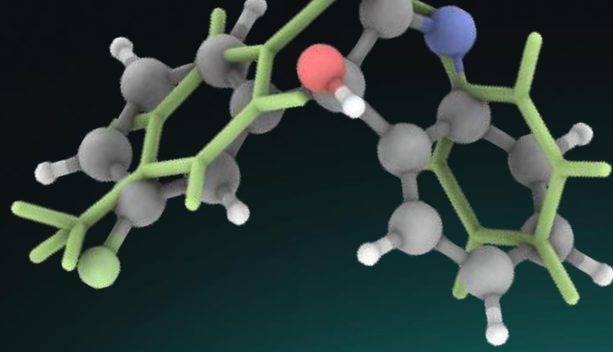
About project | Documentation | Contacts

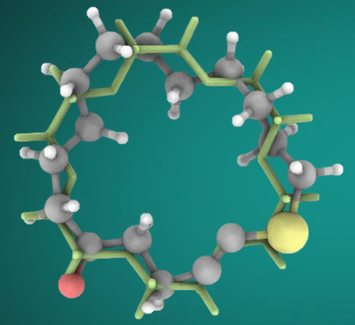
Discover Shape-Constrained Molecule Generation powered by Nvidia H100 GPU

Experience the future of molecular discovery with Quantori and Nebius.
Transform your ideas into reality with cutting-edge technology

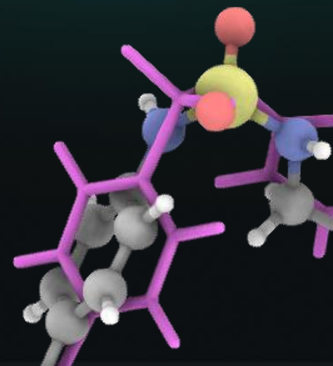
Generate molecules

QUANTORI x NEBIUS





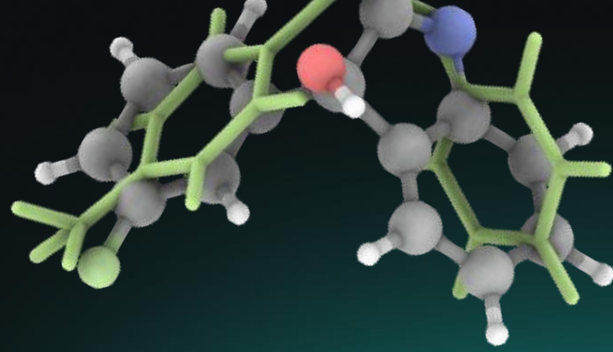
About Quantori

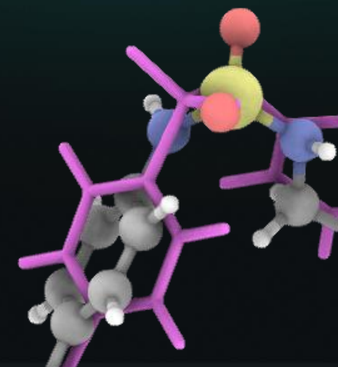


Quantori **is an asset-based informatics services company, focused exclusively on the healthcare and life sciences.**

We provide our clients with proprietary accelerators, platforms, blueprints, and customizable solutions delivered as part of our services.

With deep domain expertise and end-to-end capabilities in AI, data science, and software engineering, we help biotech, pharmaceutical, and healthcare organizations accelerate research, enhance data-driven decision-making, and improve patient outcomes.

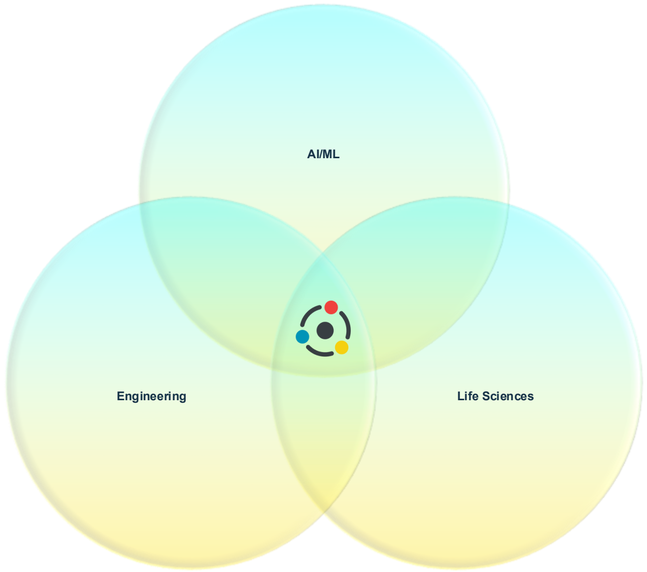




Quantori's Differentiator

NEBIUS x QUANTORI

Combination of domain knowledge, AI/ML expertise, and engineering excellence.



