XXXI Symposium on
Bioinformatics and Computer-Aided
Drug Discovery
20 - 22 October, 2025 | **Oral presentation**

Ligand Pro | Skoltech
Skolkovo Institute of Science and Technology

Do-No-Harm Molecular Generation: 12-Model Benchmark and KRAS G12D Case Study

Presented by: D. Ryabchenko[1,2], podariya12@gmail.com
Joint work with: P. Gurevich[1,2], S. Kadyrov[1], O. Tinkov[1], S. Nikolenko[1], D. Frolova[1,2], A. Shapeev[1,2], M. Pak[1,2]

[1]*Ligand Pro, Moscow, Russia;*
[2]*Skolkovo Institute of Science and Technology, Moscow, Russia*

Monday, 20th October, 2025

Moscow, 2025

INTRODUCTION

FIVE-STAGE FILTERING PIPELINE

RESULTS

CONCLUSION & DISCUSSION

INTRODUCTION

FIVE-STAGE FILTERING PIPELINE

RESULTS

CONCLUSION & DISCUSSION

Introduction

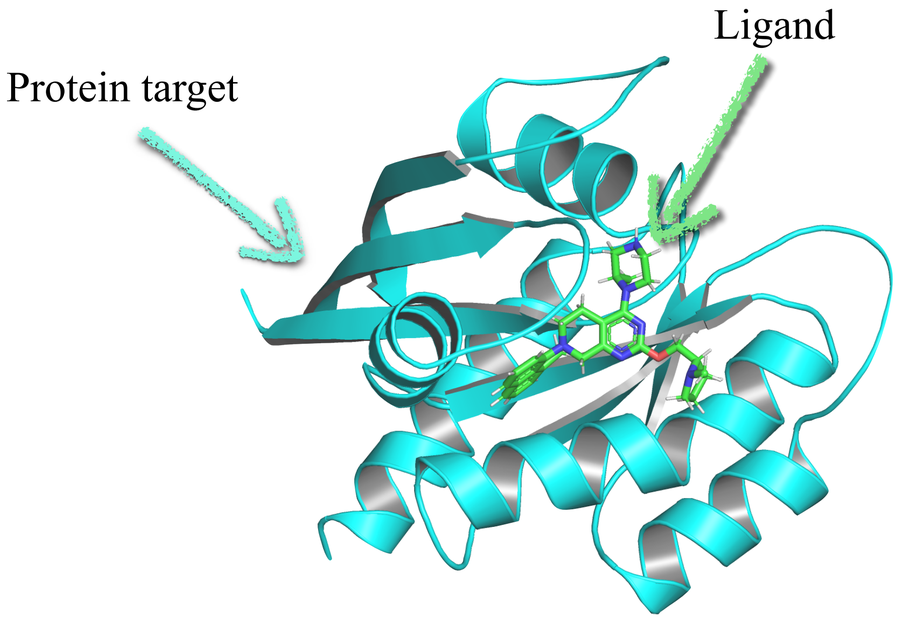
Molecule Generation

- **NECESSITY:** Accelerate the discovery of viable drug-like candidates AND save time, human, and money resources

Introduction

Molecule Generation

- **NECESSITY:** Accelerate the discovery of viable drug-like candidates AND save time, human, and money resources
- Generate a small molecule, known as *ligand*, that is
  - chemically valid, synthesizable,
    satisfying ADMET profile, binds to a target

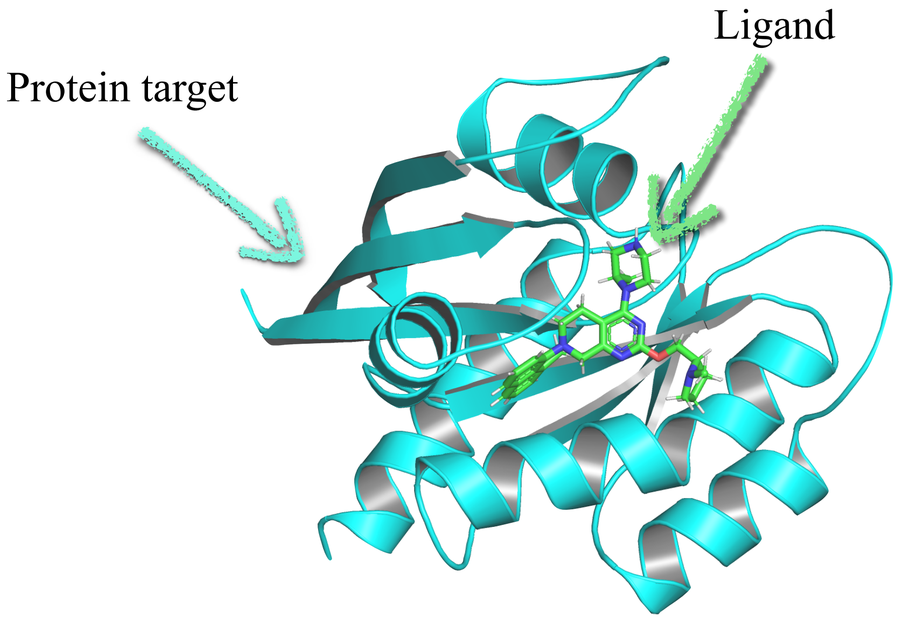


source: KRAS G12D molecule,
provided by the author, illustrated via PyMol

Introduction

Molecule Generation

- **NECESSITY:** Accelerate the discovery of viable drug-like candidates AND save time, human, and money resources
- Generate a small molecule, known as *ligand*, that is
  - chemically valid, synthesizable,
    satisfying ADMET profile, binds to a target
- **PROBLEM:** Most models show high metrics on popular generative benchmarks (GuacaMol[1], MOSES[2]), but often fail to translate into medicinally plausible compounds



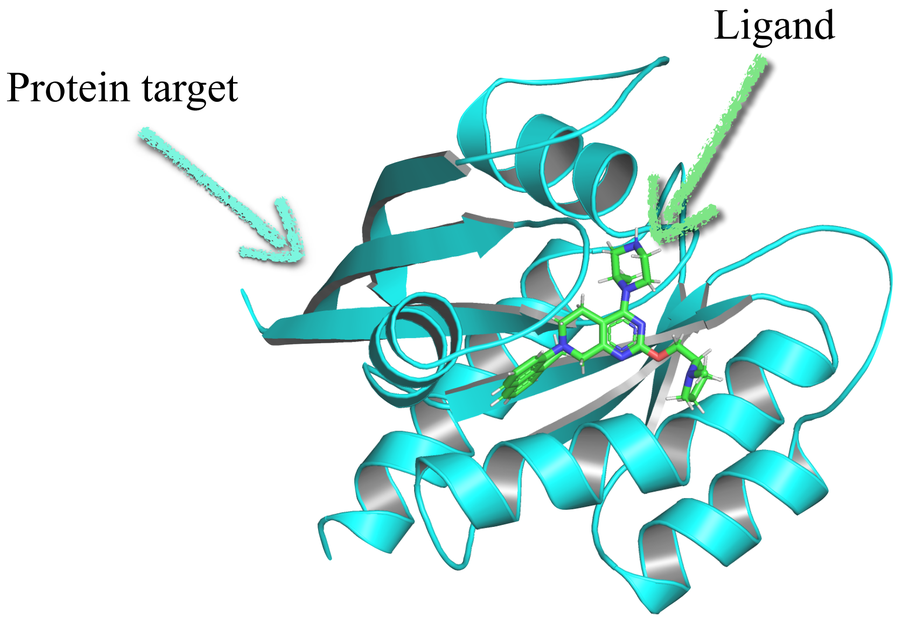
source: KRAS G12D molecule,
provided by the author, illustrated via PyMol

[1]Brown, N., Fiscato, M., Segler, M. H., & Vaucher, A. C. (2019). GuacaMol: benchmarking models for de novo molecular design. *Journal of chemical information and modeling*, *59*(3), 1096-1108.
[1]Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., ... & Zhavoronkov, A. (2020). Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in pharmacology*, *11*, 565644.

Introduction

Molecule Generation

- **NECESSITY:** Accelerate the discovery of viable drug-like candidates AND save time, human, and money resources
- Generate a small molecule, known as *ligand*, that is
  - chemically valid, synthesizable,
    satisfying ADMET profile, binds to a target
- **PROBLEM:** Most models show high metrics on popular generative benchmarks (GuacaMol[1], MOSES[2]), but often fail to translate into medicinally plausible compounds
- **SOLUTION:** Comprehensive, practice-oriented benchmark for evaluation under realistic medicinal chemistry constraints



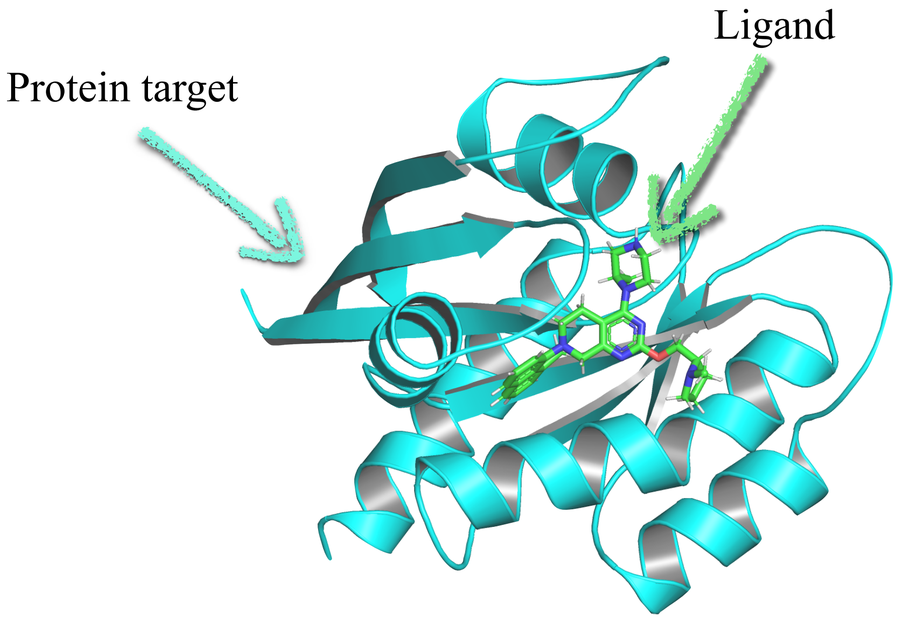
source: KRAS G12D molecule,
provided by the author, illustrated via PyMol

[1]Brown, N., Fiscato, M., Segler, M. H., & Vaucher, A. C. (2019). GuacaMol: benchmarking models for de novo molecular design. *Journal of chemical information and modeling*, *59*(3), 1096-1108.
[1]Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., ... & Zhavoronkov, A. (2020). Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in pharmacology*, *11*, 565644.

