



- Případy pro lokální LLM (Martin Čmejrek)
- Využití AI ve vývoji léků (Jan Macek)

AI Days, 23. října, 2024  
Konferenční centrum City

# MAMA AI Team History



## IBM Research (1992-2021)

Basic research in the fields of Speech recognition, Speech synthesis, Conversational AI, Machine Translation



## IBM Watson (2014-2021)

Conversational, Speech & NLP on IBM Cloud

- Watson Assistant
- Watson Speech To Text
- Watson Text To Speech
- Watson Language Translator



## The MAMA AI (2021-present)






European Group that build practical sustainable AI Products and Service using applied Artificial Intelligence






# MAMA AI builds practical AI






## Search AI (mSearch, MACHI, CHEESE)

- Hybrid/Semantic search of data  
- Many languages  
- Fast and intuitive 
- Similarity search across many sources, inc. molecules




## Communication AI (Telma, mSpeech, mVoice)

- Text, voice, omnichannel assistants
- Information gathering  
- Automatic notifications and reminders
- Call routing, email routing
- Voice Digital Twins, Tailored voices 

## Personalized AI (mStudio, ppt2video)

- Right time, right place, right amount of information  
- Recording and delivering personalized information 
- Personalized Sensing and Recommendation

## Metaverse AI (mVR)

- Virtual and augmented reality combined with AI  
- Stress situation exercises 
- Production line / Manufacturing training

## Reasoning & Generative AI (mAI)

- Data Science / Data modeling
- Predictive modeling
- LLMs incl. GPT-like architectures
- Anomaly detection

## Blockchain AI (mCrypto)

- Use AI to detect the vulnerability of Smart Contracts
- AI-enhanced blockchain

# mSearch

---

mSearch umožňuje vyhledávání v textech podle významu  
Nezávisle na jazyku prohledávaných dokumentů dostaneme odpověď ve  
stejném jazyce, jako byla otázka. Podporuje více než 100 jazyků, všechny  
jazyky EU a další.







Logged in as: public [Logout](#)

Choose document collection

German\_News\_10k

Search mode

semantic

Query

Dostaly se na trh nějaké nebezpečné potraviny?

Documents to return



Minimal confidence score

0.0

Highlight



Answer right away



Restore diacritics



Search

Answer

Export

Answer: Ano, v textu se zmiňuje o různých nebezpečných látkách v potravinách, jako jsou salmonelly v šunkách, listerie v sýrech, plísně v těstovinách nebo skleněné střepy v okurkových sklenicích. Na internetové stránce lebensmittelwarnung.de jsou zveřejňovány varování před těmito potravinami. Také se zmiňuje o obavách z kontaminace japonských potravin po havárii v jaderné elektrárně Fukushima.



Provide feedback (optional)...

News\_id

News\_content

TEXT003719

Salmonellen im Schinken Listerien im Käse Schimmelpilze in der Pasta oder Glassplitter im Gurkenglas -über die Warnungen vor derlei gesundheitsgefährdenden Lebensmitteln können Verbraucher sich erstmals auf einer zentralen Internetseite informieren. Auf lebensmittelwarnung.de veröffentlichen die zuständigen Behörden aller Bundesländer ab sofort ihre Hinweise wie Verbraucherministerin Ilse Aigner (CSU) am Freitag mitteilte.

TEXT000078

So wie die Teefirma Aiya sehen sich derzeit zahlreiche japanische Nahrungsmittelhersteller mit den Sorgen ihrer Kunden konfrontiert. Fische Meeresfrüchte grüner Meerrettich Nudelgerichte oder Sojasaucen aus dem asiatischen Land stehen unter Generalverdacht. Bislang geben

# Helpdesk tickets

Check Historic Tickets  
First!

87%  
solved\*

- Use semantic search to identify similar historic tickets
- Find solution in historic tickets
- Collect feedback and learn
- Report trends

[INC33234560](#): FM672491 | Vyhodnocení KPI | zima v 1. patře pobočky

[INC33234561](#): FM672492 | Odloženo | Nefunguje el. zámek dveří mezi chodbou a 1 ✓

Similar incidents: Rate

1. [INC33328847](#) : FM672664 | V realizaci | nefuguje zámek u dveří - 0.8485
2. [INC33375984](#) : FM672741 | V realizaci | nezavírají se dveře do pobočky - 0.8435
3. [INC32904681](#) : FM672056 | Vyhodnocení KPI | zavírání dveří do SAZO - 0.8335
4. [INC33281717](#) : FM672593 | Vyhodnocení KPI | dveře - 0.8304
5. [INC33328832](#) : FM672645 | V realizaci | dveře - 0.8282
6. [INC33187393](#) : FM672397 | V realizaci | okno na WC - 0.8230
7. [INC33187399](#) : FM672402 | Vyhodnocení KPI | Dveře na 4. patře - 0.8210
8. [INC32951838](#) : FM672145 | Odloženo | oprava omítky - 0.8089
9. [INC33187404](#) : FM672406 | V realizaci | Výměna zámkové vložky. - 0.8088
10. [INC33328845](#) : FM672661 | V realizaci | nefunkční zásuvka - 0.8069

Similar problems: Rate

[INC33234562](#): FM672493 | Zadáno dodavateli | OČTŘ ID 18895454

[INC33234564](#): FM672496 | V realizaci | závada TE-ELSI-NO-K (HJ\_3693)

[INC33234565](#): FM672497 | Předáno k vyjádření | Vadné inventéry

[INC33234566](#): FM672498 | Vyhodnocení KPI | Odstranění skvrny z koberce - místnost 7

# Tender processing

Agent technologies harnessing the full potential of Large Language Models (LLMs) to enhance document search

Eva vytěží relevantní informaci o tendru, jeho výši, akceptačních podmínkách, atd.

**MAMA** 16-09-58 Vítejte! Abychom mohli začít, tak nejdříve nahrajte Vaše PDF nebo textové soubor.

16-10-04 Soubory Svazek 1.pdf, Svazek 2.pdf, Svazek 3.pdf, Svazek 4.pdf zpracovány. Nyní můžete začít klást otázky!

[Generate Answers](#) [Upload New File](#) [Run Test Campaign](#)

**User** 16-10-09 Kolik je hodnota veřejné zakázky

Took 1 step ▾

**MAMA** 16-10-38 Předpokládaná hodnota veřejné zakázky činí celkem 17 834 950,- Kč bez DPH.

**User** 16-11-17 jaká je požadovaná kvalifikace

Took 1 step ▾

**MAMA** 16-13-09 Požadovaná kvalifikace zahrnuje prokázání základní způsobilosti, profesní způsobilosti a technické kvalifikace. Technická kvalifikace zahrnuje prokázání lidských a technických zdrojů, odborných schopností a zkušeností dodavatele. K prokázání technické kvalifikace je požadován seznam stavebních prací poskytnutých za posledních 5 let. Veškeré odborné práce musí vykonávat pracovníci dodavatele nebo jeho poddodavatelů mající příslušnou kvalifikaci.

**User** 16-14-26 jaká jsou hodnotící kritéria

Took 1 step ▾

**MAMA** 16-15-36 Hodnotící kritéria mohou zahrnovat požadavky na kvalifikaci dodavatelů, základní způsobilost dodavatele, profesní způsobilost dodavatele, technická kvalifikace dodavatele, podmínky pro podání nabídky, metoda hodnocení, dodržování podmínek rozhodnutí dotčených orgánů a organizací, a podmínky, za kterých je možné pověřit realizaci díla jinou osobu.

⏏ 🗨️ + Type your message here... 📤

# Explainability

# Chain-of-thought

# References

# Feedback

Eva poskytne vysvětlení jak k dané informaci došla..

User  
16:02:07

Kolik je hodnota veřejné zakázky

Took 1 step ^

AgentExe...

16:02:07

Took 3 steps ^

LLMChain

16:02:08

Thought: Potřebuji zjistit hodnotu veřejné zakázky. K tomu použiji nástroj mSearch, který je určený pro vyhledávání informací v dokumentech týkajících se veřejných zakázek.

Action: mSearch

Action Input: Kolik je hodnota veřejné zakázky



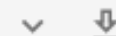
mSearch

16:02:25

Search output, here are the relevant snippets:

3.4. PŘEDPOKLÁDANÁ HODNOTA VEŘEJNÉ ZAKÁZKY

Předpokládaná hodnota veřejné zakázky činí celkem 17 834 950,- Kč bez DPH.



LLMChain

16:02:26

Výsledky vyhledávání ukazují, že předpokládaná hodnota veřejné zakázky je 17 834 950 Kč bez DPH.

Status: Success

Final Answer in Czech: Předpokládaná hodnota veřejné zakázky je 17 834 950 Kč bez DPH.



16:02:43

Předpokládaná hodnota veřejné zakázky je 17 834 950 Kč bez DPH.



MAMA  
16:02:43

Předpokládaná hodnota veřejné zakázky je 17 834 950 Kč bez DPH.





# Generate Image Descriptions



Logged in as: ct

Logout

Choose document collection

image\_gallery

Search mode

semantic

Query

pán v country klobouku

Documents to return



Minimal confidence score

0,0

Highlight



Answer right away



Restore diacritics



Search

Answer

3 selected

Export

id:

I000007



image\_description: *The image features a man wearing a cowboy hat and a brown jacket, standing in front of a sign. He appears to be the main focus of the scene. There are several other people in the background, some of them wearing ties, suggesting a formal or semi-formal gathering. The man in the cowboy hat seems to be the center of attention, possibly due to his unique attire or the fact that he is standing in front of a sign.*





```
[{'generated_text': 'two men standing on a stone wall in front of a building'}]
```

### **llava-hf/llava-1.5-7b-hf**

```
[{'generated_text': 'USER: \nDescribe what you see in the image.\nASSISTANT: The image features a man and a woman standing together on a sidewalk, posing for a picture. They are both wearing sunglasses, and the woman is wearing a blue dress. The couple is standing in front of a building, which could be a shop or a restaurant.\n\nThere are several other people in the scene, some of them standing or walking around the area. A bench is visible in the background, and a backpack can be seen placed on the ground nearby. Additionally, there are two bottles in the scene, one closer to the couple and the other further away.'}]
```

### **microsoft/kosmos-2**

```
<grounding> An image of<phrase> a couple</phrase>  
<object><patch_index_0298><patch_index_1009></delimiter_of_multi_objects/><patch_index_0292><patch_index_0971></object> standing in front of<phrase> the castle</phrase><object><patch_index_0032><patch_index_0575></object>.
```

```
An image of a couple standing in front of the castle.  
[('a couple', (12, 20), [(0.328125, 0.296875, 0.546875, 0.984375), (0.140625, 0.296875, 0.359375, 0.953125)]), ('the castle', (42, 52), [(0.015625, 0.046875, 0.984375, 0.546875)])]
```

## llava-hf/llava-1.5-7b-hf

### USER:

What are the things I should be cautious about when I visit this place?

**ASSISTANT:** When visiting this place, which is a pier or dock extending over a body of water, there are a few things to be cautious about. First, be aware of the weather conditions, as sudden changes in weather can make the pier unsafe to walk on. Second, be mindful of the water depth and any potential hazards, such as submerged rocks or debris, that could cause accidents or injuries. Additionally, be cautious of the presence of wildlife, such as birds or fish, and avoid disturbing their natural habitats. Lastly, be aware of any local regulations or guidelines for the use of the pier, as some areas may be restricted or prohibited for certain activities.