Cloud Design & Operations

We help clients migrate their data, optimize cloud presence, and provide support and maintenance for their systems.

AI, Machine Learning & LLMOps

We build AI/ML and LLM-powered solutions to enhance data accuracy, accelerate life sciences R&D, and streamline pharma operations.

Data Science & Engineering

We create the architecture, layout, workflows, and initial data systems to ensure the right data remains current and that new data can be easily added or removed.

High Performance Computing

Our HPC capabilities and accelerators enable biotech innovators to kick-start research computations with a single click.

Laboratory & Clinical Informatics

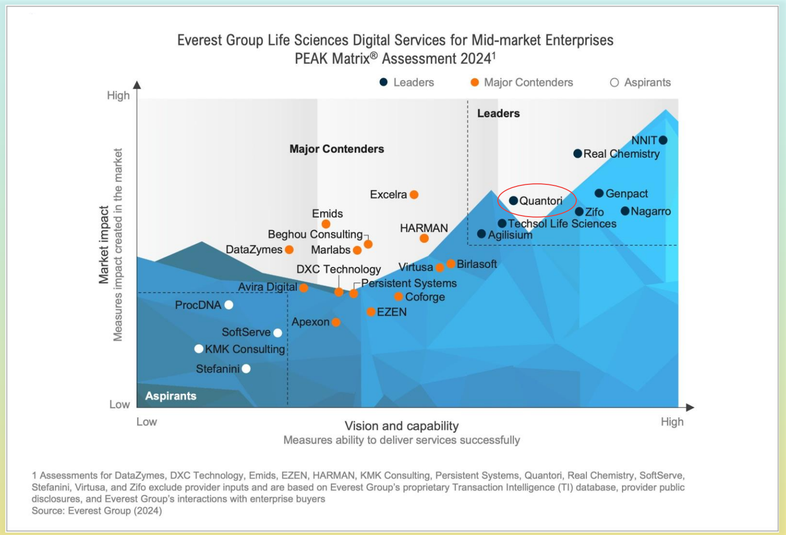
Alignment within laboratory & clinical informatics brings our clients closer to their patients through improved diagnosis, treatment methods, and drug development.

Quantori is a thought leader and trusted partner

NEBIUS x QUANTORI

30 peer-reviewed scientific articles (2019 – present)