Introduction

Molecule Generation

- **NECESSITY:** Accelerate the discovery of viable drug-like candidates AND save time, human, and money resources
- Generate a small molecule, known as *ligand*, that is
  - chemically valid, synthesizable, satisfying ADMET profile, binds to a target
- **PROBLEM:** Most models show high metrics on popular generative benchmarks (GuacaMol[1], MOSES[2]), but often fail to translate into medicinally plausible compounds
- **SOLUTION:** Comprehensive, practice-oriented benchmark for evaluation under realistic medicinal chemistry constraints



source: KRAS G12D molecule,
provided by the author, illustrated via PyMol

- **PRESENTING:** Molecule generators and their performance evaluation via *Five-Stage Filtering Pipeline*. Reproducible pipelines will be released on our GitHub (https://github.com/LigandPro)

[1]Brown, N., Fiscato, M., Segler, M. H., & Vaucher, A. C. (2019). GuacaMol: benchmarking models for de novo molecular design. *Journal of chemical information and modeling*, *59*(3), 1096-1108.
[1]Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., ... & Zhavoronkov, A. (2020). Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in pharmacology*, *11*, 565644.

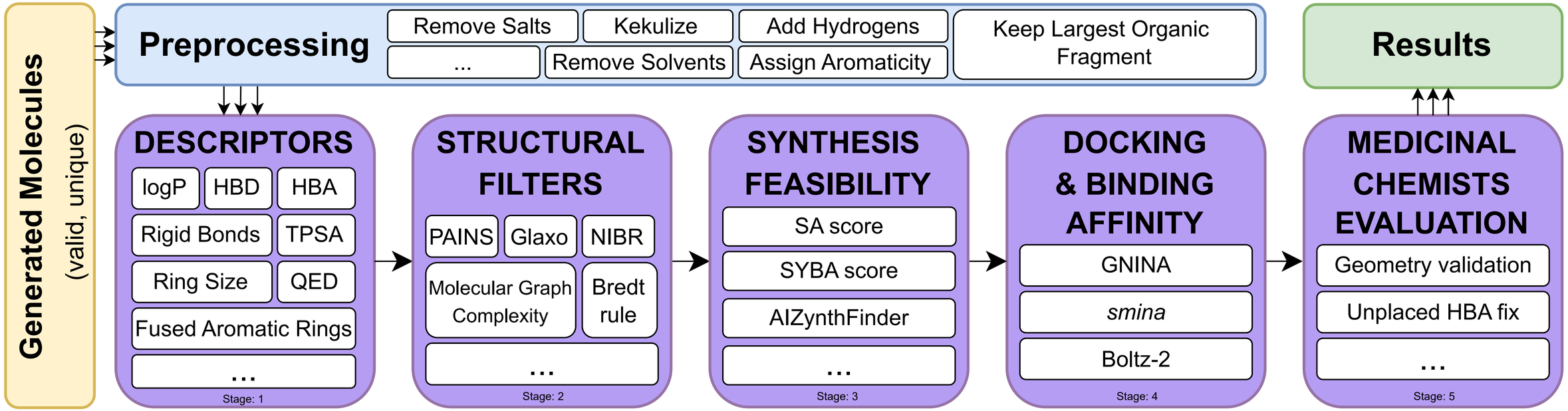
INTRODUCTION

FIVE-STAGE FILTERING PIPELINE

RESULTS

CONCLUSION & DISCUSSION

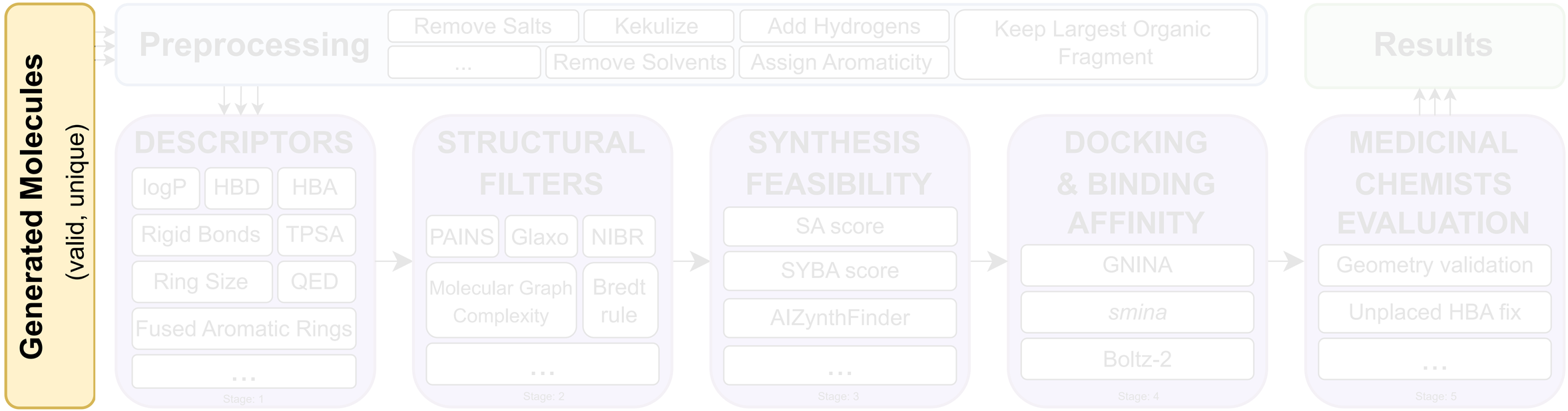
Five-Stage Filtering Pipeline



source: provided by the author

Five-Stage Filtering Pipeline

Generate Molecules

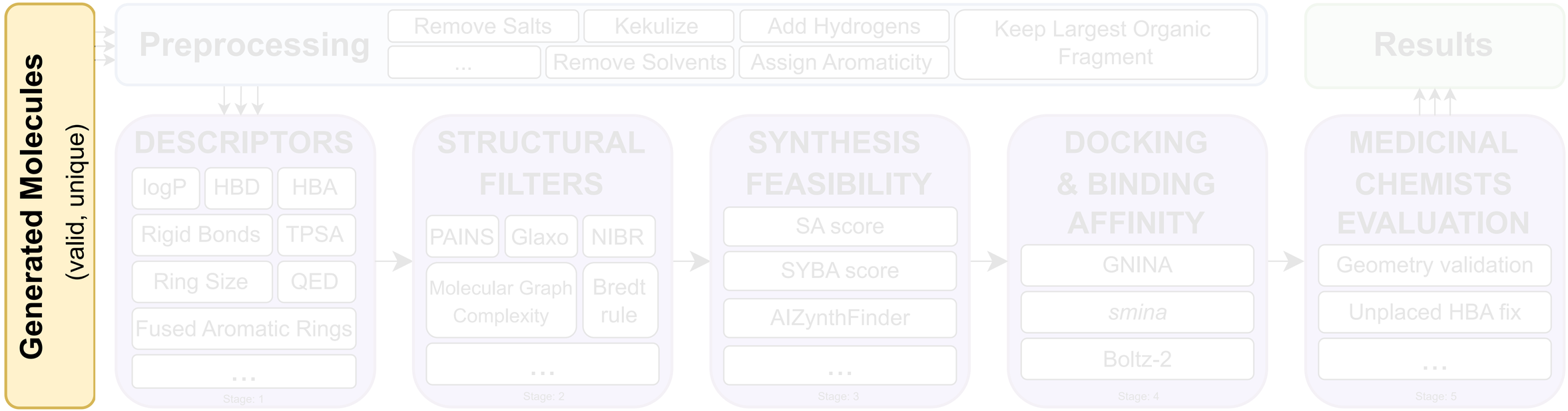


source: provided by the author

Five-Stage Filtering Pipeline

Generate Molecules

- Baseline models collection of *ligand-based,* and *protein-based* models (total 12 models).

