Multi-Agent Drug Discovery Orchestra

Dr. Andrei Dmitrenko

21.10.25

About

- Senior consultant @ D ONE (Zurich, Switzerland)
- Lead researcher @ ITMO University (St. Petersburg, Russia)

- Short vita:
  - 2018 – systems biology team lead @ BIOCAD LLC
  - 2022 – doctor of sciences @ ETH Zurich
  - 2024 – data & software engineer @ Calico Life Sciences LLC
  - 2025 – 3 posters and the Top Reviewer award @ NeurIPS conference

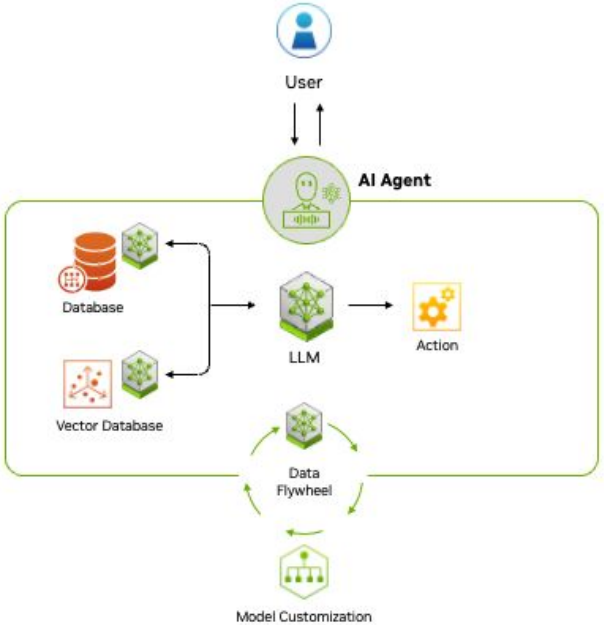
Center for AI in Chemistry @ ITMO University

- Since 2022
- Now >30 people
- A Master’s program (21 students)
- World-class research
- Academic and industrial collabs
- Core competences in:
  - Cheminformatics
  - Materials science
  - Bioinformatics
  - Artificial intelligence

Agentic AI is the new frontier

- Agent is:
  - a stateful LLM app designed to fulfil a particular role
  - equipped with tools and memory
  - free to reason, use tools, and perform actions in the real world

- Agent is, therefore, much like a human at work



Applications of agentic systems

nature › npj computational materials › articles › article

Article | Open access | Published: 23 June 2025

Agent-based multimodal information extraction for nanomaterials

R. Odobesku, K. Romanova, S. Mirzaeva, O. Zagorulko, R. Sim, R. Khakimullin, J. Razlivina, A. Dmitrenko & V. Vinogradov

npj Computational Materials **11**, Article number: 194 (2025) | Cite this article

6358 Accesses | **1** Citations | **20** Altmetric | Metrics

Benchmarking Agentic Systems in Automated Scientific Information Extraction with ChemX

Anastasia Vepreva[1] **Julia Razlivina**[1]
Maria Eremeeva[1] **Nina Gubina**[1] **Anastasia Orlova**[1] **Aleksei Dmitrenko**[1]
Ksenya Kapranova[1] **Susan Jyakhwo**[1] **Nikita Vasilev**[1] **Arsen Sarkisyan**[1]
Ivan Yu. Chernyshov[1] **Vladimir Vinogradov**[1] **Andrei Dmitrenko**[1,2]

[1]Center for AI in Chemistry, ITMO University, St. Petersburg, Russia
[2]D ONE AG, Zurich, Switzerland

dmitrenko@pish.itmo.ru

MADD: Multi-Agent Drug Discovery Orchestra

Gleb V. Solovev [*1], Alina B. Zhidkovskaya[1], Anastasia Orlova[1], Nina Gubina[1], Anastasia Vepreva[1], Rodion Golovinskii[1], Ilya Tonkii[1], Ivan Dubrovsky[1], Ivan Gurev[1], Dmitry Gilemkhanov[1], Denis Chistiakov[1], Timur A. Aliev[1], Ivan Poddiakov[2], Galina Zubkova[2], Ekaterina V. Skorb[1], Vladimir Vinogradov[1], Alexander Boukhanovsky[1], Nikolay Nikitin[1], Andrei Dmitrenko [†1,3], Anna Kalyuzhnaya [‡1], and Andrey Savchenko[2,4]