Leader in the Everest Group PEAK Matrix Assessment



12 of the 15

Largest pharmaceutical companies served

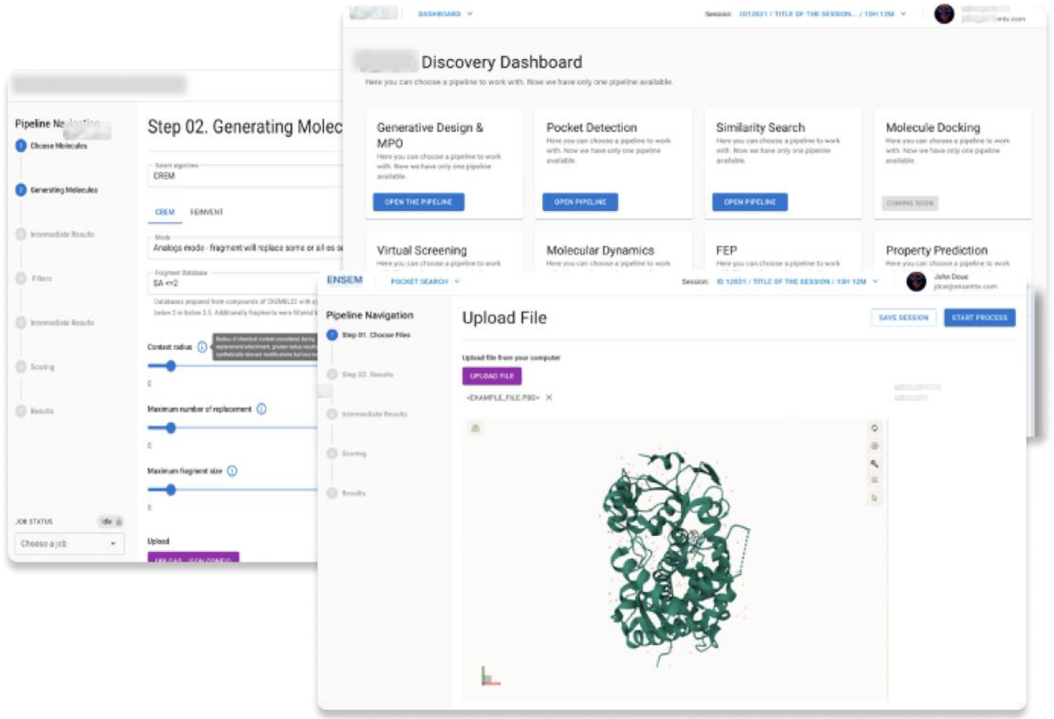
30+ Startups & Scale-ups

700+ Experts

Software engineers, data scientists, domain specialists, cloud engineers, digital health experts

Life Science Co-Pilot Platform

NEBIUS x QUANTORI



R&D Workflows

- Molecule Design & Optimization
- Target identification and validation
- Biomarker research
- Protein Engineering & Modelling
- Protein Folding & Visualization
- Molecular Dynamics
- Pocket Detection / Structure Detection
- Single-Cell Analysis and Cell State Prediction
- Custom Drug Discovery Workflows

A customizable platform brings together domain-specific intelligence, seamless data integration, and enterprise-grade governance, enabling faster insights, smarter decisions, and real-world impact across life sciences and healthcare.

Quantori’s AI-Driven R&D Service

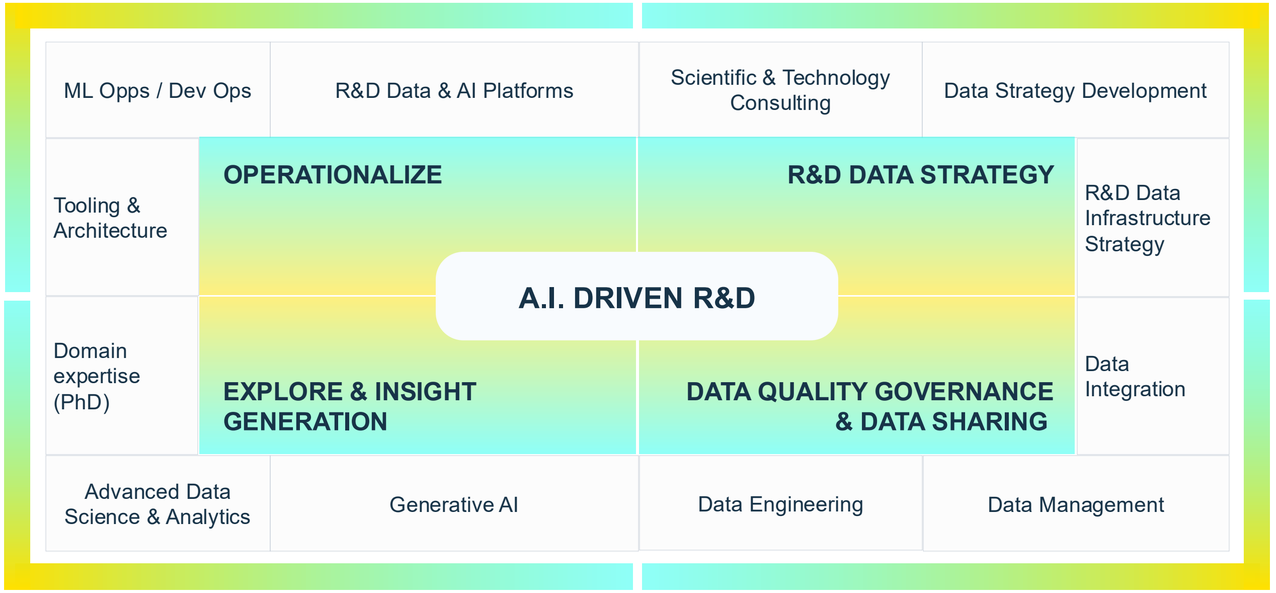
NEBIUS x QUANTORI

Operationalize

Industrialize deployment of complex, domain focused data-driven solutions at scale

Explore

Improve data science outcomes and quality in R&D by addressing gaps in capability and domain experience