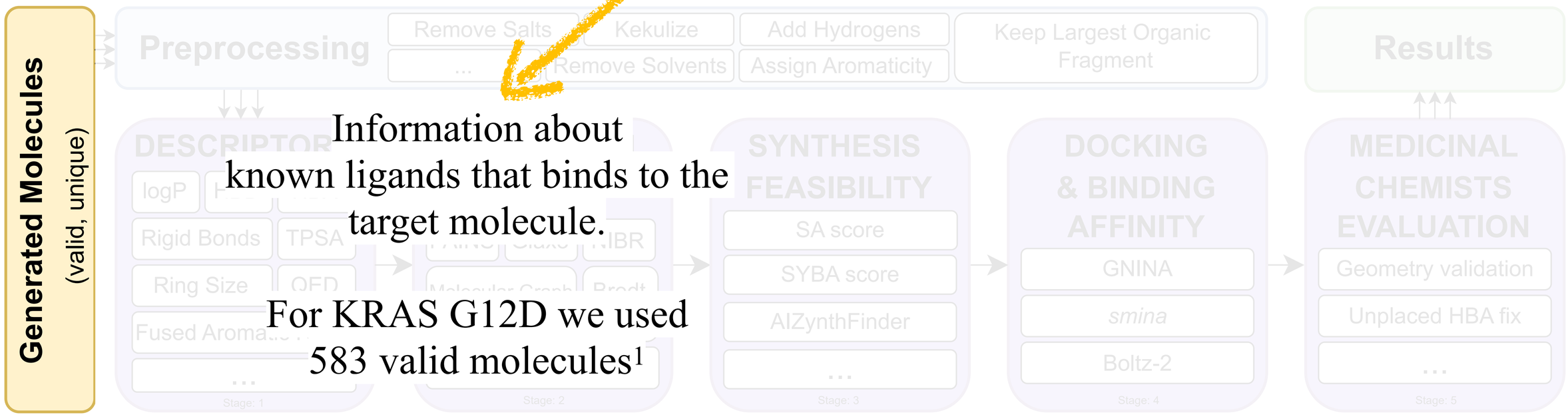


source: provided by the author

Five-Stage Filtering Pipeline

Generate Molecules

- Baseline models collection of *ligand-based,* and *protein-based* models (total 12 models).



Information about known ligands that binds to the target molecule.

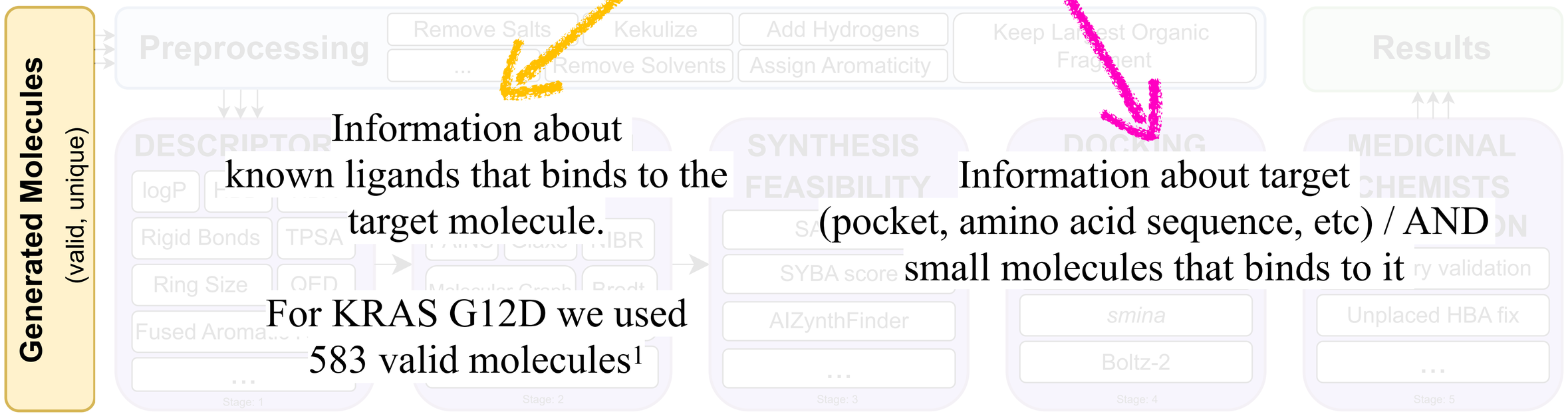
For KRAS G12D we used 583 valid molecules[1]

[1]Mohammad Ghazi Vakili, Christoph Gorgulla, Jamie Snider, AkshatKumar Nigam, Dmitry Bezrukov, Daniel Varoli, Alex Aliper, Daniil Polykovsky, Krishna M Padmanabha Das, Huel Cox Iii, et al. Quantum-computing-enhanced algorithm unveils potential kras inhibitors. *Nature Biotechnology*, pp. 1–6, 2025.

Five-Stage Filtering Pipeline

Generate Molecules

- Baseline models collection of *ligand-based,* and *protein-based* models (total 12 models).



Information about known ligands that binds to the target molecule.

For KRAS G12D we used 583 valid molecules[1]

Information about target (pocket, amino acid sequence, etc) / AND small molecules that binds to it

[1]Mohammad Ghazi Vakili, Christoph Gorgulla, Jamie Snider, AkshatKumar Nigam, Dmitry Bezrukov, Daniel Varoli, Alex Aliper, Daniil Polykovsky, Krishna M Padmanabha Das, Huel Cox Iii, et al. Quantum-computing-enhanced algorithm unveils potential kras inhibitors. *Nature Biotechnology*, pp. 1–6, 2025.

Five-Stage Filtering Pipeline

Generate Molecules

Table 1: Taxonomy of molecular generators considered in our benchmark, by model class (rows) and primary input representation (columns)

Architecture /Model Type	LIGAND-BASED	PROTEIN-BASED
Genetic Algorithm	MolFinder (Kwon & Lee, 2021)	—
Variational Autoencoder	GENTRL (Zhavoronkov et al., 2019)	—
Autoregressive	GCPG (Zou et al., 2025) PGMG (Zhu et al., 2023) REINVENT4 (Loeffler et al., 2024)	Dragonfly (Atz et al., 2024) Pocket2Mol (Peng et al., 2022) ResGen (Zhang et al., 2023)
Diffusion	—	DiffSBDD (Schneuing et al., 2024) ProtoBind-Diff (Mistryukova et al., 2025) TargetDiff (Guan et al., 2023)
Flow matching	—	DrugFlow (Schneuing et al., 2025)

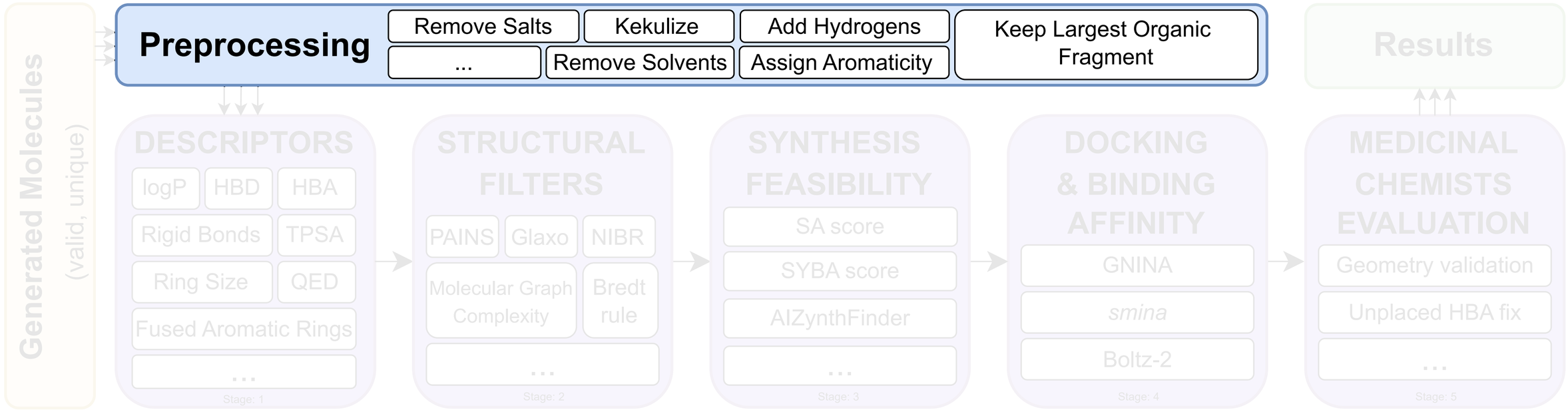
Generated Molecules (valid, unique)

Each model generates **10,000 unique and valid molecules** for further evaluation