[1]ITMO University, Saint Petersburg, Russia
[2]Sber AI Lab, Moscow, Russia
[3]D ONE AG, Zurich, Switzerland
[4]HSE University, Moscow, Russia

NEURAL INFORMATION PROCESSING SYSTEMS

EMNLP 2025 Suzhou, China 中国苏州

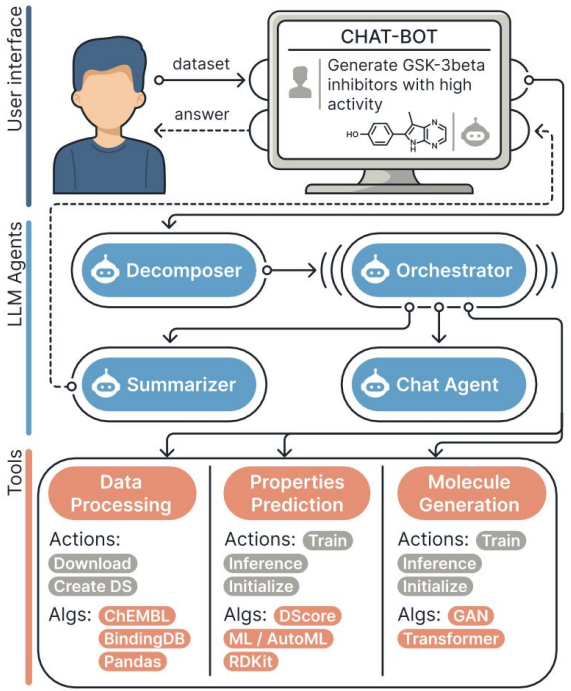
MADD

- A multi-agent system designed to address hit identification *in silico*
- Expects a user query in natural language as a default input

```
'input': 'Generate molecule of GSK-3beta
inhibitors with high docking score'
```

- Provides a list of generated molecules as a target output


```
{'Smiles': {'0': 'OC12C3C=C(Br)C4=
    NCC5(CN6CC65)C4N1CC2CC3',
          '1': 'COC1(OC)C2C3C4CCN(
              C#N)C3C41c1nncn12'},
 'Brenk': {...}, 'QED': {...}, '
    Synthetic Accessibility':
    {...},
 'LogP': {...}, 'Polar Surface Area
    ': {...}, 'H-bond Donors':
    {...},
 'H-bond Acceptors': {...}, '
    Rotatable Bonds': {...},
 'Aromatic Rings': {...}, 'Glaxo':
    {...}, 'SureChEMBL': {...},
 'PAINS': {...}, 'Validity': {...},
     'Duplicates': {...},
 'docking_score': {...}, 'IC50':
    {...}}
```

MADD approach

- Case-driven:
  - Alzheimer’s
  - Parkinson’s
  - Multiple sclerosis
  - Lung cancer
  - Dislipidemia
  - Drug resistance

  \+ *Thrombocytopenia*

- Systematic:
  - Producing two benchmarking assets: the dataset of validated user queries and 3M of scored molecules

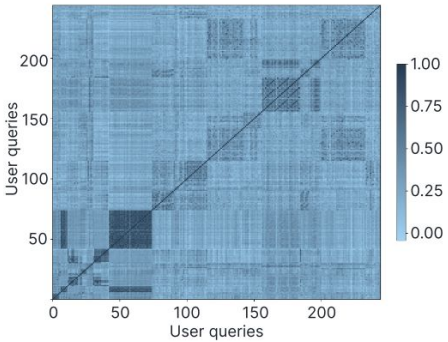
- Innovative:
  - Pioneering the application of AI-first drug design to five targets: STAT3, ABL, COMT, ACL, and PCSK9

Proposed benchmark

Queries — Dataset of >300 queries for 6 diseases

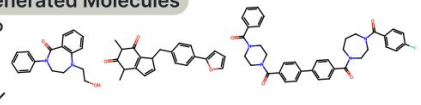
Example for Alzheimer:
Generate GSK-3beta inhibitors with high docking score and low brain-blood barrier permeability

Example for Dislipidemia:
Can you suggest molecules that inhibit Proprotein Convertase Subtilisin/Kexin Type 9 with enhanced bioavailability and the ability to cross the BBB?