Strategy

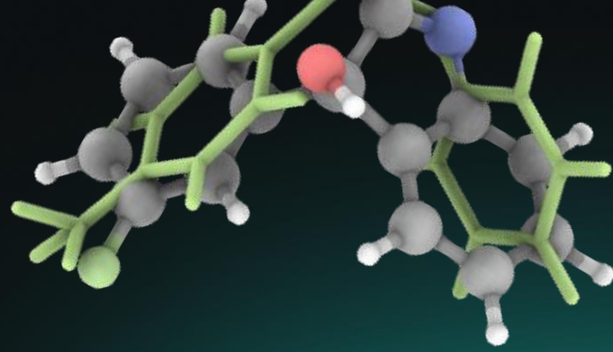
Development of data management and application strategy

Data

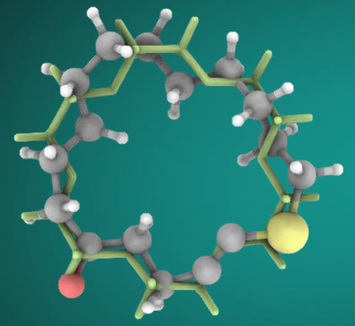
Robust data governance and enable the creation of a strong data ecosystem

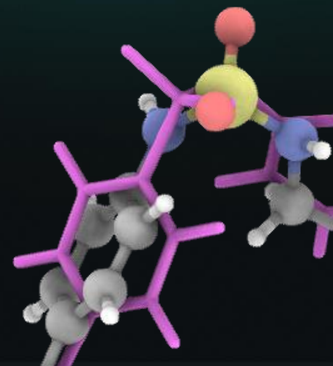
Quantori offers AI strategy consulting and builds proprietary AI platforms to increase workflow efficiency and boost company valuation.

QUANTORI x NEBIUS



MLConfGen: Shape Constrained Molecular Generation





Usage of physically-grounded shape descriptors for efficient training and creative inference

Practical landscape of structure-based molecule generation

NEBIUS x QUANTORI

Structure-based molecular generation

- Designing conformers that are both chemically valid and geometrically meaningful is highly complex.
- A classic example is ligand design: molecules must fit into protein binding pockets and form key interactions.
- Other applications include:
  - Catalysis
  - Macromolecular engineering
  - Material science

Use in real-world drug design workflows

- ML models are supportive tools, not standalone drug generators.
- Typically used to:
  - Expand compound libraries
  - Modify known scaffolds
  - Propose new ideas based on shape/property goals
- Human experts still play a crucial role — blending intuition with AI-generated suggestions.
- A core challenge remains: designing molecules that mimic a specific 3D shape.

Equivariant diffusion as an efficient strategy for molecule generation

NEBIUS x QUANTORI

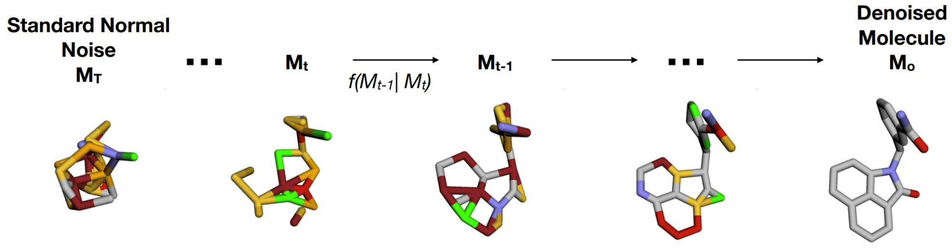
A class of generative models that **learn to reverse a noise process** applied to molecular structures.

Based on **score-based diffusion** or **denoising** frameworks.

Equivariance ensures predictions respect molecular symmetries: Translation, rotation, and permutation.

Why it works well for molecules

- Molecules exist in **3D space** — equivariant models naturally preserve physical meaning.
- Text-based Generative models may struggle with 3D consistency or require complex constraints.
- EDMs **learn geometry directl**y, generating conformers with:
  - Chemically valid structures
  - Realistic spatial arrangements
  - Good coverage of conformational diversity



Why descriptors matter

NEBIUS x QUANTORI

Shape descriptors define **what the model is optimizing for** — and how efficiently it can learn it.