source: provided by the author

Five-Stage Filtering Pipeline

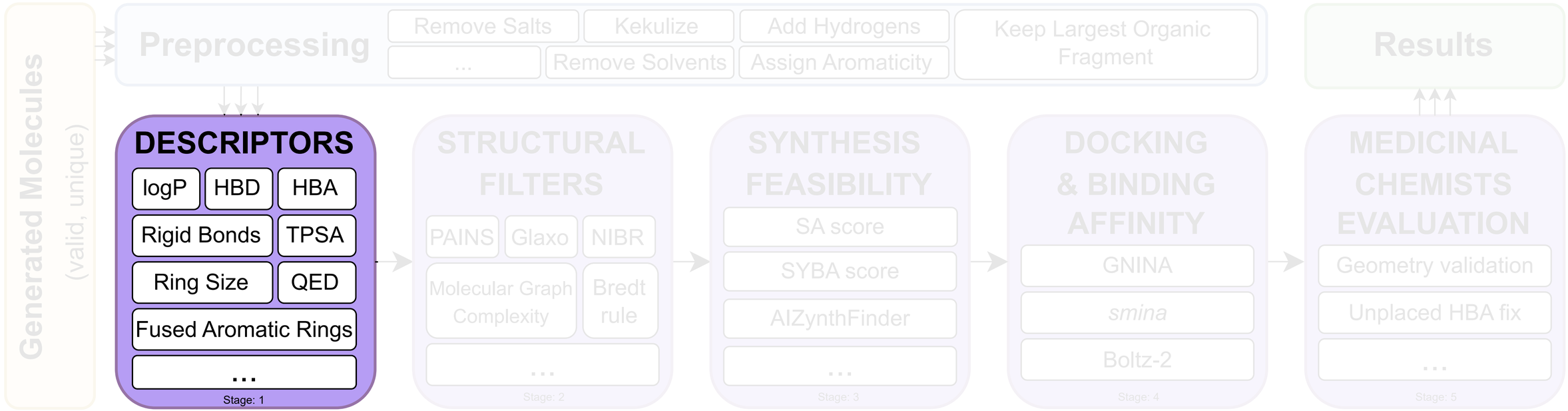
Molecules Preprocessing



source: provided by the author

Five-Stage Filtering Pipeline

Stage 1: Descriptors

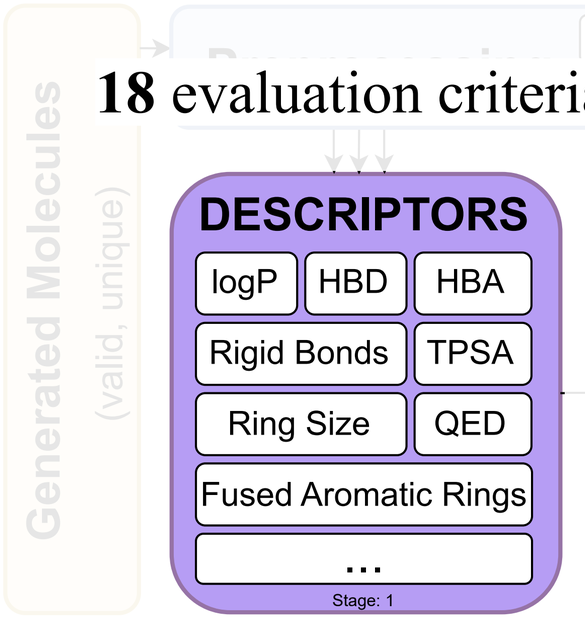


source: provided by the author

Five-Stage Filtering Pipeline

Stage 1: Descriptors

18 evaluation criteria



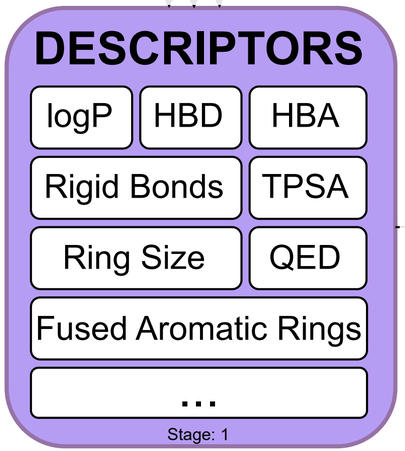
Metric	Threshold
Chars	$[C, N, S, O, F, Cl, Br, H]$
#atoms	$[10, 100]$
#heavy atoms	$[10, 50]$
#hetero atoms	$[2, 15]$
#Nitrogen atoms	$[0, 10]$
MW	$[100, 1000]$
logP	$[-2, 10]$
ring size	$[3, 12]$
#rings	$[0, 8]$
#aroma rings	$[0, 5]$
#fused aroma rings	$[0, 2]$
#rigid bonds	$[0, 30]$
#RotB	$[2, 8]$
HBD	≤ 5
HBA	$[3, 12]$
f_{sp^3}	$[0.15, 0.8]$
TPSA	$[40, 140]$
QED	$[0.3, 1]$

source: provided by the author

Five-Stage Filtering Pipeline

Stage 1: Descriptors

18 evaluation criteria



Metric	Threshold
Chars	$[C, N, S, O, F, Cl, Br, H]$
#atoms	$[10, 100]$
#heavy atoms	$[10, 50]$
#hetero atoms	$[2, 15]$
#Nitrogen atoms	$[0, 10]$
MW	$[100, 1000]$
logP	$[-2, 10]$
ring size	$[3, 12]$
#rings	$[0, 8]$
#aroma rings	$[0, 5]$
#fused aroma rings	$[0, 2]$
#rigid bonds	$[0, 30]$
#RotB	$[2, 8]$
HBD	≤ 5
HBA	$[3, 12]$
f_{sp^3}	$[0.15, 0.8]$
TPSA	$[40, 140]$
QED	$[0.3, 1]$

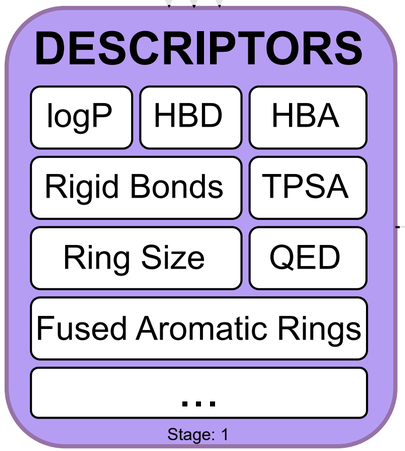
A molecule should pass all physicochemical descriptors **simultaneously** to pass the 1st filtering stage

source: provided by the author

Five-Stage Filtering Pipeline

Stage 1: Descriptors

18 evaluation criteria



Metric	Threshold
Chars	$[C, N, S, O, F, Cl, Br, H]$
#atoms	$[10, 100]$
#heavy atoms	$[10, 50]$
#hetero atoms	$[2, 15]$
#Nitrogen atoms	$[0, 10]$
MW	$[100, 1000]$
logP	$[-2, 10]$
ring size	$[3, 12]$
#rings	$[0, 8]$
#aroma rings	$[0, 5]$
#fused aroma rings	$[0, 2]$
#rigid bonds	$[0, 30]$
#RotB	$[2, 8]$
HBD	≤ 5
HBA	$[3, 12]$
f_{sp^3}	$[0.15, 0.8]$
TPSA	$[40, 140]$
QED	$[0.3, 1]$

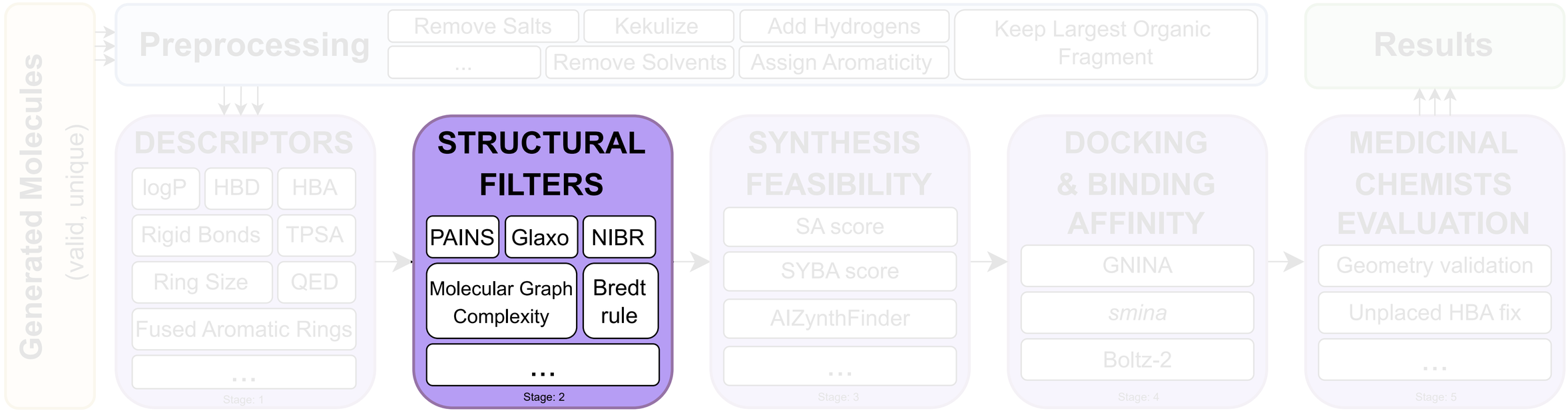
source: provided by the author

A molecule should pass all physicochemical descriptors **simultaneously** to pass the 1st filtering stage