Generated SMILES — >5000 molecules for 6 diseases

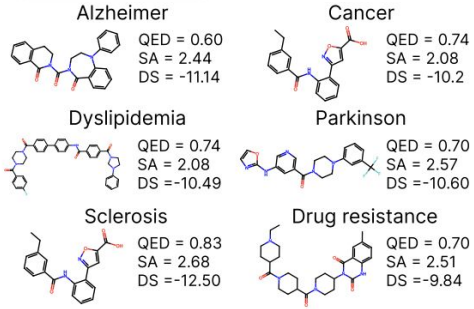
Generated Molecules



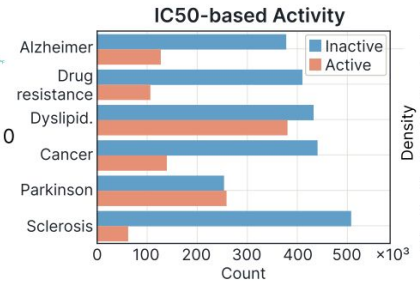
Filters

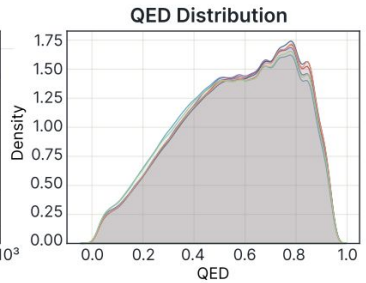
- DS <= -7
- IC50 = 1
- SA Score <= 3
- Brenk = 0
- SureChEMBL = 0
- Glaxo = 0
- PAINS = 0
- QED > 0.6

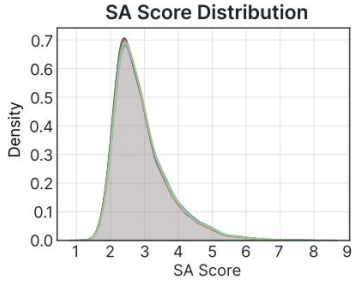
Selected Molecules

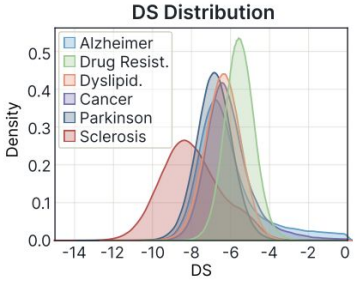


Molecule properties — Molecular properties dataset (DS, IC50, SA, QED) for >3M molecules









Agentic system design

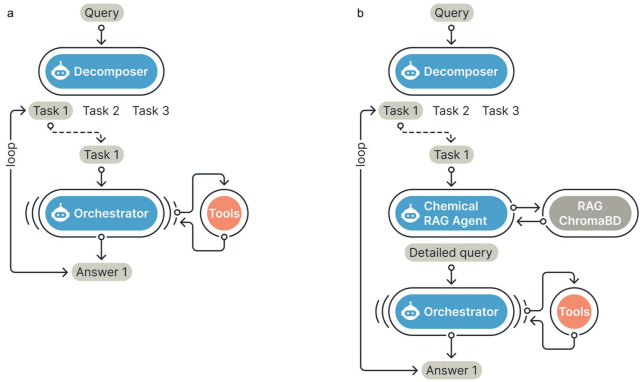


Figure 12: Visualisation of MADD-v2A and MADD-v3 systems.

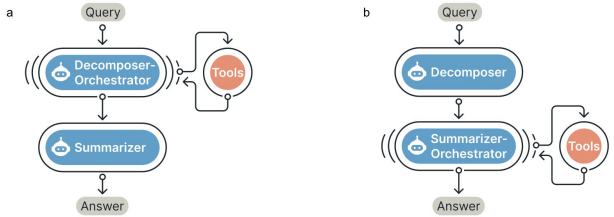
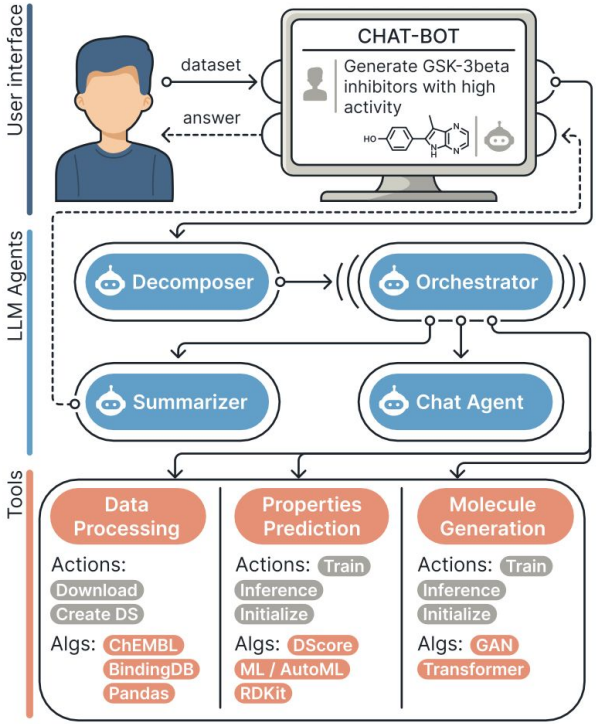