Why the choice of descriptor is important

- **Training Efficiency:** Descriptors must be **quick to compute** for large datasets and repeated epochs.
- **Inference Speed:** At generation time, we want shape-conditioned outputs **without bottlenecks**.
- **Model Performance:** Descriptors must actually reflect **relevant geometric properties** — if they don't encode shape well, performance suffers.

Simplicity and meaningfulness over high-resolution

NEBIUS x QUANTORI

$[a_1 , a_2 , \ldots, a_n]$ n: 128 - 1024

Most models encode shape using **voxelized volumes** or **molecular surfaces**, processed through **heavy encoder networks**.

While aiming at precision, they are **slow, memory-hungry, and opaque.**

Slow to compute at training time and inference

High memory usage

Low interpretability

$[I_{xx} , I_{yy} , I_{zz}]$

The Moment of Inertia (MoI) tensor gives us a compact, physically grounded way to represent molecular shape — just **3 values** that capture how mass is distributed in 3D space.

Fast to compute — ideal for both training and inference

Rotation-aware and **interpretable**

Competitive precision

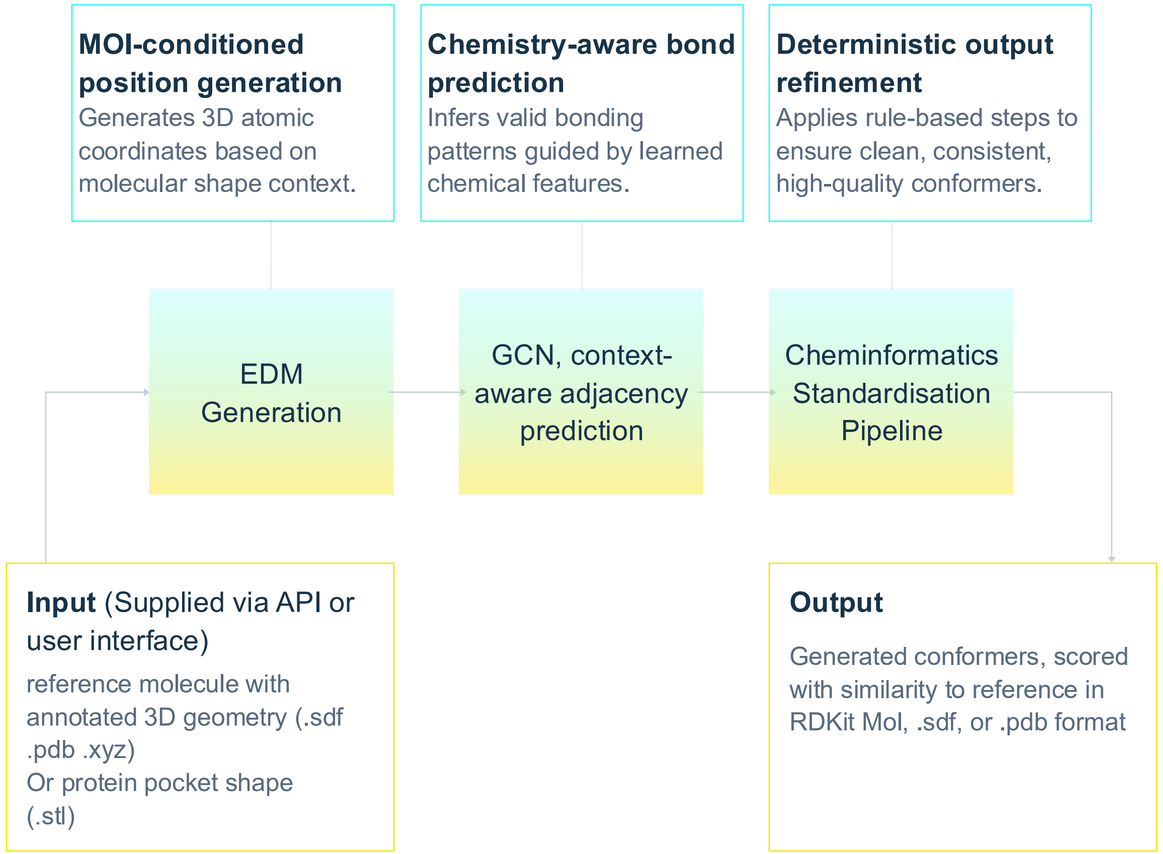
Physically meaningful — unlocks math tricks for smart inference

A descriptor that's not just smart — but usable at scale.