Molecules passed *Stage 1* go to the *Stage 2*

Five-Stage Filtering Pipeline

Stage 2: Structural Filters



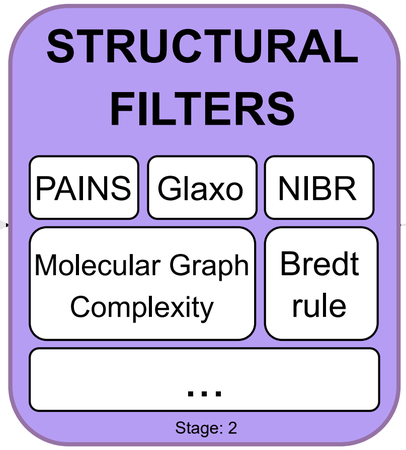
source: provided by the author

Five-Stage Filtering Pipeline

Stage 2: Structural Filters

5 rule sets and
~2500 SMARTS patterns

Rule set
Common Alerts
MolGraph statistics
MolComplexity
NIBR
Bredt



source: provided by the author

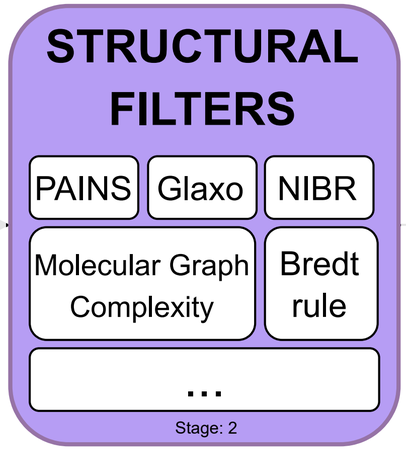
Five-Stage Filtering Pipeline

Stage 2: Structural Filters

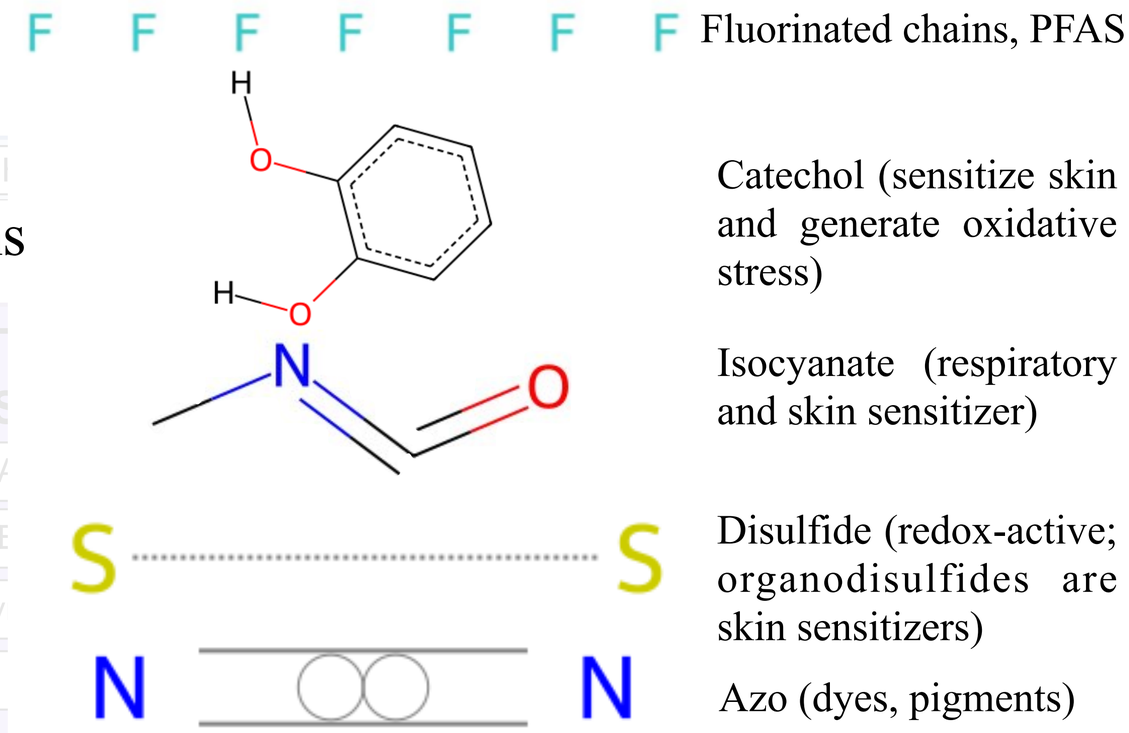
5 rule sets and
~2500 SMARTS patterns

Rule set

Common Alerts
MolGraph statistics
MolComplexity
NIBR
Bredt



Toxic or reactive chemical substructures



Fluorinated chains, PFAS

Catechol (sensitize skin and generate oxidative stress)

Isocyanate (respiratory and skin sensitizer)

Disulfide (redox-active; organodisulfides are skin sensitizers)

Azo (dyes, pigments)

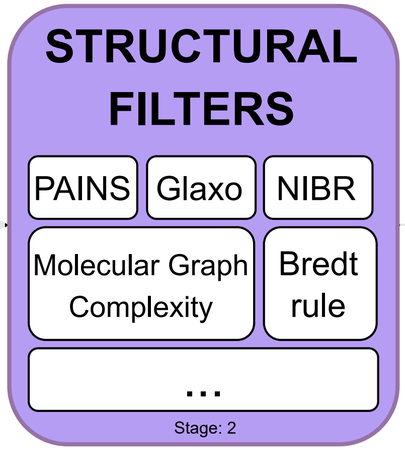
source: provided by the author

Five-Stage Filtering Pipeline

Stage 2: Structural Filters

5 rule sets and
~2500 SMARTS patterns

Rule set
Common Alerts
MolGraph statistics
MolComplexity
NIBR
Bredt



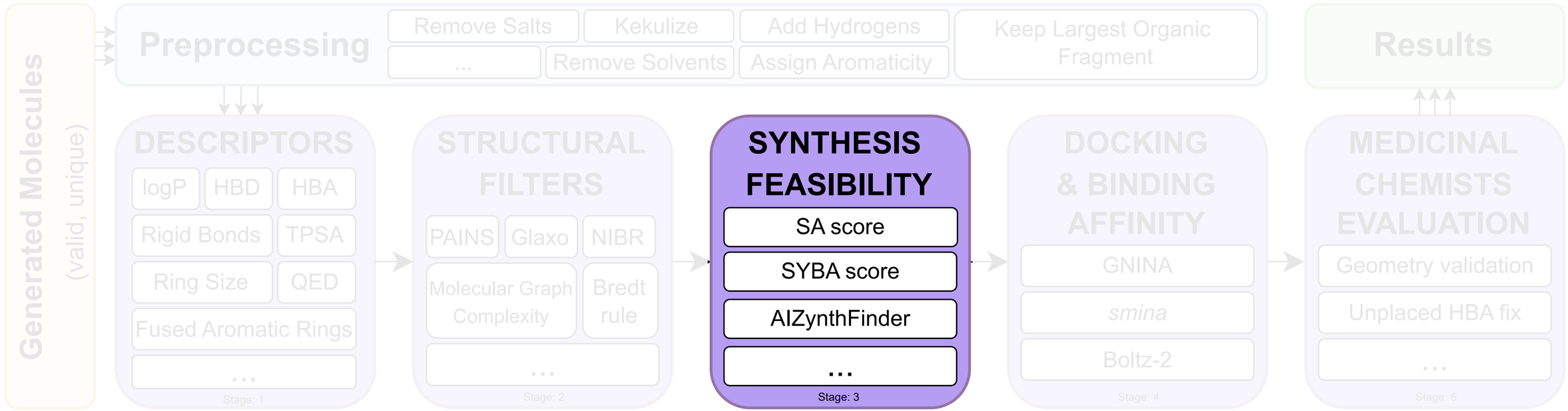
A molecule should not contain **ANY** of prohibited SMARTS patterns to pass the 2nd filtering stage

source: provided by the author

Molecules passed *Stage 2* go to the *Stage 3*

Five-Stage Filtering Pipeline

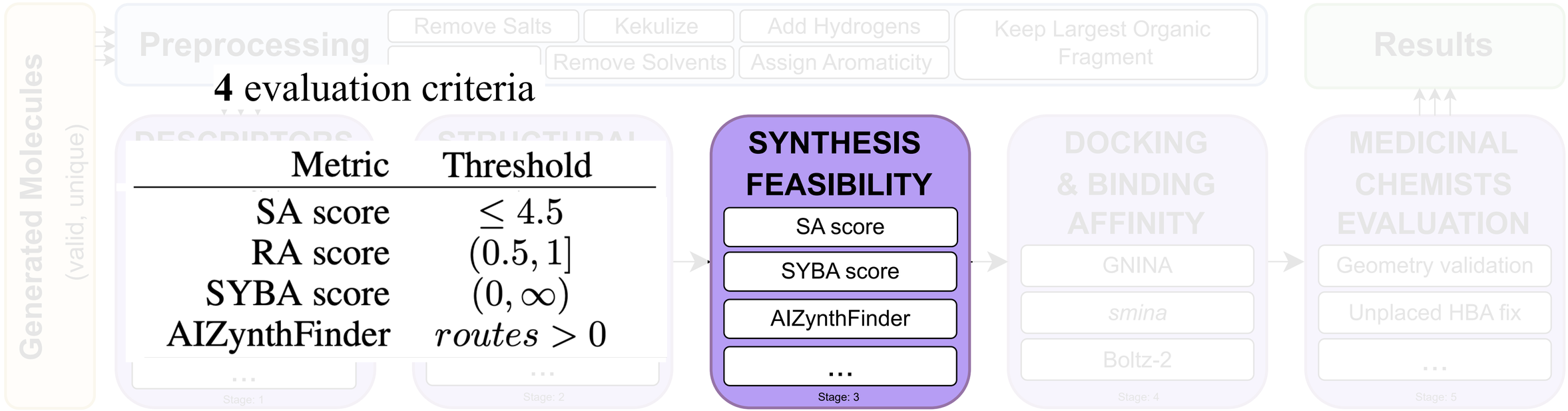
Stage 3: Synthesis Feasibility



source: provided by the author

Five-Stage Filtering Pipeline

Stage 3: Synthesis Feasibility



source: provided by the author

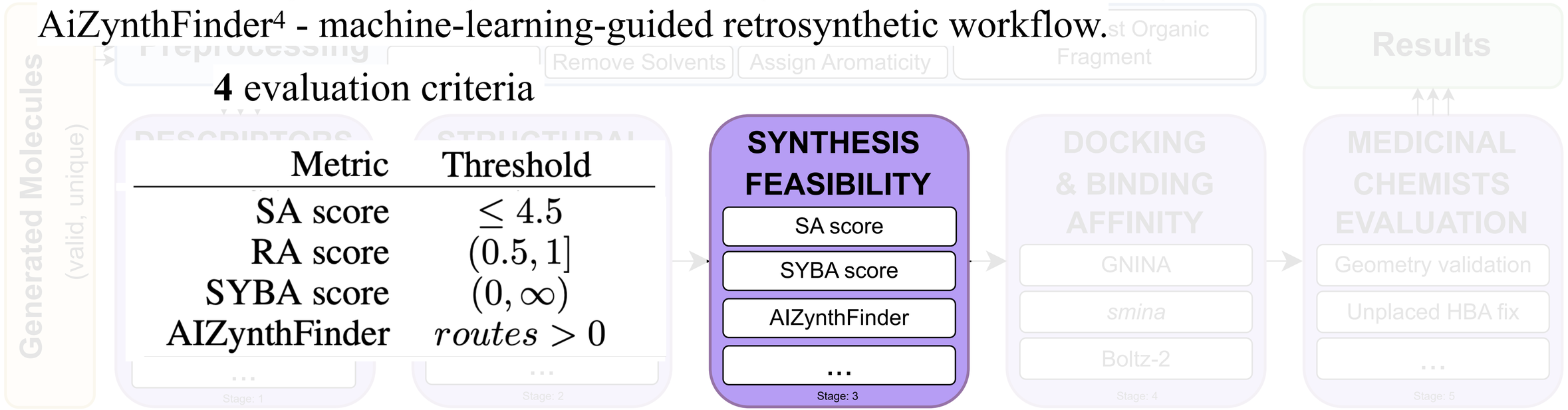
Five-Stage Filtering Pipeline

Stage 3: Synthesis Feasibility

SA score - Synthetic Accessibility score[1] - synthetic complexity score from 1 (easy) to 10 (hard).
RA score - Retrosynthetic Accessibility score[2] - probability of being a synthetic path for a compound.
SYBA score - Synthetic Bayesian Accessibility score[3] - classifier as easy or hard to synthesize.
AiZynthFinder[4] - machine-learning-guided retrosynthetic workflow.