# Contract compliance verification

Armed Forces	NO	N/A	N/A	The provided contract wording does not contain any explicit reference to an 'Armed Forces' exclusion or anything that resembles this exclusion. The exclusions mentioned focus on various other aspects like nuclear risks, terrorism, and pollution, but do not specifically address vehicles of armed forces or similar military vehicles.
Asbestos	YES	"Any actual or alleged liability whatsoever for any claim or claims in respect of loss or losses directly or indirectly arising out of, resulting from or in consequence of, or in any way involving asbestos / silicosis, or any materials containing asbestos / silicosis in whatever form or quantity."	<a name="15_asbestos_silicosis"></a>	The contract explicitly includes an Asbestos/Silicosis exclusion which excludes any claim or liability related to losses arising out of or in consequence of asbestos or products containing asbestos. The presence of specific language about excluding liability in relation to asbestos indicates that this exclusion is present in the contract.
Communicable Diseases	YES	"This Reinsurance contract excludes any loss, damage, liability, claim, cost or expense of whatsoever nature, directly or indirectly caused by, contributed to by, resulting from, arising out of, or in connection with a Communicable Disease or the fear or threat (whether actual or perceived) of a Communicable Disease regardless of any other cause or event contributing concurrently or in any other sequence thereto."	<a name="exclusion_list"></a>	The contract includes a specific section labeled 'Communicable Disease Exclusion' that outlines the exclusion from coverage for any loss, damage, liability, claim, cost or expense caused by or in connection with a communicable disease. It clearly defines a 'Communicable Disease' in accordance with the typical industry standards, including references to transmission by infectious agents and the potential for public health impact.
Delegation of underwriting authority (Line Slips, Binding Authorities, Open Covers)	YES	"13. The delegation of underwriting authority on any basis."	<a name="exclusion_list"></a>	The exclusion regarding the delegation of underwriting authority is included in the contract as explicitly specified in the exclusion list. Specifically, the contract states that '13. The delegation of underwriting authority on any basis' is excluded, which aligns directly with the given definition and examples of delegation of underwriting authority.
Downstream Exclusion and Limited Writeback	NO	N/A	N/A	The provided contract wording does not include a specific mention of a 'Downstream Exclusion and Limited Writeback.' Instead, it lists several exclusions but does not detail any specific exclusions that correspond to onshore energy risks or related operations. Additionally, the term 'Downstream Exclusion' does not appear directly in the text, and its components related to oil refineries or onshore facilities are not explicitly stated. Thus, I conclude that this specific exclusion is not present.
Employers Liability/Workmen's Compensation	NO	N/A	N/A	The contract does not explicitly mention 'Employers Liability' or 'Workmen's Compensation' in the exclusion list. However, it does contain a comprehensive list of exclusions, none of which align with the criteria indicating that an exclusion of liability for bodily injury to employees or workers is explicitly stated. The closest reference is the statement that the contract covers 'any kind of third party liability coverage, other than the liability sections of CAR/EAR policies', which does not indicate an exclusion for employers' liability even implicitly.
Event Definition	YES	"The words 'Loss occurrence' shall mean all individual losses arising out of and directly occasioned by one event."	<a name="2_definition_of_loss_occurrence"></a>	The exclusion 'Event Definition' is indeed present in the contract, specifically under 'Definition of Loss Occurrence'. This section clearly defines 'Loss occurrence' as all individual losses arising out of and directly occasioned by one event, outlining the unities of locality, time, and causality essential for determining what constitutes a single event or loss occurrence.



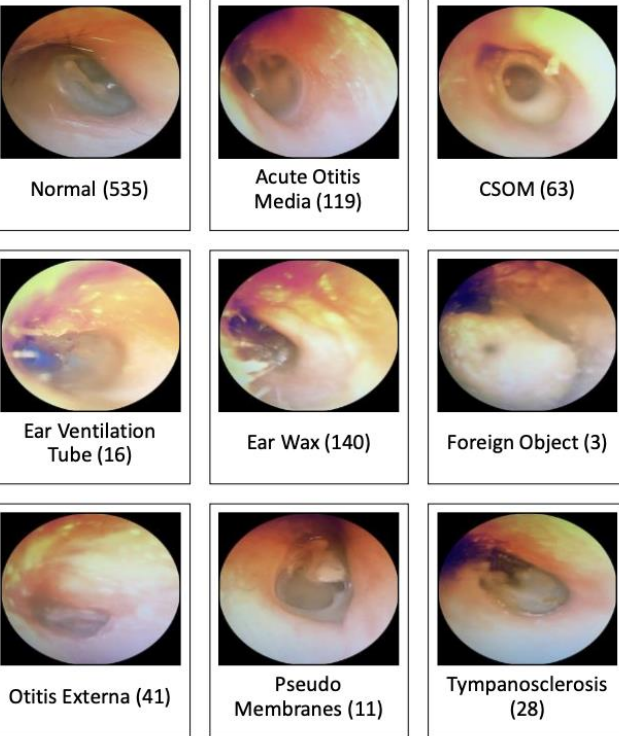


Image Class	Original Train Data	Over sampling Ratio	Updated Train Data
Acute Otitis Media	83	1	166
Chronic Otitis Media	44	1	88
Earwax Plug	98	1	196
Otitis Externa	28	3	112
Foreign Object	2	4	12
Ventilation Tube	11	3	55
Pseudo Membrane	7	5	35
Tympanosclerosis	19	3	76

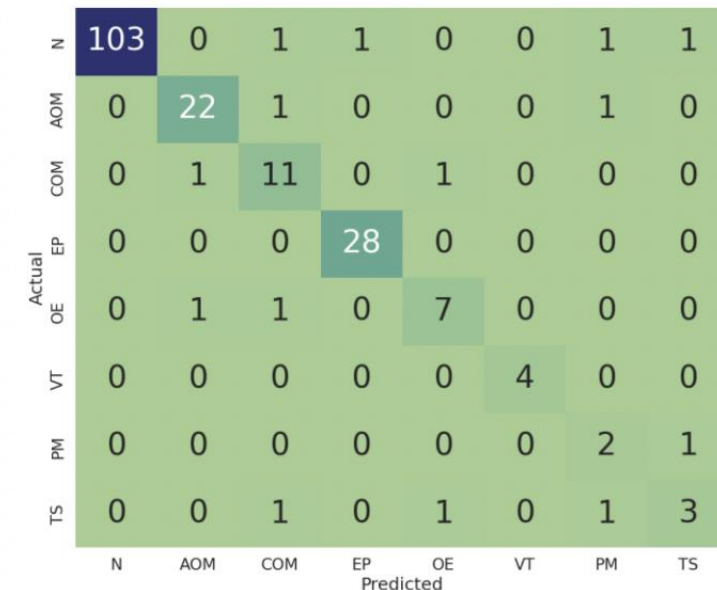
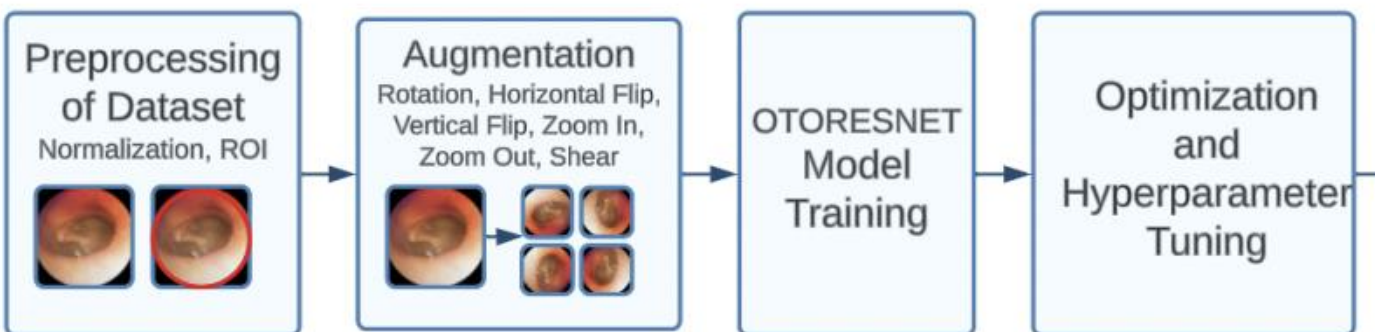


FIGURE 8. OTONet Confusion Matrix depicting the classification results for different ear conditions. [ N (Normal), AOM (Acute Otitis Media), COM (Chronic Otitis Media), EP (Earwax Plug), OE (Otitis Externa), VT (Ventilation Tube), PM (Pseudo Membrane), TS (Tympanosclerosis)]



- nedostatek snímků všech diagnóz
- nevybalancované diagnózy
- heterogenní datové sady
  - různé kliniky používají různé přístroje, zaměřují se na jiné diagnózy
  - na základě kvality obrázku lze predikovat podmnožinu diagnóz



# .csv DataChat

Agent technologies harnessing the full potential of Large Language Models (LLMs) to enhance search capabilities within structured data

## Query Titanic passenger list



Show examples

Enter your query:

What was the survival chance for passenger classes? Respond with a table. Then speculate about the reasons

106/256

Answer

### Answer:

The survival chance for the passenger classes was as follows:

PClass	Survival Chance
1st	0.599
2nd	0.426
3rd	0.194

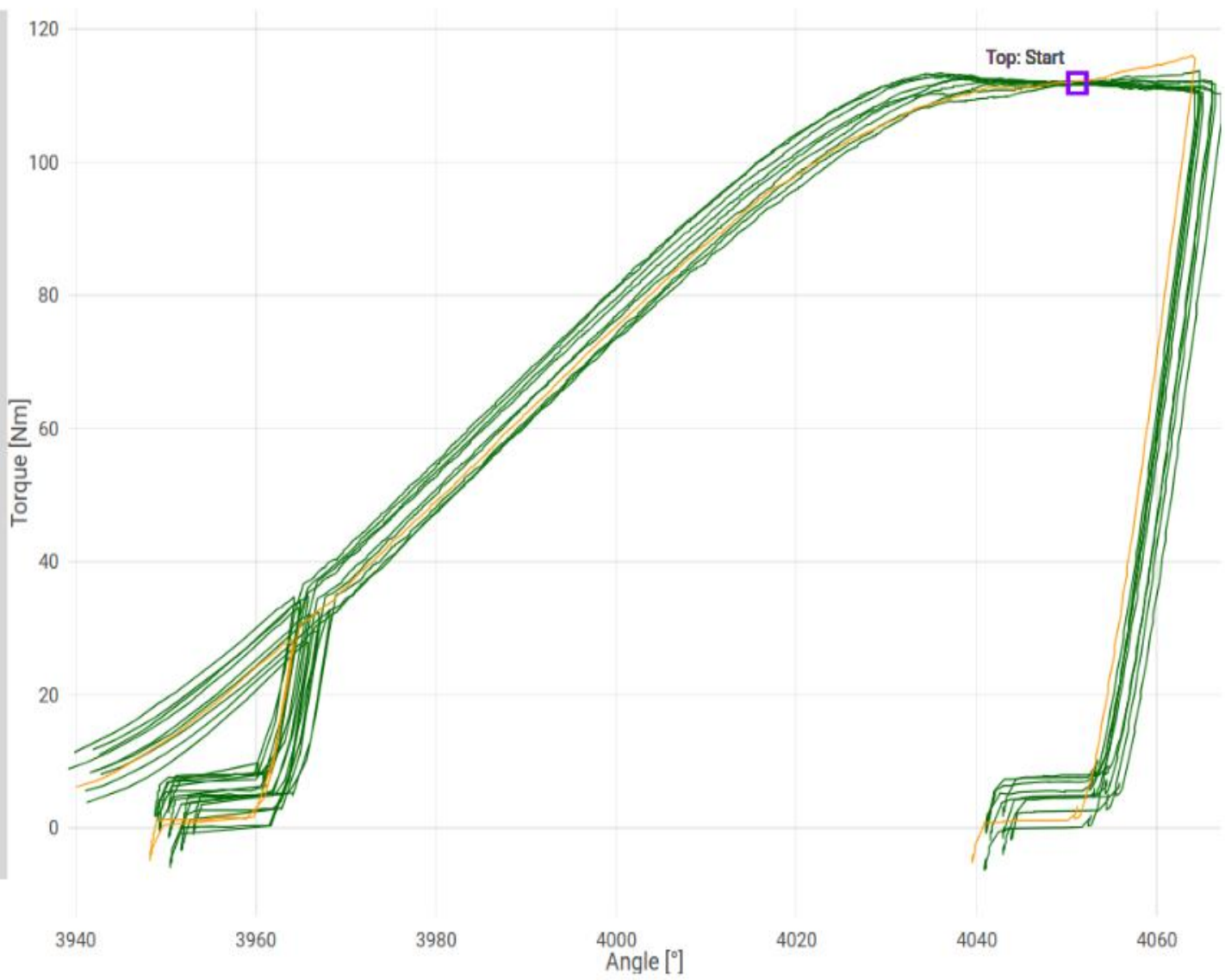
This means that passengers in the 1st class had the highest survival chance, followed by the 2nd class, and then the 3rd class.