ML Conformer Generator – a smarter way to generate molecules

NEBIUS x QUANTORI

The model was **trained on 1.6M of ChEMBL molecules** on Nebius Cloud with 8× NVIDIA H200 — large-scale, fast, and efficient



ML Conformer Generator raw benchmarks

NEBIUS x QUANTORI

Technical Performance

- The average time for the generation of 50 samples is 11.46 sec (NVidia H100) and 250.46 sec (CPU)
- Average Generation speed (NVidia H100) - 4.18 molecule/sec (GPU) and 0.20 molecule/sec (CPU)
- Estimated memory Consumption per single Generation thread (50 samples) - up to 4.0 GB

Generation Quality

Average Shape Tanimoto similarity	**53.32%**
Maximum Shape Tanimoto similarity	**99.69%**
Average Chemical Tanimoto similarity (Morgan FP size - 2048)	**10.87%**
% Of chemically unique molecules in reference to training dataset (not found in training dataset)	**99.84%**
% Of valid molecules in generated batch (as defined by the standardisation pipeline)	**48%**
% Of chemically unique molecules within the generated set	**99.94%**
Freschet Fingerprint Distance (Morgan FP size - 2048)	
ChEMBL	**4.13**
PubChem	**2.64**
ZINC (250k drugs subset)	**4.95**

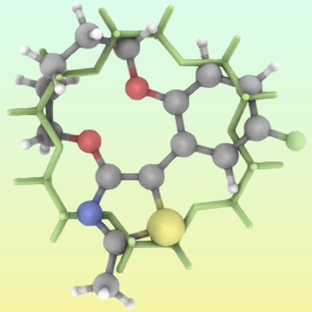
ML Conformer Generator: general use case

NEBIUS x QUANTORI

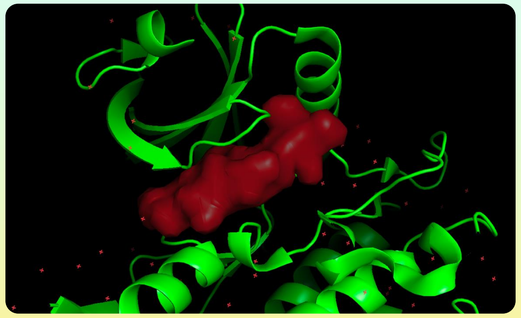
Seamless integration with existing Cheminformatic tools empowers users to rapidly generate novel molecules with high shape similarity to reference conformers or directly design structures based on a target pocket shape, driving efficient scaffold innovation and accelerating ideation.

- Faster generation of molecules tailored to specific binding sites or 3D shapes.
- Chemically sound and geometrically accurate molecules that support high-quality drug design.
- 3D graph-based representations for more precise molecular behavior prediction.
- Accelerated time-to-market for new molecules.

Generation from reference molecules



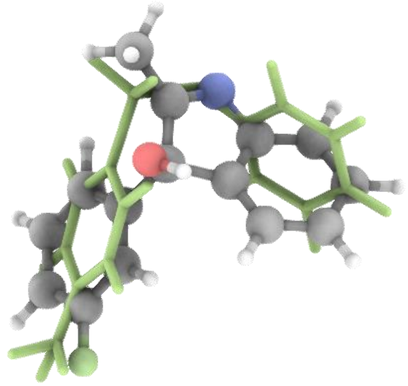
Generation from protein pocket shapes



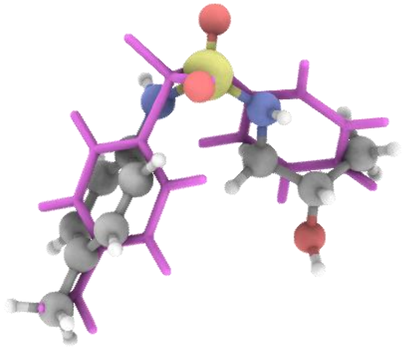
ML Conformer Generator: generation examples