4 evaluation criteria

Metric	Threshold
SA score	≤ 4.5
RA score	$(0.5, 1]$
SYBA score	$(0, \infty)$
AIZynthFinder	$routes > 0$



[1]Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1(1):8, 2009.
[2]Amol Thakkar, Veronika Chadimova´, Esben Jannik Bjerrum, Ola Engkvist, and Jean-Louis Reymond. Retrosynthetic accessibility score (rascore)–rapid machine learned synthesizability classification from ai driven retrosynthetic planning. *Chemical science*, 12(9):3339–3349, 2021.
[3]Milan Vors˘ila´k, Michal Kola´˘r, Ivan C˘melo, and Daniel Svozil. Syba: Bayesian estimation of synthetic accessibility of organic compounds. *Journal of cheminformatics*, 12(1):35, 2020.
[4]Samuel Genheden, Amol Thakkar, Veronika Chadimova´, Jean-Louis Reymond, Ola Engkvist, and Esben Bjerrum. Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of cheminformatics*, 12(1):70, 2020.

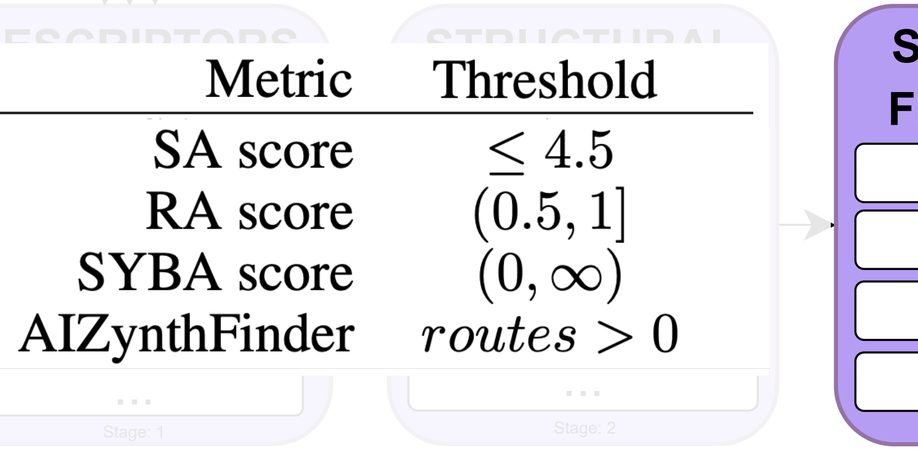
Five-Stage Filtering Pipeline

Stage 3: Synthesis Feasibility

SA score - Synthetic Accessibility score[1] - synthetic complexity score from 1 (easy) to 10 (hard).
RA score - Retrosynthetic Accessibility score[2] - probability of being a synthetic path for a compound.
SYBA score - Synthetic Bayesian Accessibility score[3] - classifier as easy or hard to synthesize.
AiZynthFinder[4] - machine-learning-guided retrosynthetic workflow.

4 evaluation criteria

Metric	Threshold
SA score	≤ 4.5
RA score	$(0.5, 1]$
SYBA score	$(0, \infty)$
AIZynthFinder	$routes > 0$



A molecule should pass all synthetic feasibility criteria **simultaneously** to pass the 3rd filtering stage

[1]Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1(1):8, 2009.
[2]Amol Thakkar, Veronika Chadimova´, Esben Jannik Bjerrum, Ola Engkvist, and Jean-Louis Reymond. Retrosynthetic accessibility score (rascore)–rapid machine learned synthesizability classification from ai driven retrosynthetic planning. *Chemical science*, 12(9):3339–3349, 2021.
[3]Milan Vors˘ila´k, Michal Kola´˘r, Ivan C˘melo, and Daniel Svozil. Syba: Bayesian estimation of synthetic accessibility of organic compounds. *Journal of cheminformatics*, 12(1):35, 2020.
[2]Samuel Genheden, Amol Thakkar, Veronika Chadimova´, Jean-Louis Reymond, Ola Engkvist, and Esben Bjerrum. Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of cheminformatics*, 12(1):70, 2020.

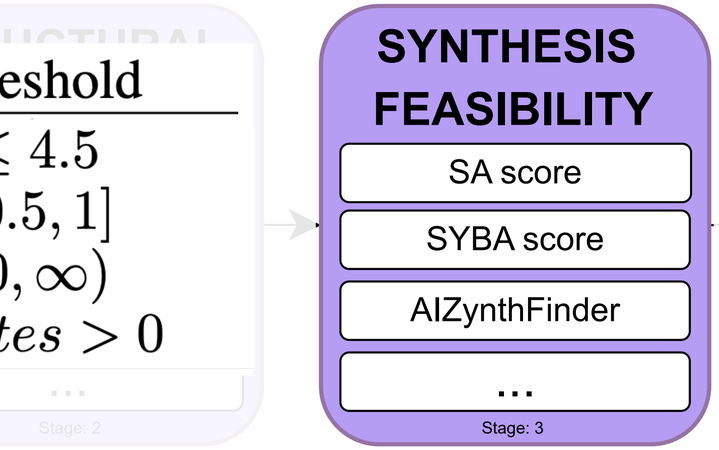
Five-Stage Filtering Pipeline

Stage 3: Synthesis Feasibility

SA score - Synthetic Accessibility score[1] - synthetic complexity score from 1 (easy) to 10 (hard).
RA score - Retrosynthetic Accessibility score[2] - probability of being a synthetic path for a compound.
SYBA score - Synthetic Bayesian Accessibility score[3] - classifier as easy or hard to synthesize.
AiZynthFinder[4] - machine-learning-guided retrosynthetic workflow.

4 evaluation criteria

Metric	Threshold
SA score	≤ 4.5
RA score	$(0.5, 1]$
SYBA score	$(0, \infty)$
AIZynthFinder	$routes > 0$



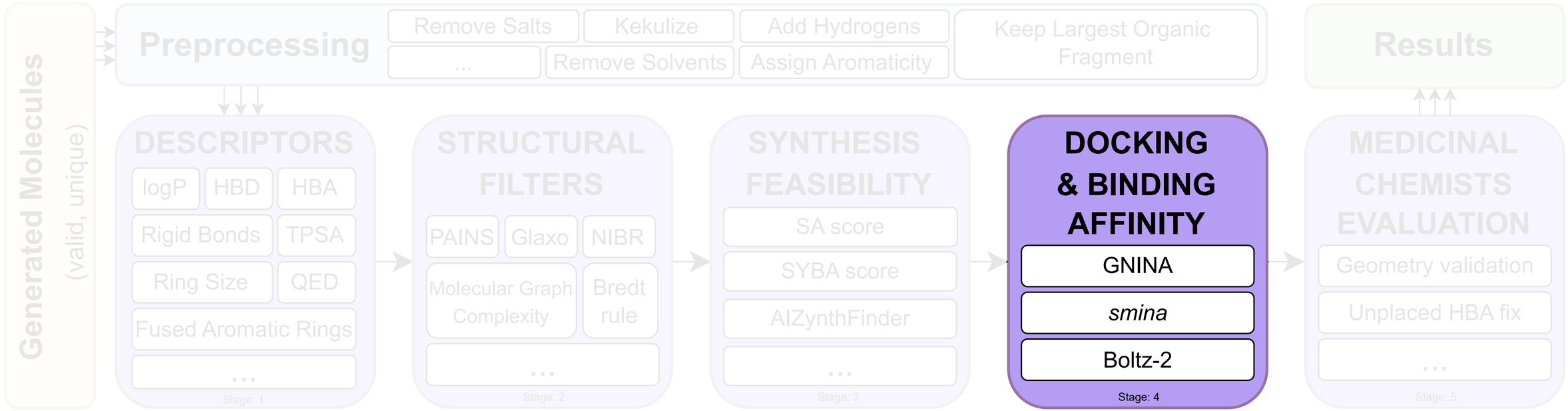
A molecule should pass all synthetic feasibility criteria **simultaneously** to pass the 3rd filtering stage

Molecules passed *Stage 3* go to the *Stage 4*

[1]Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1(1):8, 2009.
[2]Amol Thakkar, Veronika Chadimova´, Esben Jannik Bjerrum, Ola Engkvist, and Jean-Louis Reymond. Retrosynthetic accessibility score (rascore)–rapid machine learned synthesizability classification from ai driven retrosynthetic planning. *Chemical science*, 12(9):3339–3349, 2021.
[3]Milan Vorsˇila´k, Michal Kola´ˇr, Ivan Cˇmelo, and Daniel Svozil. Syba: Bayesian estimation of synthetic accessibility of organic compounds. *Journal of cheminformatics*, 12(1):35, 2020.
[2]Samuel Genheden, Amol Thakkar, Veronika Chadimova´, Jean-Louis Reymond, Ola Engkvist, and Esben Bjerrum. Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of cheminformatics*, 12(1):70, 2020.