The reasons for this could be many. One possible explanation could be that the 1st class cabins were located in the upper parts of the ship, which would have given those passengers more time to reach the lifeboats when the ship started sinking. Additionally, it's possible that the crew prioritized the safety of 1st class passengers over others due to their higher social status.

However, these are just speculations and the actual reasons could be different. To get a more accurate understanding, we would need to look at more detailed data and possibly conduct a more thorough analysis.

Part Nr. <b>0608820115</b>	Operation <b>0815</b>	Spindle <b>0.2</b>
-------------------------------	--------------------------	-----------------------

- >  OK - 1
- >  OK - 2
- >  OK - 3
- >  OK - 4
- >  OK - 5
- >  OK - 6
- >  OK - 7
- >  OK - 8
- >  OK - 9
- >  OK, ABER



MCHAT
✕

Young start to the End.

align them to top end

[21] Sure, here are the curves aligned at the start of the top region.

this is pathological, the momentum still rises instead of a plateau

Thank you for your analysis. To summarize: - We have looked at the momentum curve, specifically focusing on the part after the Eindrehen region. - We compared this curve with some of the past curves and aligned them at the start of the top region. - Based on your expertise, you have determined that this is a pathological situation since the momentum continues to rise instead of forming a plateau as expected. I will flag this anomaly for further investigation. Thanks again for your help. Goodbye!

Start typing here...
➤



<https://themama.ai>  
[mama@themama.ai](mailto:mama@themama.ai)



CHEESE

*AI-based platform and CRO services  
for drug discovery*

October 2024

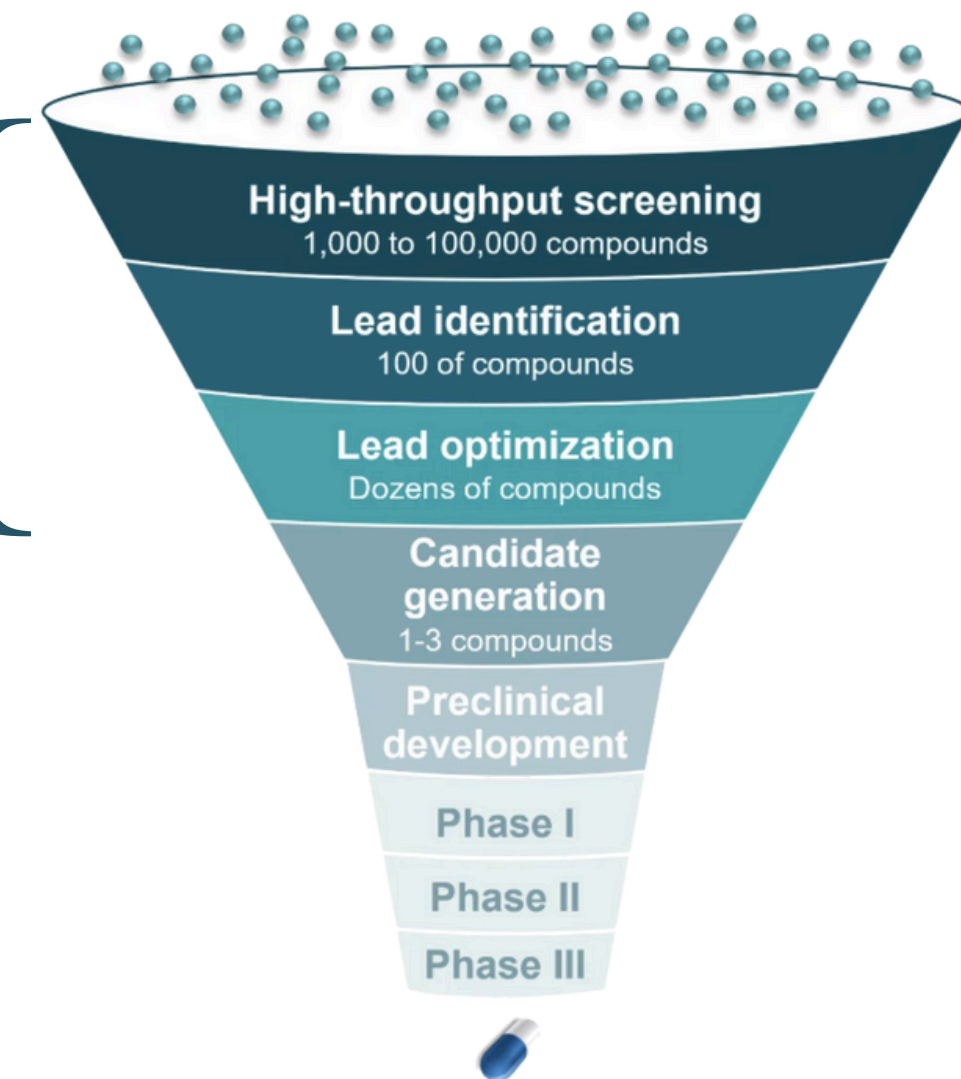
Contact: Jan Macek, CEO, Deep MedChem a.s.  
[jan.macek@deepmedchem.com](mailto:jan.macek@deepmedchem.com)

# The Problem

The field of drug discovery faces enormous challenges:

- **Vast 1T+ molecule databases**
- Missing **integrated property prediction**
- **Public and proprietary data** sources
- Inflexible **deployment options**

Deep MedChem  
CHEESE Platform



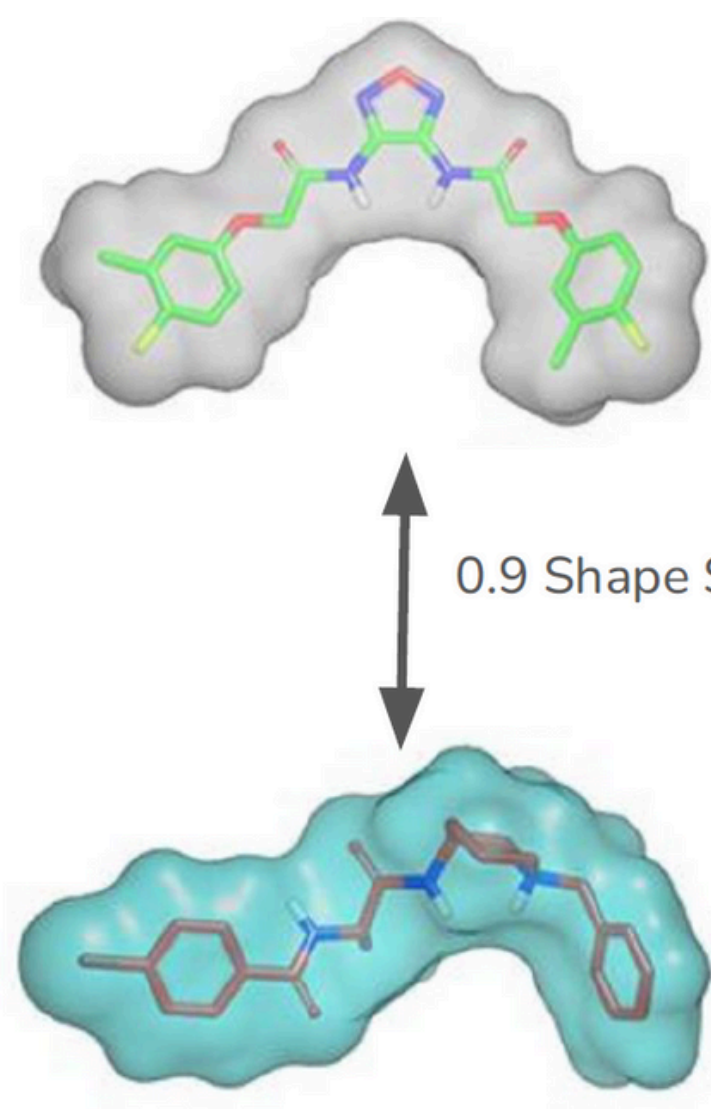
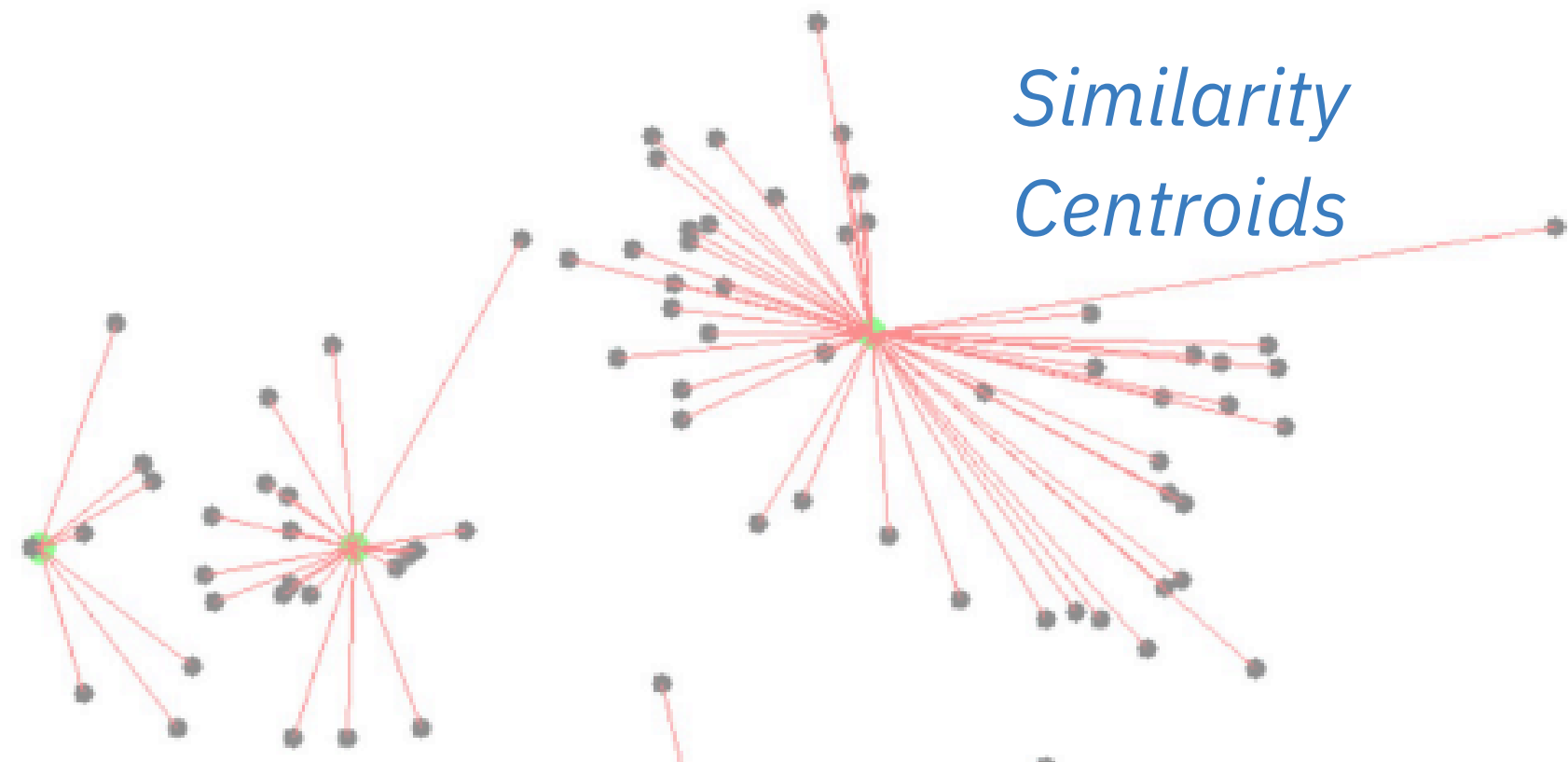
# The Solution