NEBIUS x QUANTORI

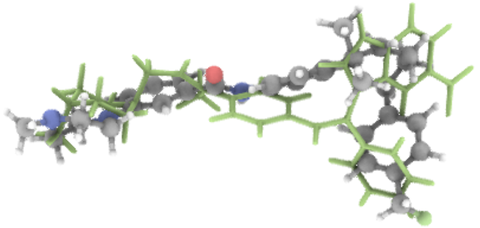
From Reference Molecular Conformer



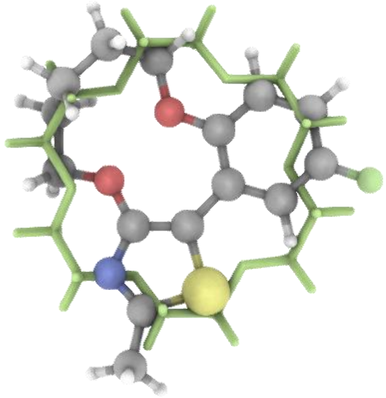
Shape similarity – 0.66



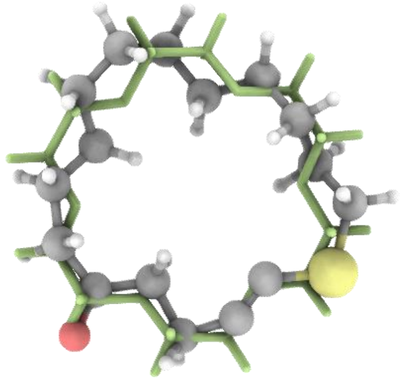
Shape similarity – 0.63



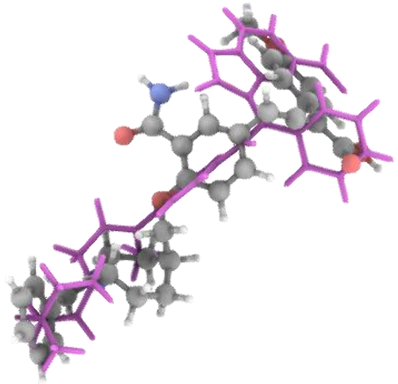
Shape similarity – 0.54



Shape similarity – 0.65



Shape similarity – 0.85

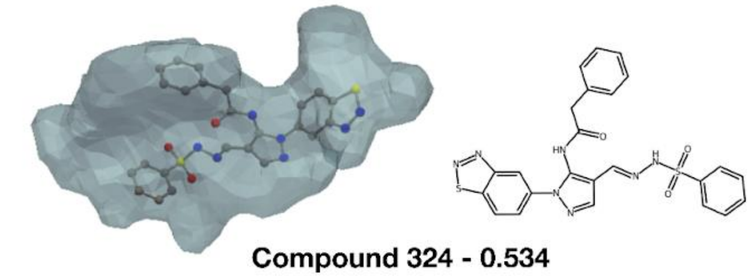


Shape similarity – 0.53

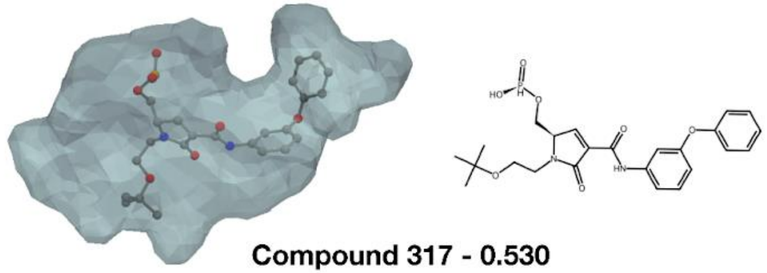
ML Conformer Generator: generation examples