Figure 13: Visualisation of MADD-v2C and MADD-v2B systems.

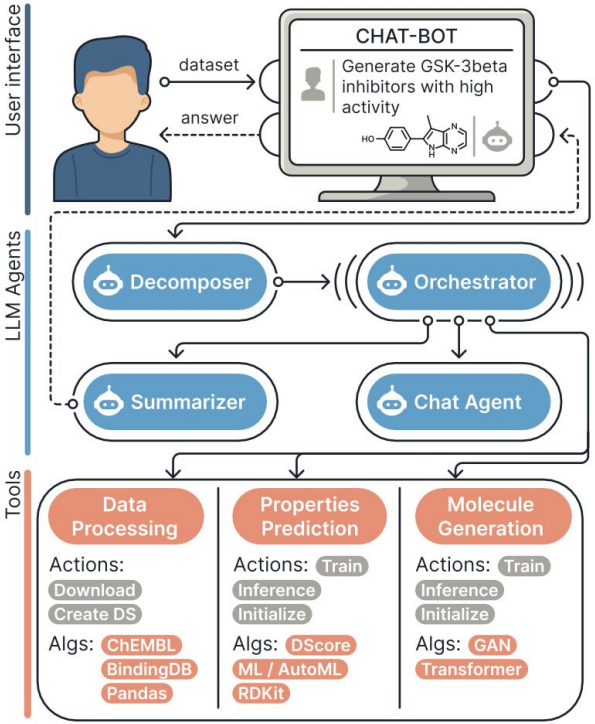


Agentic system design

Table 1: Comparison of MADD with other candidate architectures for multi-agent systems by TS, SSA, and FA metrics.

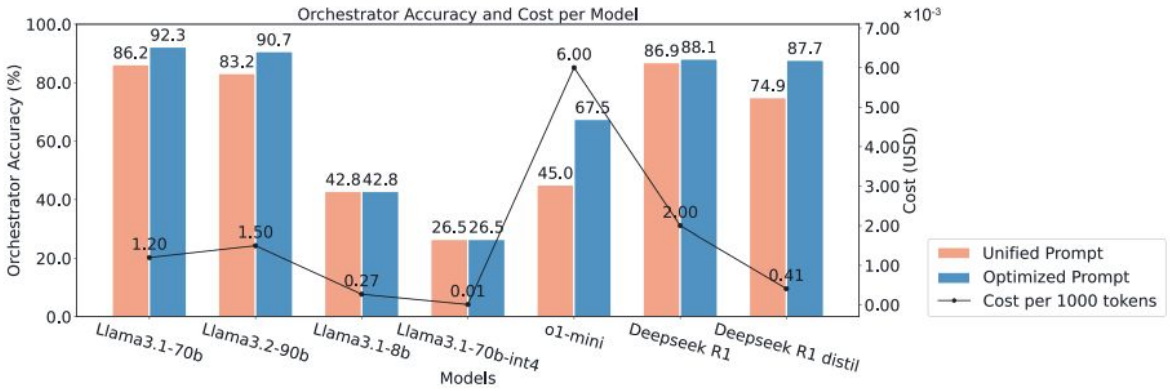
Metric	TS	SSA	FA
MADD	**83.7**	**95.3**	**79.8**
MADD-v1	42.5	70.0	29.8
MADD-v2A	35.0	65.1	22.8
MADD-v2B	76.7	95.3	73.0
MADD-v2C	81.4	53.7	43.7
MADD-v3	46.5	75.6	35.2

- TS – Tool Selection
- SSA – System Summarization Accuracy
- FA – Final Accuracy