Five-Stage Filtering Pipeline

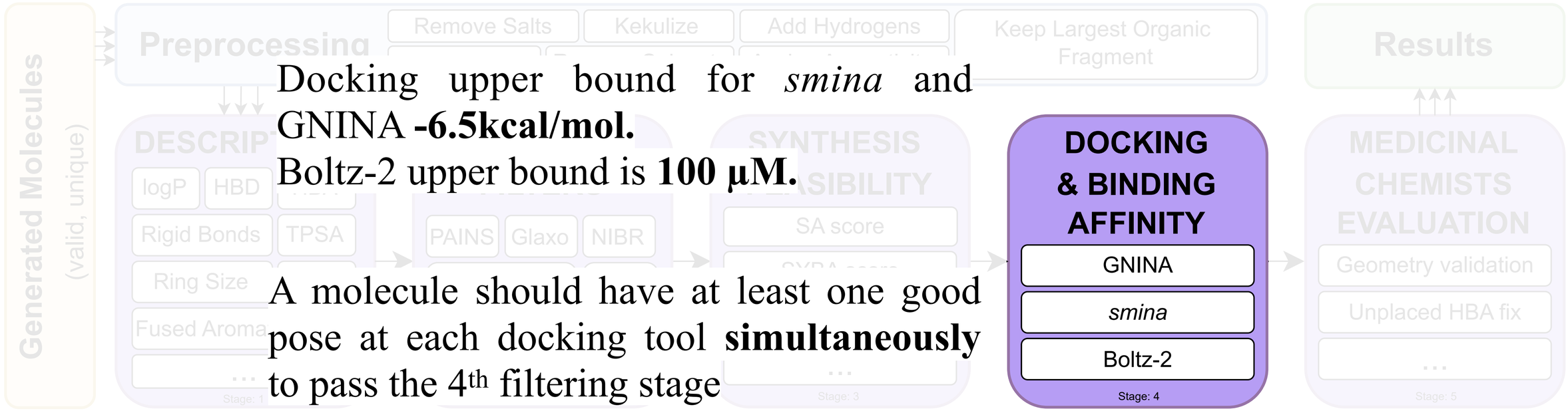
Stage 4: Docking and Binding Affinity Estimation



source: provided by the author

Five-Stage Filtering Pipeline

Stage 4: Docking and Binding Affinity Estimation



Docking upper bound for *smina* and GNINA **-6.5kcal/mol.**
Boltz-2 upper bound is **100 μM.**

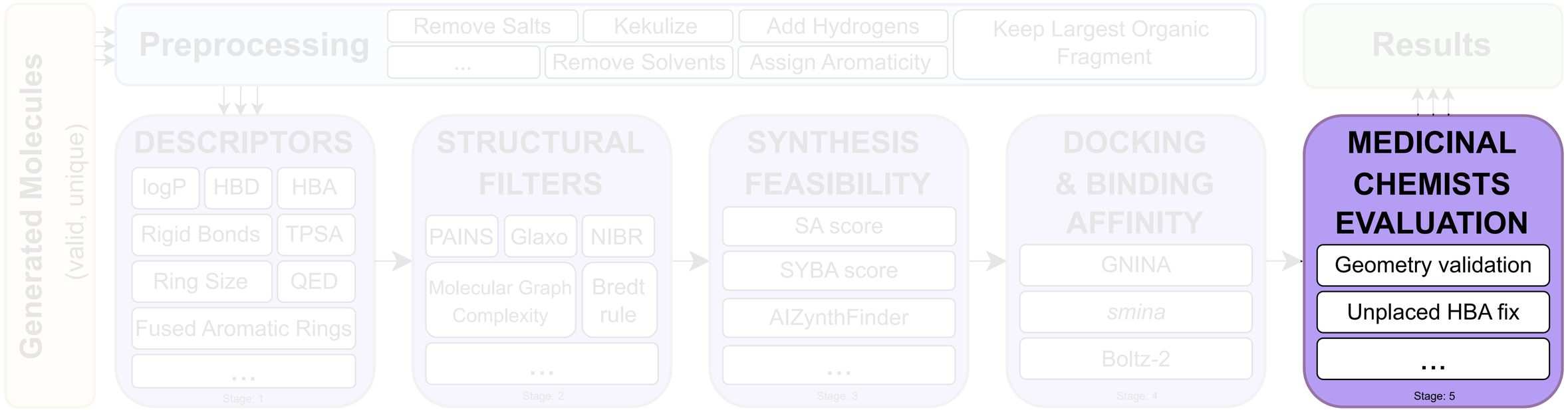
A molecule should have at least one good pose at each docking tool **simultaneously** to pass the 4th filtering stage

source: provided by the author

Molecules passed *Stage 4* go to the *Stage 5*

Five-Stage Filtering Pipeline

Stage 5: Medicinal Chemists Evaluation



source: provided by the author

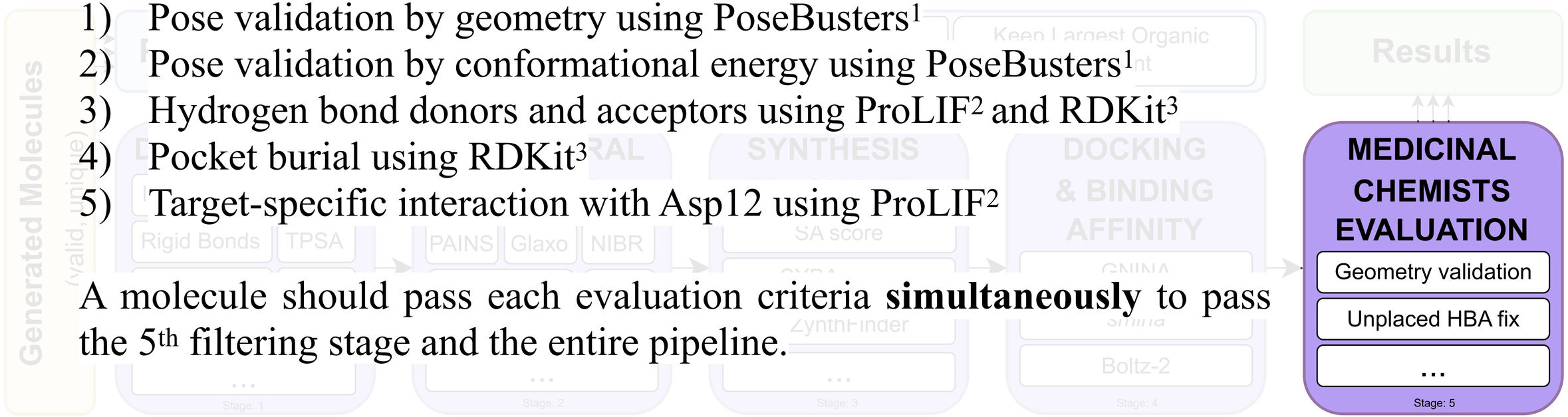
Five-Stage Filtering Pipeline

Stage 5: Medicinal Chemists Evaluation

Evaluation criteria:

1) Pose validation by geometry using PoseBusters[1]
2) Pose validation by conformational energy using PoseBusters[1]
3) Hydrogen bond donors and acceptors using ProLIF[2] and RDKit[3]
4) Pocket burial using RDKit[3]
5) Target-specific interaction with Asp12 using ProLIF[2]

A molecule should pass each evaluation criteria **simultaneously** to pass the 5th filtering stage and the entire pipeline.



[1]Martin Buttenschoen, Garrett M Morris, and Charlotte M Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 2024.
[2]Ce´dric Bouysset and Se´bastien Fiorucci. Prolif: a library to encode molecular interactions as fingerprints. Journal of cheminformatics, 13(1):72, 2021.
[3]Greg Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013.

INTRODUCTION

FIVE-STAGE FILTERING PIPELINE

RESULTS

CONCLUSION & DISCUSSION

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**

REINVENT4[1] (V, vanilla): unmodified, out-of-the-box model
REINVENT4 (P, prior): provided prior with mol2mol medium Tanimoto similarity threshold of 0.7
REINVENT4 (TL, transfer-learning): fine-tuned REINVENT4 (V) on 583 known KRAS G12D inhibitors[2]

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**

Dragonfly[3]: unmodified, out-of-the-box model
Dragonfly (b, biased): condition sampling on target compound descriptors

[1]Hannes H Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H Mervin, and Ola Engkvist. Reinvent 4: modern ai–driven generative molecule design. *Journal of Cheminformatics*, 16(1):20, 2024.
[2]Mohammad Ghazi Vakili, Christoph Gorgulla, Jamie Snider, AkshatKumar Nigam, Dmitry Bezrukov, Daniel Varoli, Alex Aliper, Daniil Polykovsky, Krishna M Padmanabha Das, Huel Cox Iii, et al. Quantum-computing-enhanced algorithm unveils potential kras inhibitors. *Nature Biotechnology*, pp. 1–6, 2025.
[3]Kenneth Atz, Leandro Cotos, Clemens Isert, Maria Ha˚kansson, Dorota Focht, Mattis Hilleke, David F Nippa, Michael Iff, Jann Ledergerber, Carl CG Schiebroek, et al. Prospective de novo drug design with deep interactome learning. *Nature Communications*, 15(1):3408, 2024.