**CHEESE** platform fulfills all above needs and more. **CHEESE Search** is already praised by medicinal chemists and drug discovery teams.

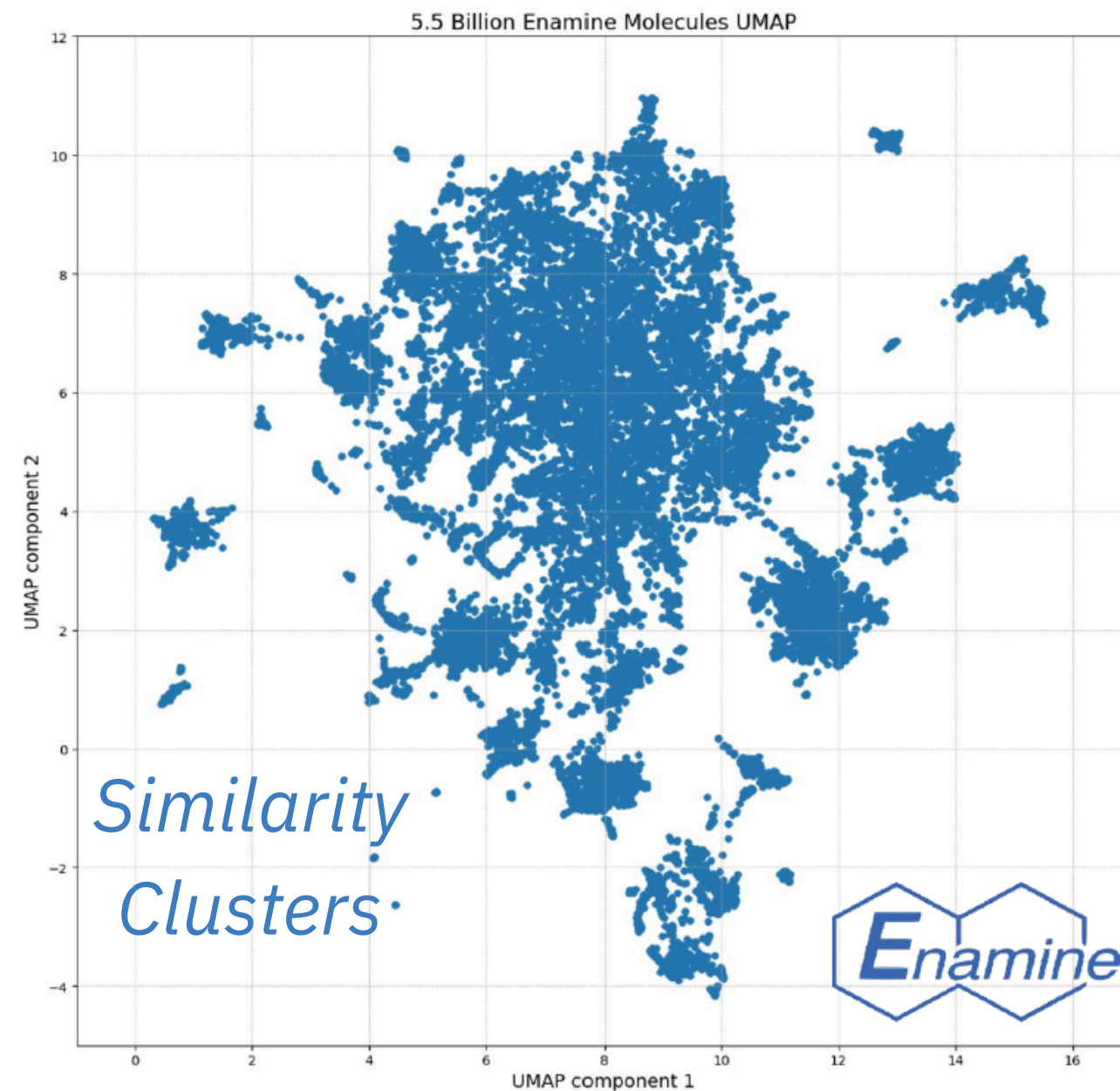
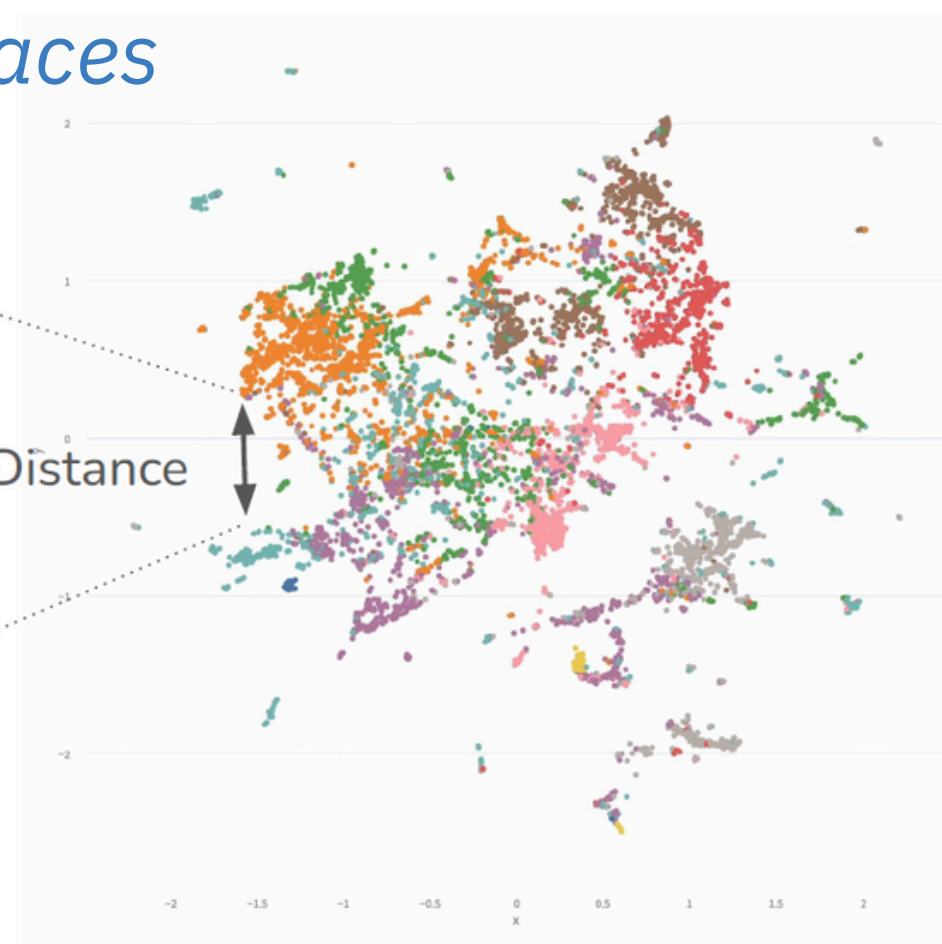


# The Solution

## CHEESE Vector Spaces



*Trained Molecular Metric Spaces*



# The Product



## Search

<https://cheese.deepmedchem.com>

## Modeller

<https://www.deepmedchem.com>

## Electrostatics

<https://electrostatics.deepmedchem.com>

## Explorer

<https://explorer.deepmedchem.com>

Search super-fast by 3D-shape and 3D-electrostatic. (incl. indexing custom molecular libraries)

Optimise (extremely fast) your hit molecules by property prediction

Model interaction with target by ESP/RESP partial charges

Explore the vast chemical space in the vicinity of your molecules

### Partners



ÚOCHB <sup>AV</sup><sub>ČR</sub>  
IOCB PRAGUE



**eMolecules**





*AI-based Molecular Similarity Search  
with 3D-Shape and 3D-Electrostatics*

## Key benefits

- Search by similarity in billion-scale databases in seconds
- Property prediction for million-scale datasets in a fraction of a second
- Universal representation able to capture arbitrary similarity metric

## Availability

- CHEESE is available as-a-service now (<https://www.deepmedchem.com/>)
- Can be deployed on private cloud and on premises
- Can be fine tuned for proprietary molecular databases







smiles    

 CONSENSUS
  2D FINGERPRINT
  3D SHAPE
  **3D ELECTROSTATIC**
 ACTIVE PAIRS

Number of results:

30

Search Database :

Select database  
ENAMINE-REAL

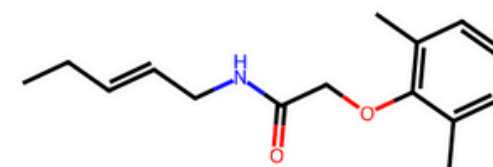
Search type:

Very accurate

Accurate

**Fast**

Molecular Weight	247.157
Formal Charge	0
N. of Heavy Atoms	18
Hydrogen Bond Acceptors	2
Hydrogen Bond Donors	1



N. of Rotatable Bonds	6
N. of Rings	1
Molar Refractivity	73.852
N. of Atoms	18
Topological Polar Surface Area	38.330

Absorption			Distribution			Metabolism			Excretion			Toxicity		
Caco2 Permeability (Log cm/s)	-4.258		Plasma Protein Binding Rate (%)	61.757		CYP2C9 Inhibition	0.004		Clearance Hepatocyte (mL/min/g)	105.135		LD50 in Rat (-Log mol/kg)	1.892	
Lipophilicity (LogD)	1.665		Volume of Distribution (L/kg)	1.615		CYP2D6 Inhibition	0.001		Clearance Microsome (mL/min/g)	34.054		AMES Mutagenicity	0.286	
Solubility (LogS)	-3.000		Blood-Brain Barrier			CYP3A4 Inhibition	0.032					Drug Induced Liver Injury	0.059	