Agentic system design

- Orchestration is all you need



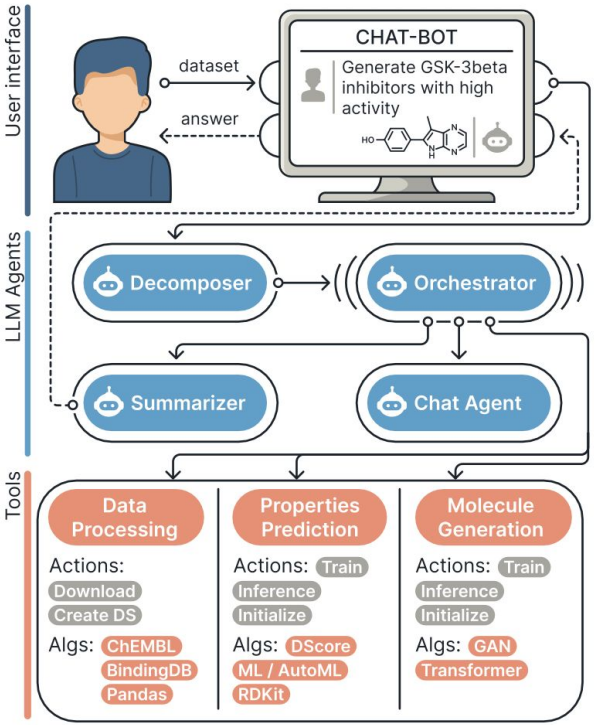
Levels of complexity

Table 2: Comparison of the Final Accuracy (%) of MADD and baseline methods on datasets of different complexity.

Dataset	S	M	L
MADD	**86.9**	**84.3**	**79.8**
ChemAgent	12.4	15.3	16.4
LlasMol	0.46	0.24	0
X-Lora-Gemma	0.44	0.12	0
ChemDFM	5.31	0.33	0

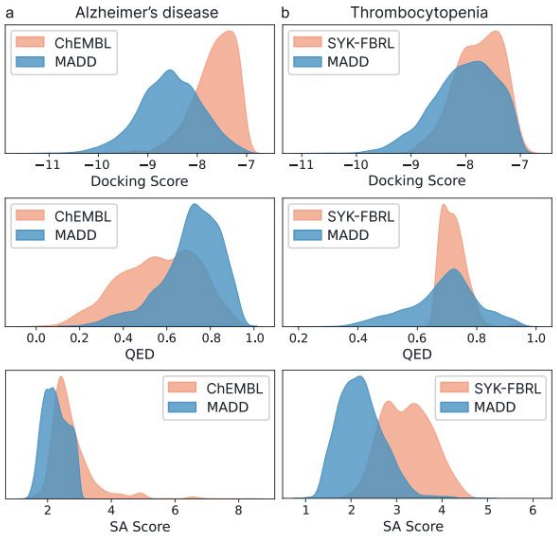
- S – only single-task queries
- M – up to 3 tasks in a query
- L – up to 5 tasks in a query

* ***Phoenix*** *from the FutureHouse Platform managed to create a single molecule that satisfied the 2/5 filter groups, but no more. The molecules produced by* ***TxGemma*** *did not pass any filters.*



Validation case studies

- Reproducing experimentally validated molecules for validation



Case	Alzheimer		Thrombocytopenia	
Model	MADD	MADD Auto	SYK-FBRL	MADD Auto
Mean DS	-7.46	**-7.57**	-7.76	**-8.02**
Novelty	**78.21**	73.47	100	100
Validity	87.47	**89.5**	**100**	90.71
GR1,%	**20.30**	15.99	0.18	**1.54**
GR2,%	**17.56**	14.43	0.07	**1.35**
GR3,%	**13.72**	13.14	0.06	**1.32**
GR4, %	**13.40**	12.34	0.06	**1.32**
GR5,%	**13.40**	12.34	0.06	**1.32**

What MADD is and isn't

Is:

- A first-of-its-kind effective *application* of multi-agent systems to early drug discovery
- A powerful *showcase* for the next-generation pharma R&D tools
- A versatile *interface* for the wet-lab researchers

Is not:

- A universal framework to solve drug design
- A commercially available product
- A definitive solution for hit identification – *a lot more to come*

Multi-Agent Drug Discovery Orchestra

Gleb Solovev*, Alina Zhidkovskaya, Anastasia Orlova, Nina Gubina, Anastasia Vepreva, Rodion Golovinskii, Ilya Tonkii, Ivan Dubrovsky, Ivan Gurev, Dmitry Gilemkhanov, Denis Chistiakov, Timur Aliev, Ivan Poddiakov, Galina Zubkova, Ekaterina Skorb, Vladimir Vinogradov, Alexander Boukhanovsky, Nikolay Nikitin, **Andrei Dmitrenko***, Anna Kalyuzhnaya*, Andrey Savchenko