NEBIUS x QUANTORI

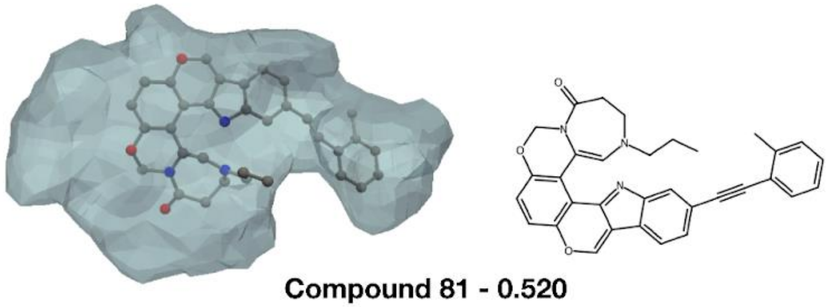
From Protein Pocket



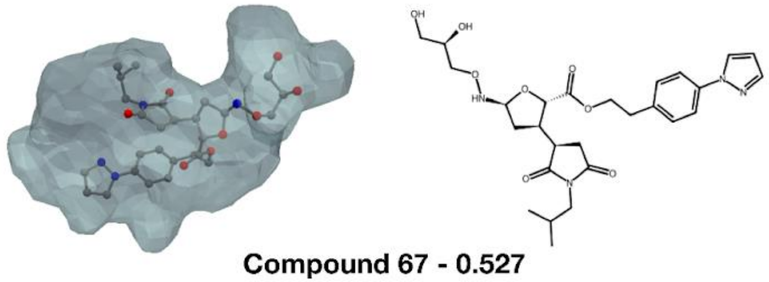
Compound 324 - 0.534



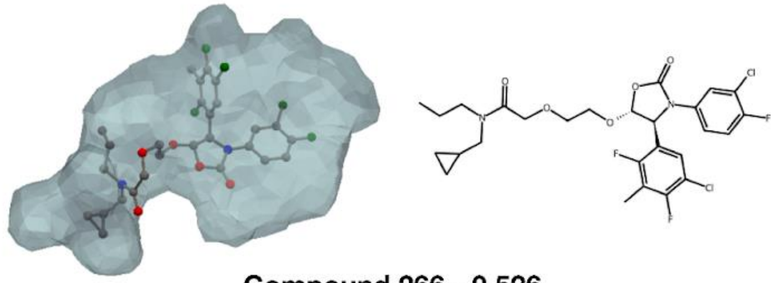
Compound 317 - 0.530



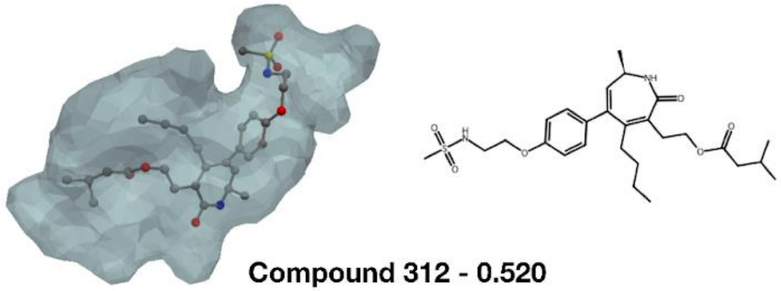
Compound 81 - 0.520



Compound 67 - 0.527

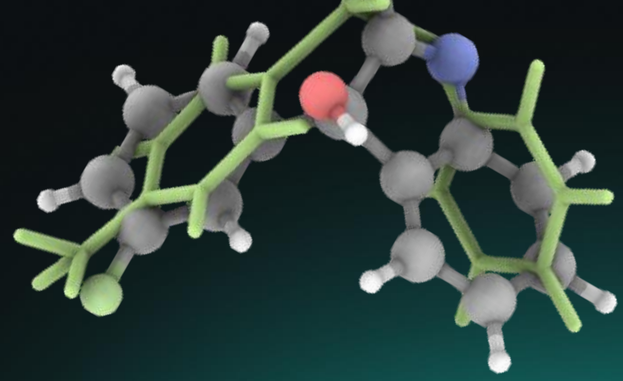


Compound 266 - 0.526

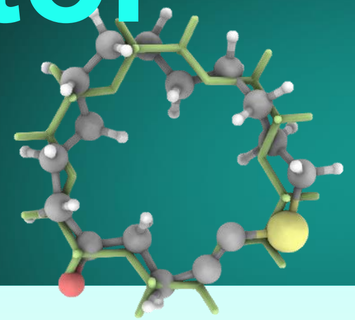


Compound 312 - 0.520

QUANTORI x NEBIUS



Bring ML Conformer Generator Into Your Pipeline



Inference as a Service

Access ready-to-use conformer generation through **Nebius Marketplace** or other cloud platforms — no setup required.

Open Source Access

Use our PyPI package with pretrained weights (non-commercial license) to evaluate and experiment freely.

GitHub

pip install mlconfgen

Custom Training & Fine-Tuning

Adapt the model to your own chemistry, fine-tuned on your proprietary data for maximum relevance and performance.

Custom Dataset Generation

Need specialized conformers for a specific scaffold class or shape constraint? We generate high-quality training or inference datasets to match your needs.

NEBIUS | QUANTORI

ml_conformer_generator

Learn more about Nebius for Life Sciences

Test Molecule Generation online

NEBIUS | QUANTORI

ml_conformer_generator

Learn more about Nebius for Life Sciences

Q&A

NEBIUS | QUANTORI

Artem Elmuratov on LinkedIn

Denis Sapegin on LinkedIn

Thank you and let’s stay in touch