# Benchmark: Side Chain Virtual Screening

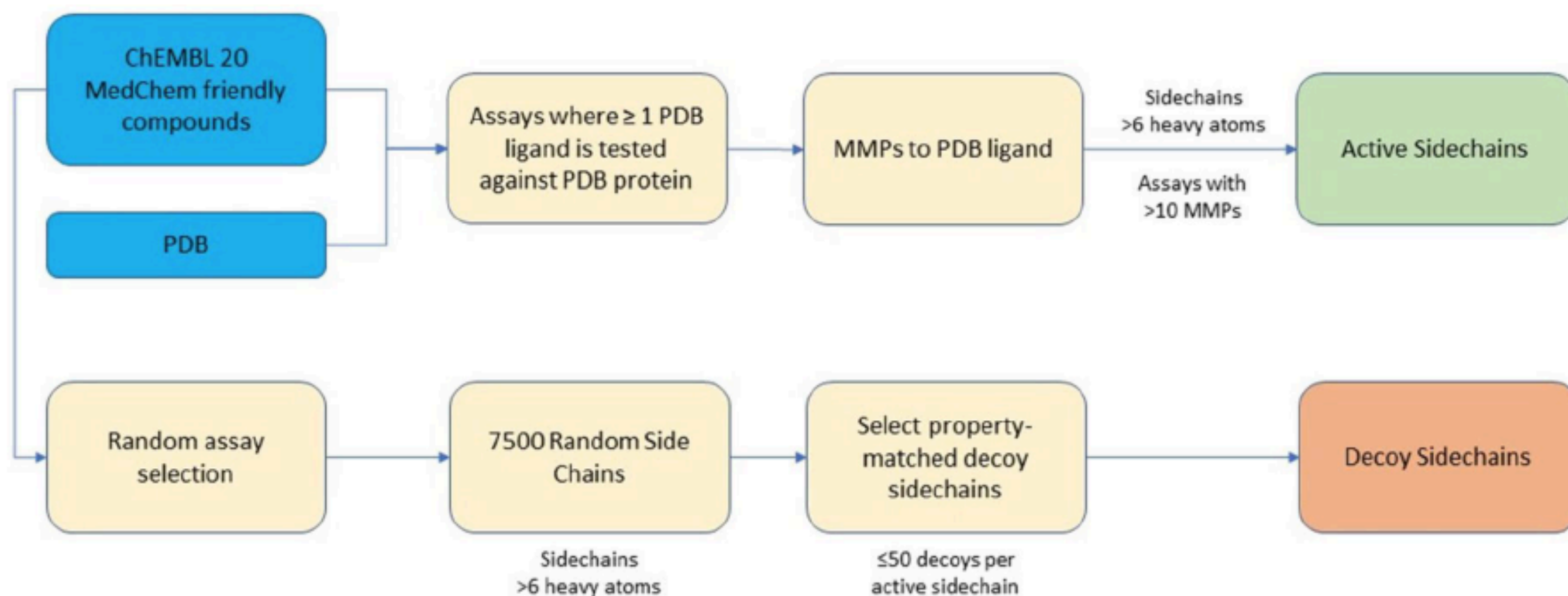
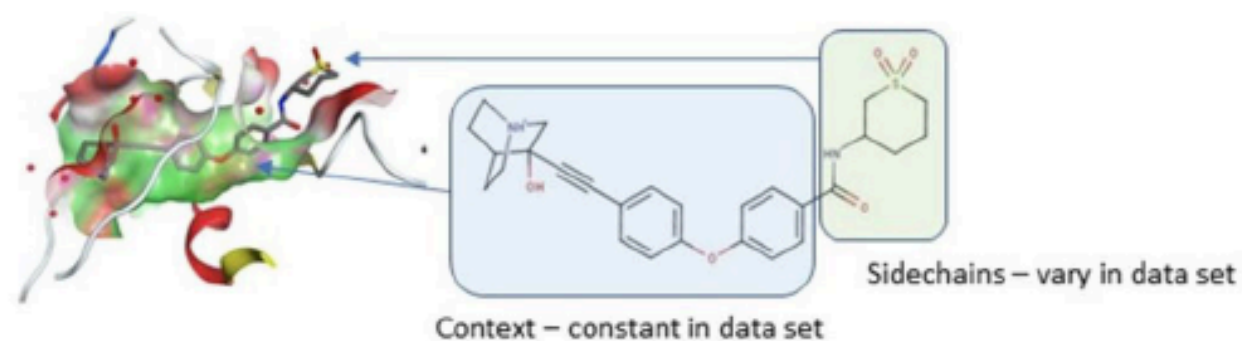


Table 1: Performance comparison of various methods based on Mean ROC area and Variance.

Method	Mean ROC area	Variance
Spark (with random conf)	0.790	0.177
Spark	0.783	0.166
Spark (with protein)	0.768	0.169
CHEESE Shapesim	0.768	0.190
CHEESE Espsim+Shapesim	0.759	0.217
CHEESE Espsim	0.755	0.218
2D Similarity (RDKit)	0.699	0.212
RDKit alignment with Smina	0.639	0.174
Glide docking	0.636	0.178
Smina docking	0.373	0.223

202 Targets

402 Assays

Baumgartner, M. P. & Evans, D. A. Side chain virtual screening of matched molecular pairs: a PDB-wide and ChEMBL-wide analysis. *J. Comput.-Aided Mol. Des.* **34**, 953–963 (2020).