EMNLP 2025 Suzhou, China 中国苏州

Thank you

Work & collaboration:
dmitrenko@pish.itmo.ru

Appendix: molecule generation

		GR1	GR2	GR3	GR4	GR5	Diversity
Drug resistance	GAN	0.23	0.15	0.10	0.10	0.10	0.39
	Transformer	**8.32**	**6.92**	**6.14**	**6.05**	**6.05**	0.77
	RL	0.63	0.52	0.40	0.38	0.38	0.13
	MTDD-EF	0.00	0.00	0.00	0.00	0.00	0.11
	ChemTSv2	0.14	0.14	0.09	0.09	0.09	0.43
Dyslipidemia	GAN	7.27	6.15	4.92	4.72	4.72	0.34
	Transformer	**28.87**	**28.27**	**25.07**	**24.50**	**13.16**	0.21
	RL	0.02	0.02	0.02	0.02	0.02	0.05
	MTDD-EF	0.00	0.00	0.00	0.00	0.00	0.06
	ChemTSv2	0.12	0.12	0.06	0.06	0.06	0.44
Lung cancer	GAN	5.53	4.41	3.43	3.31	3.31	0.39
	Transformer	6.12	5.72	**4.97**	**4.76**	**4.76**	0.8
	RL	0.57	0.53	0.51	0.51	0.51	0.09
	MTDD-EF	1.00	1.00	1.00	1.00	1.00	0.11
	ChemTSv2	**6.65**	**6.65**	4.06	3.53	3.53	0.43

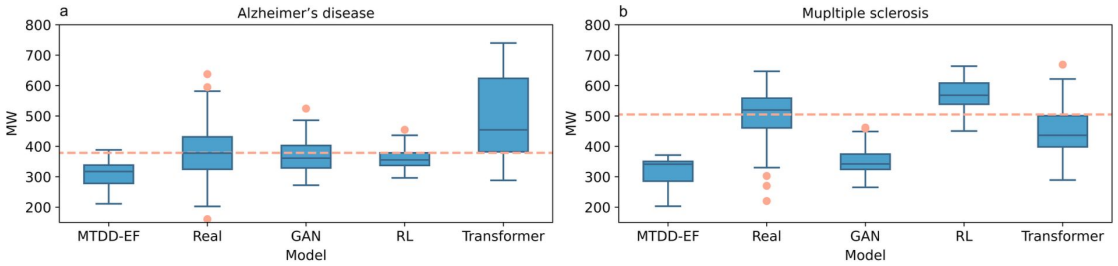


Figure 11: Comparison of the molecular mass of experimentally validated molecules ("Real") with generated molecules using MTDD-EF and integrated models for: a) Alzheimer's disease, b) Multiple sclerosis.

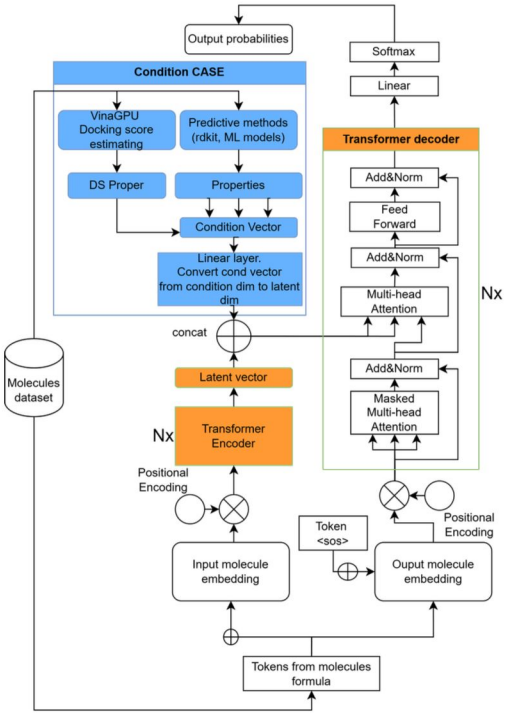


Figure 10: Our CVAE transformer architecture

Model	GAN	Transformer
GPU memory (GB)	**6.40**	8.43
Training time (hours)	**2.82**	23.73
Generation time (ms/molecule)	**0.19**	295.00