# Benchmark: DUD-E

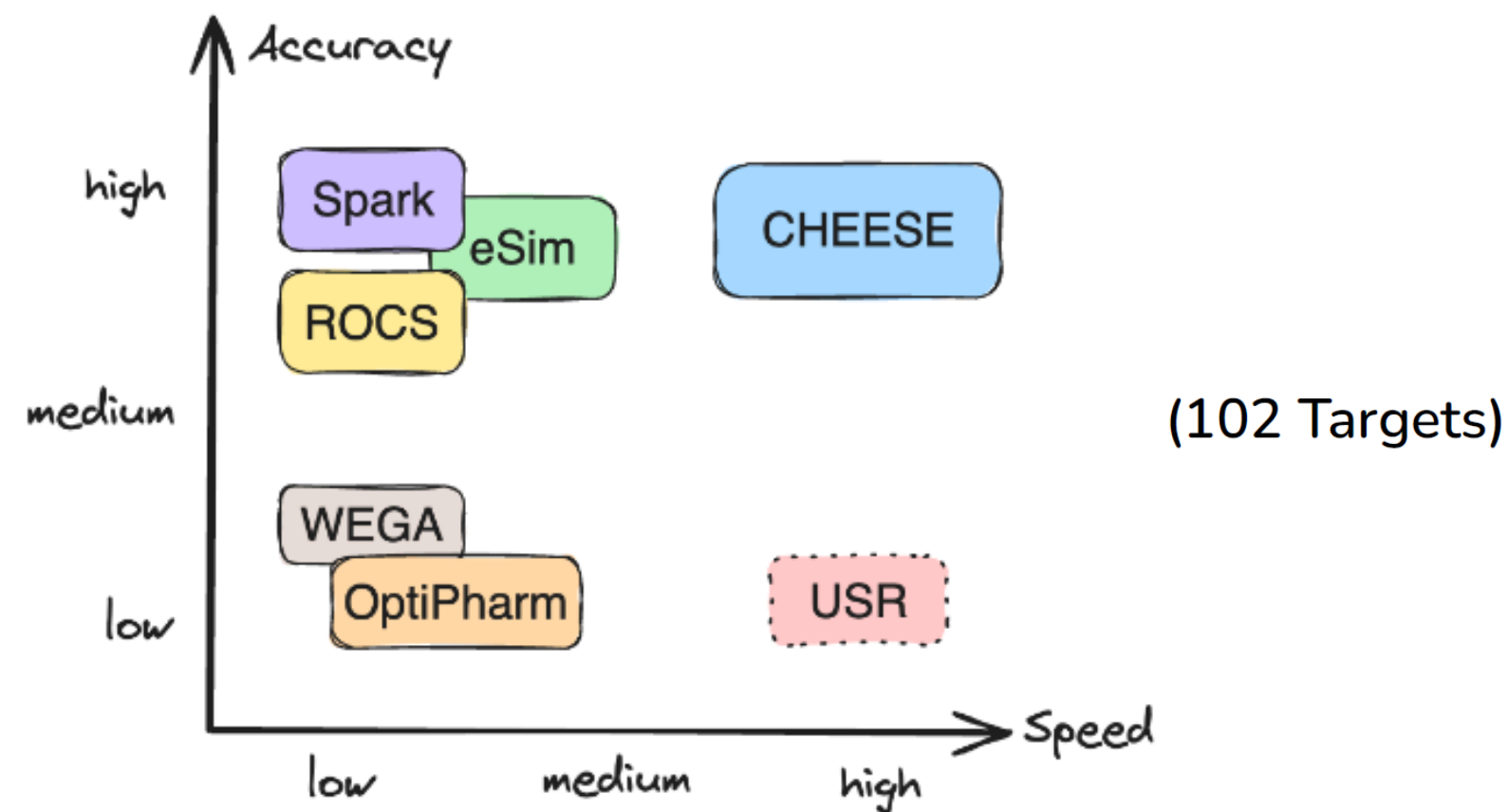
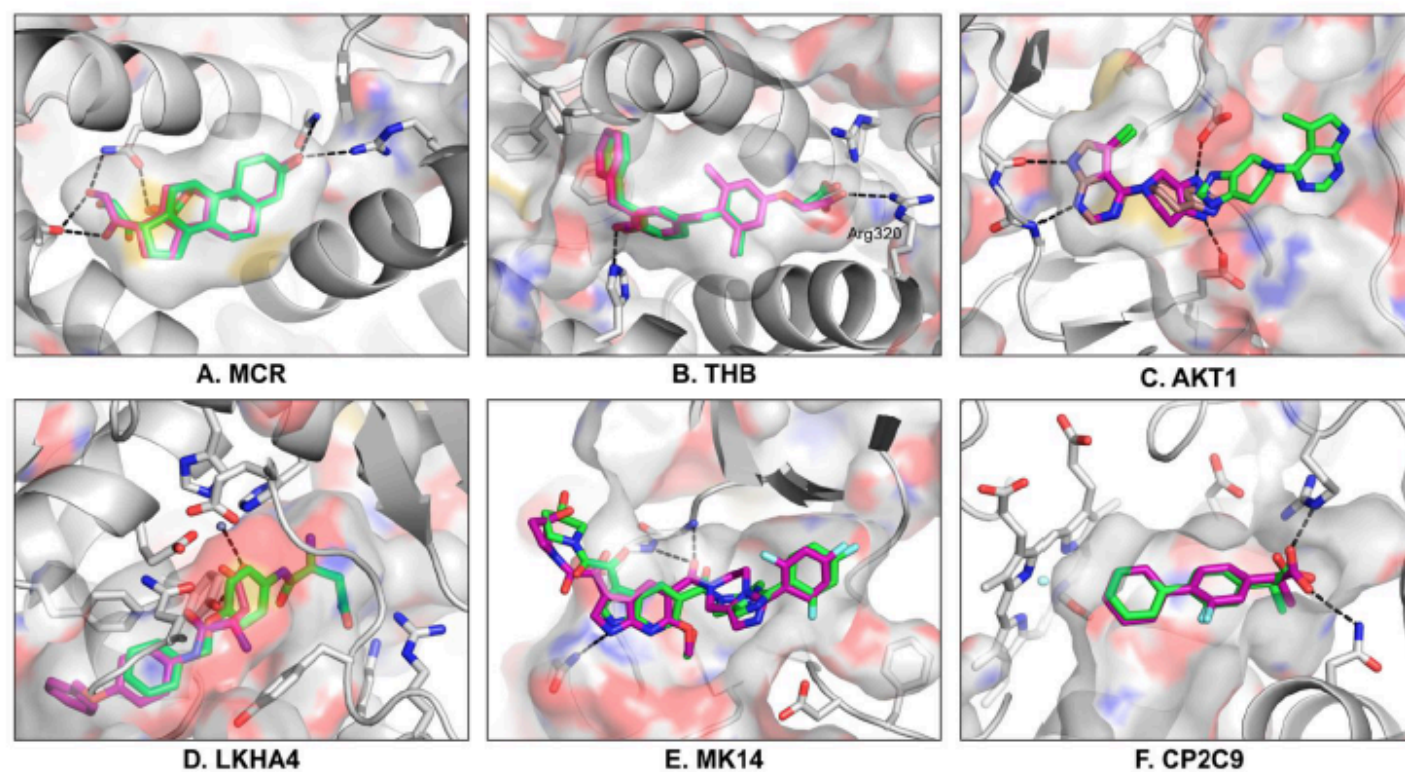
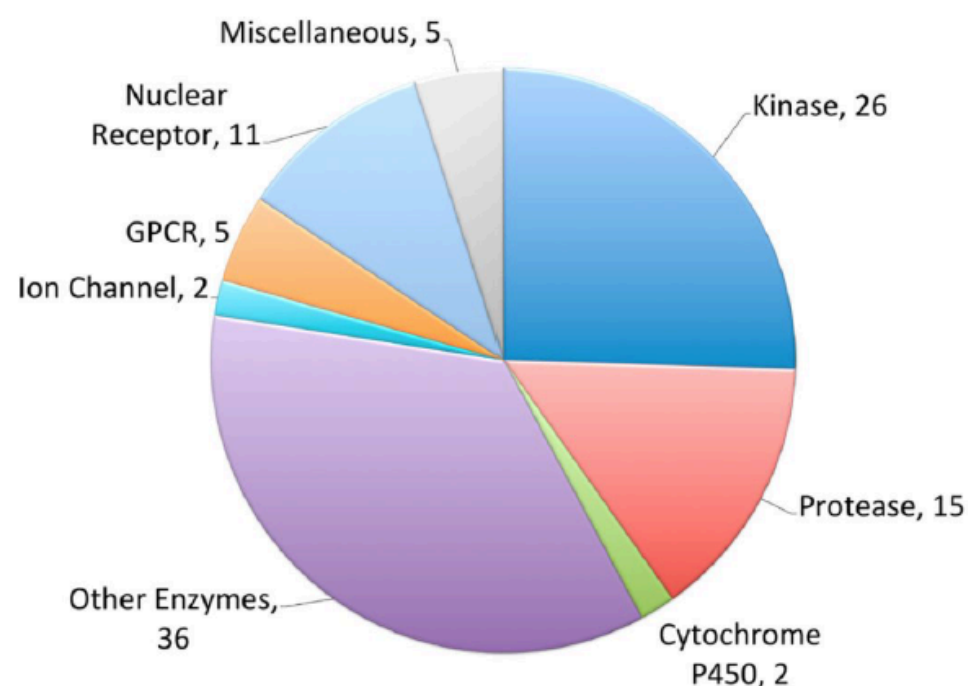
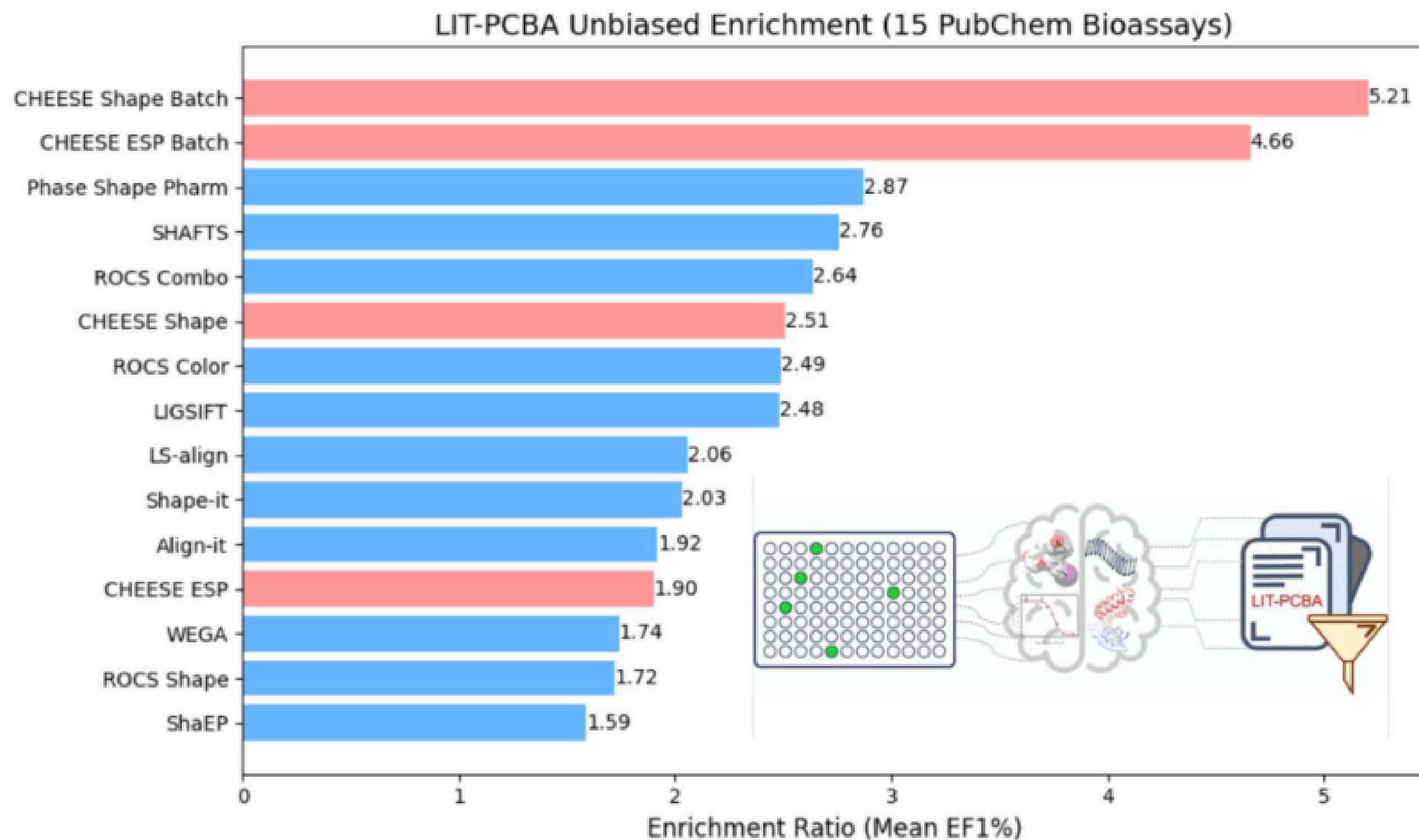


Table 1: Performance comparison of various methods. Mean ROC area and percentage of AUC values above thresholds are reported.

Method	Mean ROC	< 0.50	≥ 0.60	≥ 0.70	≥ 0.80	≥ 0.90	≥ 0.95	Time (mol/s)
CHEESE Esp+Shape	0.773	2	78	67	39	16	10	6000 *
CHEESE Espsim	0.764	3	75	65	37	17	9	6000 *
eSim-pscreen	0.755	5	81	69	43	17	8	12.3
eSim-pfast	0.736	9	82	62	34	14	5	61.2
CHEESE Shapesim	0.731	6	73	50	31	15	4	6000 *
eSim-pfastf	0.706	5	79	53	26	6	3	274.9
ROCS	0.663	18	66	44	21	9	3	50.0
ROCS (shape)	0.596	31	54	25	12	1	0	50.0
SF-Imprint	0.570	36	37	19	7	1	0	200 000.0
WEGA	0.564	31	44	19	6	0	0	26.7
OptiPharm (robust)	0.560	32	41	15	5	0	0	12.0
VAMS	0.560	36	40	14	3	0	0	5000.0
USR	0.554	35	43	20	5	1	0	6000.0
USR (shape)	0.520	43	28	13	2	0	0	6000.0

\* The processing speed for CHEESE in practice exceeds 1 billion molecules per second, thanks to the construction of vector database with molecule embeddings.

# Benchmark: CHEESE Batch Search



**LIT-PCBA: An Unbiased  
Data Set for Machine  
Learning and Virtual  
Screening**

# Customer Quotes

“

Andrea Brancale\*

**The search is extremely fast.**

“

Andrea Brancale\*

**People will definitely use it.**

“

Medicinal Chemist at BIO Int.

**I prefer this tool over the one developed by my company**

“

Medicinal Chemist at BioVaria

**The property predictions are also very useful and the prediction is surprisingly fast.**

“

**With other tools, I would go for a coffee while getting predictions.**

“

Jiri Damborsky

**AI for pharma: great tool!!!**

\*Prof. Jiri Damborsky, Masaryk University, Brno, h-index 69

\*Prof. Andrea Brancale, Univ of Chem Tech, Prague, h-index 47



**CHEESE**  
M O D E L L E R

*Unlocks value of experimental data*  
*Accelerates the pre-screening process*

## Overview

- Accelerate pre-screening process with high speed and accurate predictive models
- Run super-fast prediction of molecular properties with pre-trained ADMET and activity predictive models
- Unlock the value in your data with custom predictive models at unparalleled speed



Deep MedChem





**CHEESE**  
M O D E L L E R

*Molecular property prediction with unmatched speed (and with custom models)*

## Key Features

- Fast property prediction
- Employ custom experimental data
- Out-of-the-box pre-trained models  
(ADMET, Activity)
- Ease of use (No AI expert needed)
- Total data privacy

## Benefits

- Drastically shorter pre-screening times
- Reduced computing costs
- Lower personnel expenses
- Better lead selection and diversity
- Accelerated Time to Market
- **ROI > 300%\***

*\* based on a customer case study*



**Deep MedChem**



**CHEESE**  
MODELLER

# ADME Prediction

- Therapeutic Data Commons
- SOTA on 4 out of 22 benchmarks
- Solid Speed & Accuracy Tradeoff

Table 1: Performance comparison of SOTA (May 2023) and CHEESE Modeller on ADME benchmarks.

Dataset	Metric	SOTA	Modeller
Ames Mutagenicity [2]	AUROC ↑	0.865	0.845
Blood Brain Barrier [3]	AUROC ↑	0.923	0.839
Oral Bioavailability [4]	AUROC ↑	0.748	0.674
Volume of Distribution [5]	Spearman ↑	0.612	0.622
Hepatocyte Clearance [6]	Spearman ↑	0.491	0.490
Microsome Clearance [6]	Spearman ↑	0.625	0.444
CYP P450 2C9 Substrate [7]	AUPRC ↑	0.437	0.328
CYP P450 2C9 Inhibition [8]	AUPRC ↑	0.794	0.697
CYP P450 2D6 Substrate [7]	AUPRC ↑	0.711	0.634
CYP P450 2D6 Inhibition [8]	AUPRC ↑	0.721	0.598
CYP P450 3A4 Substrate [7]	AUPRC ↑	0.680	0.659
CYP P450 3A4 Inhibition [8]	AUPRC ↑	0.882	0.826
Drug Induced Liver Injury	AUROC ↑	0.937	0.848
Terminal Phase Half Life [9]	Spearman ↑	0.416	0.437
hERG Inhibition	AUROC ↑	0.875	0.759
Human Intestinal Absorption [10]	AUROC ↑	0.988	0.953
P-glycoprotein Inhibition [11]	AUROC ↑	0.940	0.920
Plasma Protein Binding Rate [6]	MAE ↓	8.251	10.149
Aqueous Solubility [12]	MAE ↓	0.727	1.002
Caco-2 Effective Permeability [13]	MAE ↓	0.285	0.370
Acute Toxicity LD50	MAE ↓	0.588	0.626
Lipophilicity [6]	MAE ↓	0.533	0.690



**CHEESE**  
M O D E L L E R

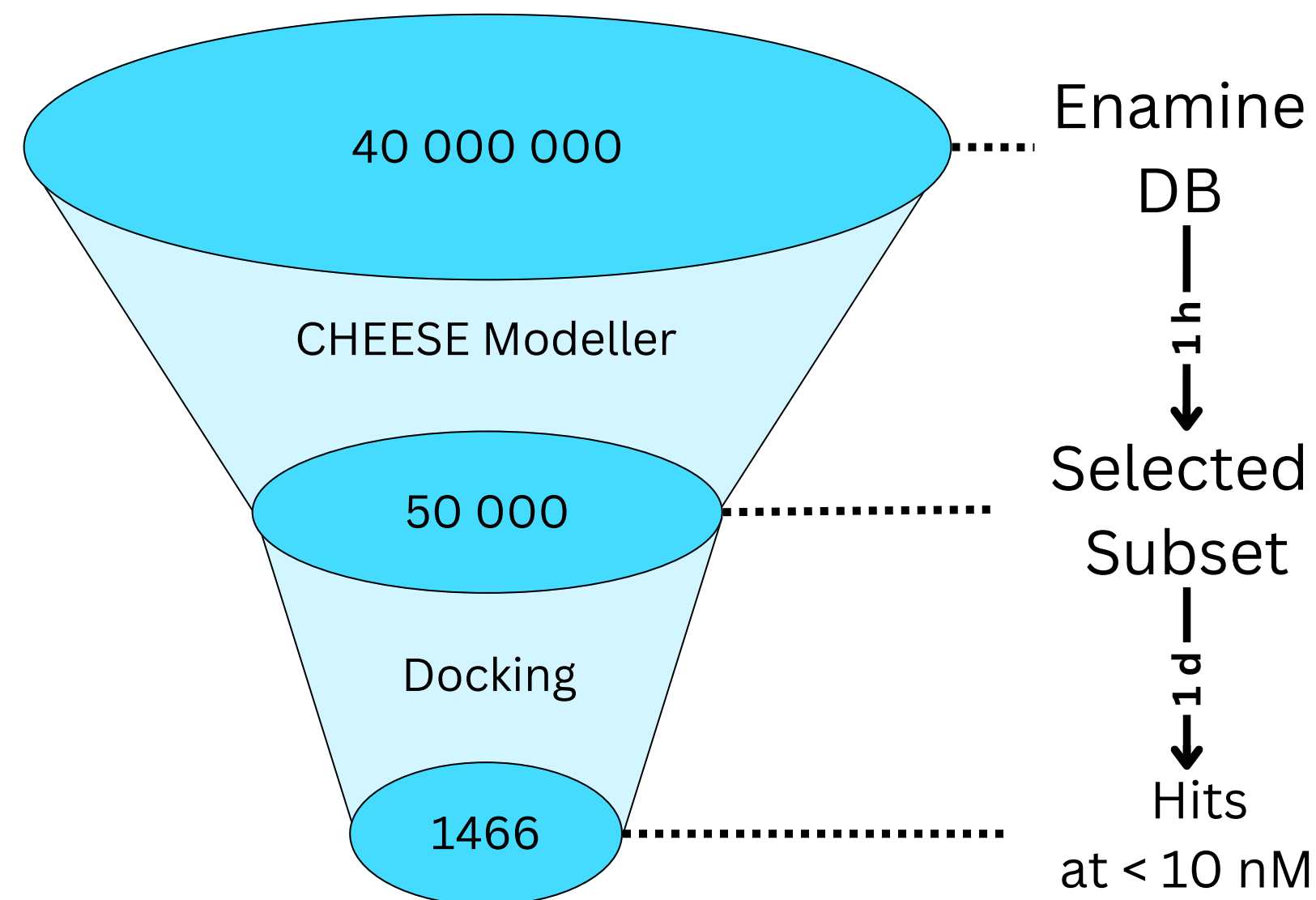
*Molecular property prediction with unmatched speed (and with custom models)*

## Case Study



Prescreening of molecules for docking to a target with partner biotech company:

- Up to **147x** improved **<10nm hit-rate** (identified binders)
- Speedup: **25 hours** vs 200 days
- **2400x Faster** than docking of 40M molecules



Deep MedChem





# CHEESE

MODEL LER

*Molecular property prediction with unmatched speed (and with custom models)*

Approach	AI Threshold	<1000nM pred. potency	<100nM pred. potency	<10 nM pred. potency
Docking of random 250K compounds out of 34M	-	<b>1.7% hit rate</b> (4185/250K)	<b>0.7% hit rate</b> (1907/250K) 228 clusters / 515 scaff. +	<b>0.02% hit rate</b> (43/250K) 22 clusters / 19 scaffolds +
Docking-AI-Docking 50K*	<1000 nM ROC area 0.96	<b>29% hit rate (17x)</b> (14K/50K)	<b>14% hit rate (20x)</b> (6.9K/50K) 116 clusters / 1276 scaff. +	<b>2.65% hit rate (133x)</b> (1327/50K) ~23% extraction^ 48 clusters / 352 scaffolds +
Docking-AI-Docking 300K*	<1000 nM ROC area 0.96	<b>17% hit rate (10x)</b> (51K/300K)	<b>7% hit rate (10x)</b> (20K/300K) 293 clusters / 3335 scaff. +	<b>1.04% hit rate (52x)</b> (3130/300K) ~54% extraction^ 109 clusters / 760 scaffolds +
Docking-AI-Docking 50K*	<100nM ROC area	<b>24% hit rate (14x)</b> (12K/50K)	<b>15% hit rate (22x)</b> (7.6K/50K) 133 clusters / 1464 scaff. +	<b>2.93% hit rate (147x)</b> (1466/50K) ~25% extraction^ 62 clusters / 410 scaffolds +
Docking-AI-Docking 300K*	<100nM ROC area	<b>13% hit rate (8x)</b> (40K/300K)	<b>7% hit rate (10x)</b> (20K/300K) 286 clusters / 3482 scaff. +	<b>1.14% hit rate (57x)</b> (3434/300K) ~59% extraction^ 132 clusters / 854 scaffolds +

\*1) Taking data above

2) CHEESE: binary approach (active/inactive based on specified threshold) and predicting activity for all 34M

3) Taking 50K or 300K top from CHEESE and docking those

^ Extraction of the 34M space - assuming equal and random distribution

(43 hits \* 34M/250K = 5848 hypothetical hits in 34M space)

\*Clustering (Data Warrior-highest similarity below 0.8)

Murco Scaffolds (DataWarrior)





**CHEESE**  
EXPLORER

## Visualisation & Clustering in Billion-scale Molecular Spaces

- Interpretable (Isometric Embeddings)
- Scalable
- 11+ Databases
- Integration with CHEESE Search & Modeller

Database	Type	Molecules
<a href="#">ChEMBL 34</a>	Enumerative	100,000
<a href="#">DrugBank 5</a>	Enumerative	8,282
<a href="#">eXplore Diverse 2024</a>	Combinatorial	100,000
<a href="#">CHEMRIYA 1.2</a>	Combinatorial	100,000
<a href="#">GDB17</a>	Enumerative	100,000
<a href="#">PubChem 2024</a>	Enumerative	99,962
<a href="#">SureChEMBL 2024</a>	Enumerative	99,754
<a href="#">ZINC22</a>	Enumerative	100,000
<a href="#">ChEBI</a>	Enumerative	99966
<a href="#">COCONUT</a>	Enumerative	99902
<a href="#">FooDB</a>	Enumerative	60569
<a href="#">Enamine Diverse 2024*</a>	Combinatorial	99,999
<a href="#">Freedom Diverse 2024*</a>	Combinatorial	100,000

\* coming soon



# CHEESE

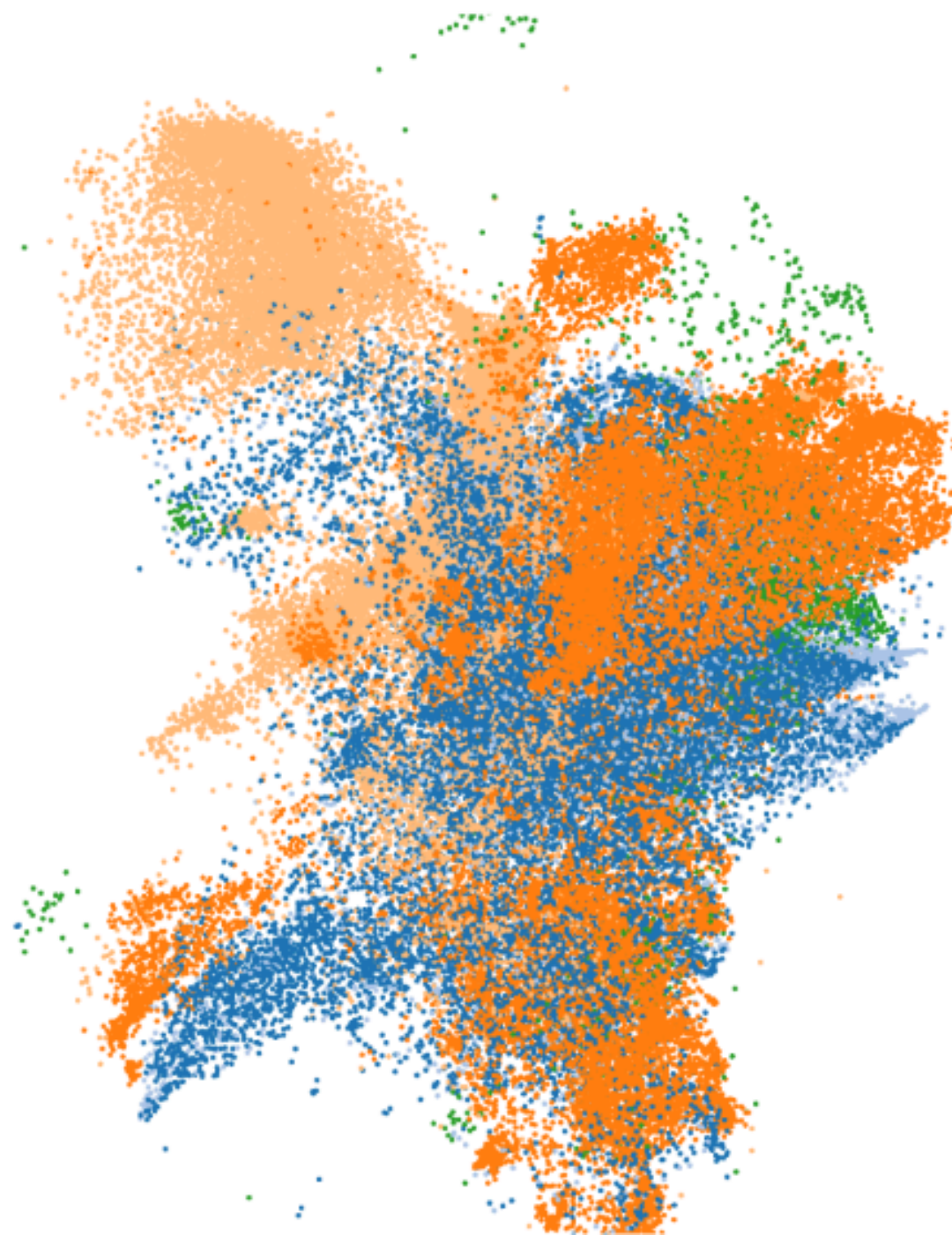
E X P L O R E R

explorer.deepmedchem.com

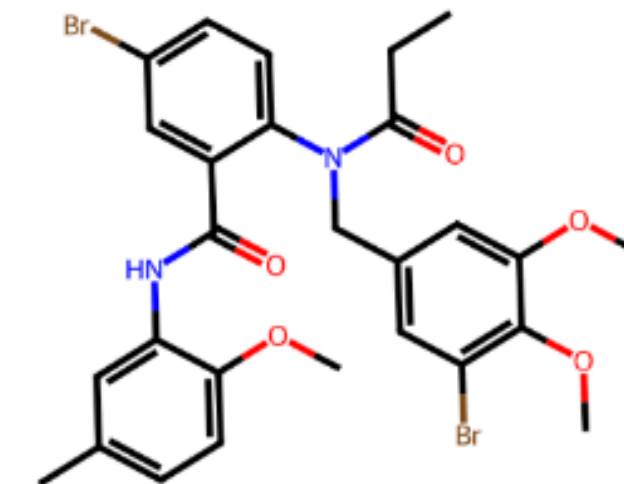
Electrostatic Shape Fingerprint Activity 2D 3D Low High Resolution Select a property Visualise Get Explorer

- ChEMBL 34
- DrugBank 5
- eXplore Diverse 2024
- CHEMRIYA 1.2
- GDB17
- PubChem 2024
- SureChEMBL 2024
- ZINC22
- ChEBI
- COCONUT
- FooDB

- 4 Representations
- 100k Points
- 2D and 3D
- Custom Datasets (premium)



Download CSV



ID:  
YAT\_R038\_part2|7014713410.0|1366672|998877  
Dataset: chemriya





# CHEESE

## ELECTROSTATICS

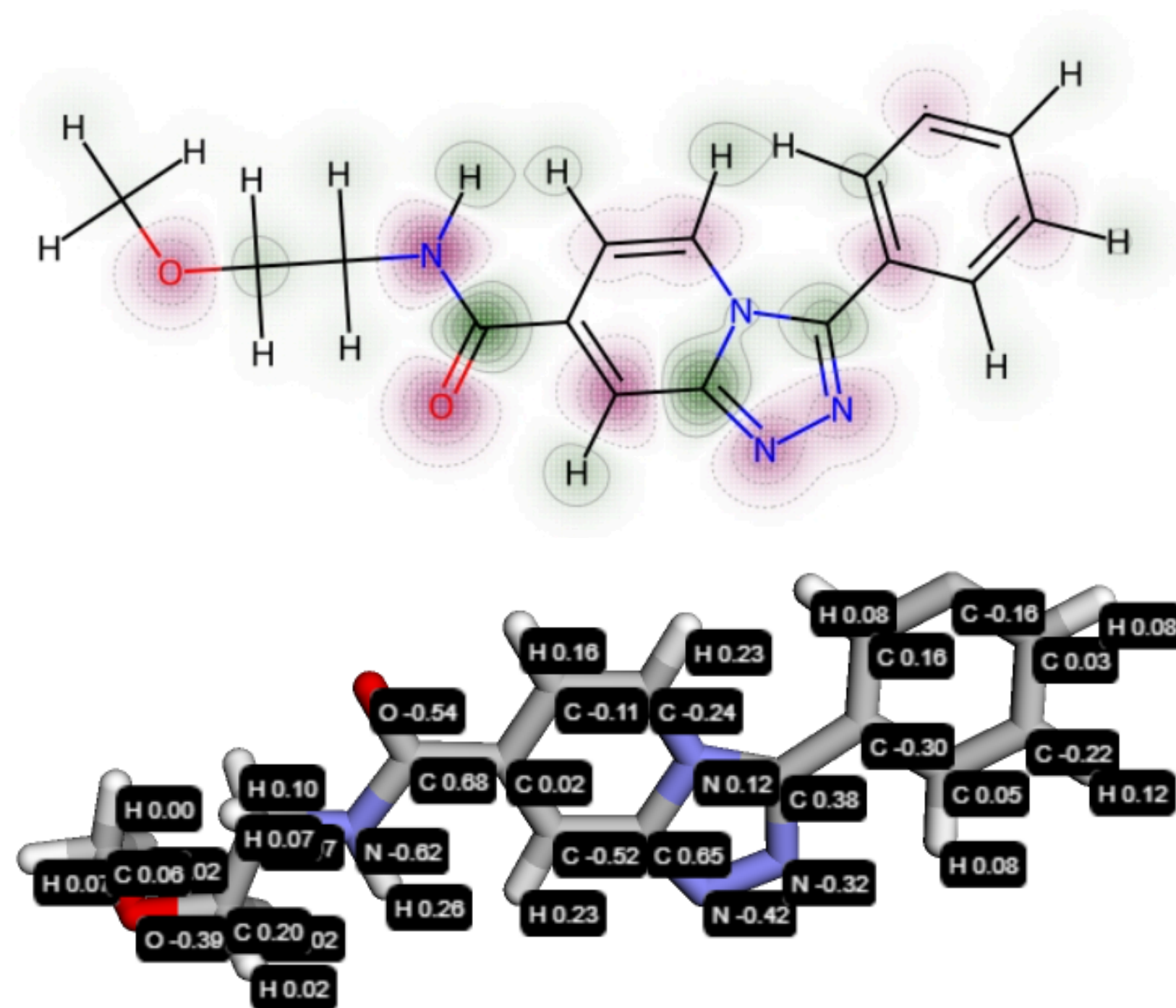


Deep MedChem

**Much faster and yet accurate estimation of molecular electrostatic potential**

- Both restrained (RESP) and unrestrained (ESP)
- Trainable AI models trained on DFT charges for difference base functions.
- **In seconds vs. minutes/hours/days** (vs. traditional DFT)
- Keeps the accuracy (0.98 correlation) and generalises well
- Available online in *BETA*

<http://electrostatics.deepmedchem.com>





# CHEESE

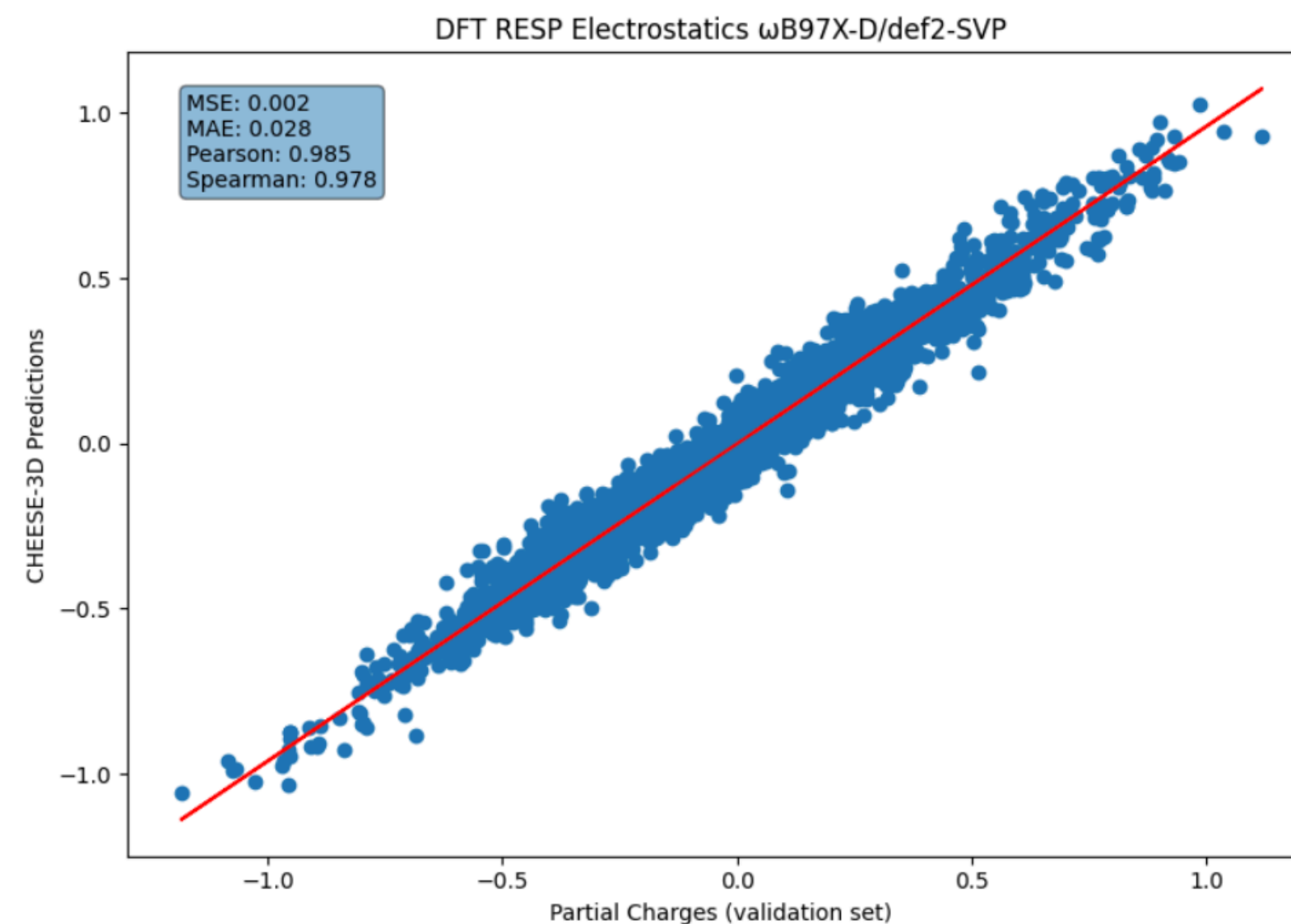
## ELECTROSTATICS



Deep MedChem

### Applications of Electrostatic Partial Charges:

- Molecular Docking
- Molecular Dynamics Simulations
- Quantum Mechanical Calculations
- Geometry Optimisation
- Electrostatic Similarity
- Electrostatic Complementarity



**Validation:** out-of-distribution scaffold-split test set (molecules with less than 0.3 Morgan Tanimoto similarity so called "Hard split").



# Great Flexibility in Collaboration Modes

## ✔ Self-Service Mode

- Independent use of CHEESE
- All benefits of the platform
- Detailed documentation
- Great tutorials
- Customized SLAs

## ✔ CRO Mode

- AI/ML Experts with DD experience
- Data analytics
- Results interpretation
- Model tuning and adaptation
- Integration to your workflows



# Great Flexibility in Deployment Options



**on**

**public cloud**



**@**

**aws**



**on premise**



**Deep MedChem**



Deep MedChem

**Thank you for your time!**

[www.deepmedchem.com](http://www.deepmedchem.com)