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)	OVERALL PASS, %
Descriptors	**6616**	5669	1592	195	4089	936	1204	20,301 / 70,000 = 29.0014
Structural Filters	**4168**	1925	366	37	1325	593	413	8,827 / 70,000 = 12.61
Synthesis Feasibility	**1064**	303	265	22	918	222	276	3,070 / 70,000 = 4.3857
Docking & Binding Aff.	**648**	238	200	19	518	72	164	1,859 / 70,000 = 2.6557
Med.Chem. Evaluation	**110**	24	7	4	93	17	32	287 / 70,000 = 0.41
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**	

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF	OVERALL PASS, %
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444	21,577 / 80,000 = 26.9713
Structural Filters	197	**1459**	218	1392	682	195	255	136	4,534 / 80,000 = 5.6675
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4	2,027 / 80,000 = 2.5338
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0	1,119 / 80,000 = 1.3988
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0	318 / 80,000 = 0.3975
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**	

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MOLFINDER	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)	OVERALL PASS, %
Descriptors	**6616**	5669	1592	195	4089	936	1204	20,301 / 70,000 = 29.0014
Structural Filters	**4168**	1925	366	37	1325	593	413	8,827 / 70,000 = 12.61
Synthesis Feasibility	**1064**	303	265	22	918	222	276	3,070 / 70,000 = 4.3857
Docking & Binding Aff.	**648**	238	200	19	518	72	164	1,859 / 70,000 = 2.6557
Med.Chem. Evaluation	**110**	24	7	4	93	17	32	287 / 70,000 = 0.41
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**	

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

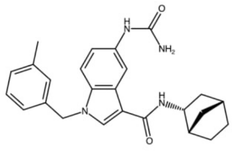
Stage /Model	DIFFSBDD	DRAGONFLY	DRAGONFLY (B)	DRUGFLOW	POCKET2MOL	PROTOBIND-DIFF	RESGEN	TARGETDIFF	OVERALL PASS, %
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444	21,577 / 80,000 = 26.9713
Structural Filters	197	**1459**	218	1392	682	195	255	136	4,534 / 80,000 = 5.6675
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4	2,027 / 80,000 = 2.5338
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0	1,119 / 80,000 = 1.3988
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0	318 / 80,000 = 0.3975
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**	

Results

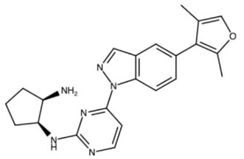
12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

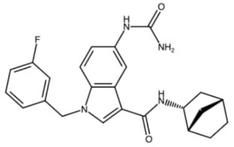
Stage /Model	GCPG	GENTRL	MolFinder	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**



GCPG-07684
consensus score = 0.830



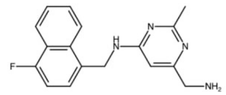
GCPG-00516
consensus score = 0.779



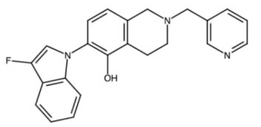
GCPG-06646
consensus score = 0.742

Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

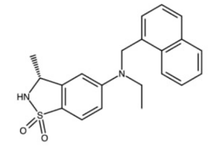
Stage /Model	DiffSBDD	Dragonfly	Dragonfly (b)	DrugFlow	Pocket2Mol	ProtoBind-Diff	ResGen	TargetDiff
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**



DrugFlow-03063
consensus score = 0.722



DrugFlow-08439
consensus score = 0.716



DrugFlow-04620
consensus score = 0.715

Results

12 models and 15 model setups

Table 2: Comparison of ligand-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	GCPG	GENTRL	MolFinder	PGMG	REINVENT4 (V)	REINVENT4 (P)	REINVENT4 (TL)
Descriptors	**6616**	5669	1592	195	4089	936	1204
Structural Filters	**4168**	1925	366	37	1325	593	413
Synthesis Feasibility	**1064**	303	265	22	918	222	276
Docking & Binding Aff.	**648**	238	200	19	518	72	164
Med.Chem. Evaluation	**110**	24	7	4	93	17	32
Pass	**110**	**24**	**7**	**4**	**93**	**17**	**32**

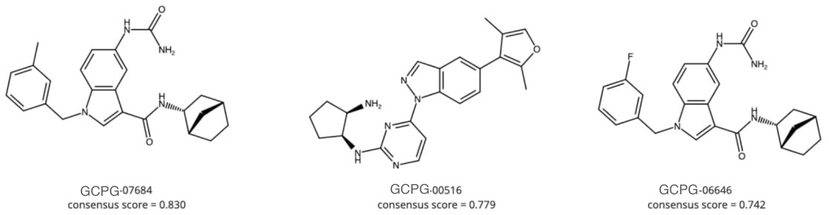
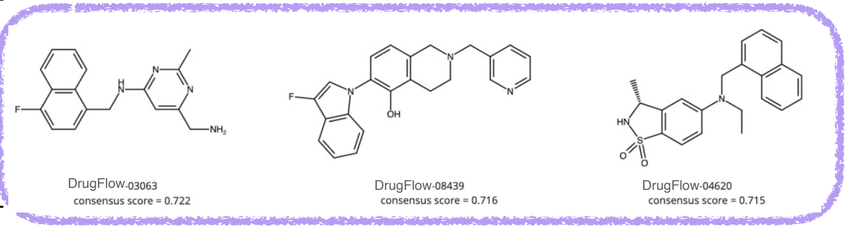


Table 3: Comparison of protein-based models, each with initial number of N_{gen} = 10,000 molecules

Stage /Model	DiffSBDD	Dragonfly	Dragonfly (b)	DrugFlow	Pocket2Mol	ProtoBind-Diff	ResGen	TargetDiff
Descriptors	3665	2779	1022	**5464**	2657	1466	1080	3444
Structural Filters	197	**1459**	218	1392	682	195	255	136
Synthesis Feasibility	24	**1207**	38	453	137	102	62	4
Docking & Binding Aff.	13	**575**	15	344	69	66	37	0
Med.Chem. Evaluation	0	**227**	4	62	12	7	6	0
Pass	**0**	**227**	**4**	**62**	**12**	**7**	**6**	**0**



Small number of left molecules does not highlight overall poor performance. DrugFlow[1] molecules are 4th, 5th and 6th top molecules according to consensus score

[1]Arne Schneuing, Ilia Igashov, Adrian W Dobbelstein, Thomas Castiglione, Michael Bronstein, and Bruno Correia. Multi-domain distribution learning for de novo drug design. *arXiv preprint arXiv:2508.17815*, 2025.

INTRODUCTION

FIVE-STAGE FILTERING PIPELINE

RESULTS

CONCLUSION & DISCUSSION

Conclusion and Discussion

- **Standard generative benchmarks are not good** proxies for real-world performance; optimizing for validity, synthesis or pocket fidelity independently is insufficient for actionable chemical space
- **Rigorous filtration** is essential for improving success rates and reducing costs by collecting and proceeding those molecules that meet chemical, medicinal, and task-dependent criteria
- Our *Five-Stage Filtering Pipeline* prioritizes **stress-testing** molecular generators against constraints that matter in early drug discovery
- Under our *Five-Stage Filtering Pipeline*, only a small fraction (less that 1%) of generated molecules pass all filters and remain applicable for future work

Research Team

https://ligandpro.ru/en

Ligand Pro | Skoltech

Skolkovo Institute of Science and Technology

References

Yongbeom Kwon and Juyong Lee. Molfinder: an evolutionary algorithm for the global optimization of molecular properties and the extensive exploration of chemical space using smiles. *Journal of cheminformatics*, 13(1):24, 2021.

Alex Zhavoronkov, Yan A Ivanenkov, Alex Aliper, Mark S Veselov, Vladimir A Aladinskiy, Anas- tasiya V Aladinskaya, Victor A Terentiev, Daniil A Polykovskiy, Maksim D Kuznetsov, Arip Asadulaev, et al. Deep learning enables rapid identification of potent ddr1 kinase inhibitors. *Na- ture biotechnology*, 37(9):1038–1040, 2019.

Yurong Zou, Tao Guo, Zhiyuan Fu, Zhongning Guo, Weichen Bo, Dengjie Yan, Qiantao Wang, Jun Zeng, Dingguo Xu, Taijin Wang, et al. A structure-based framework for selective inhibitor design and optimization. *Communications Biology*, 8(1):422, 2025.

Huimin Zhu, Renyi Zhou, Dongsheng Cao, Jing Tang, and Min Li. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nature Communications*, 14(1):6234, 2023.

Hannes H Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H Mervin, and Ola Engkvist. Reinvent 4: modern ai–driven generative molecule design. *Journal of Chem- informatics*, 16(1):20, 2024.

Kenneth Atz, Leandro Cotos, Clemens Isert, Maria Ha˚kansson, Dorota Focht, Mattis Hilleke, David F Nippa, Michael Iff, Jann Ledergerber, Carl CG Schiebroek, et al. Prospective de novo drug design with deep interactome learning. *Nature Communications*, 15(1):3408, 2024.

References

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Effi- cient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pp. 17644–17655. PMLR, 2022.

Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, et al. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, 2023.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L Blundell, Pietro Lio, et al. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12):899–909, 2024.

Lukia Mistryukova, Vladimir Manuilov, Konstantin Avchaciov, and Peter O Fedichev. Protobind- diff: A structure-free diffusion language model for protein sequence-conditioned ligand design. *bioRxiv*, pp. 2025–06, 2025.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equiv- ariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

Arne Schneuing, Ilia Igashov, Adrian W Dobbelstein, Thomas Castiglione, Michael Bronstein, and Bruno Correia. Multi-domain distribution learning for de novo drug design. *arXiv preprint arXiv:2508.17815*, 2025.

References

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Effi- cient molecular sampling based on 3d protein pockets. In *International Conference on Machine Learning*, pp. 17644–17655. PMLR, 2022.

Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, et al. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, 2023.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L Blundell, Pietro Lio, et al. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12):899–909, 2024.

Lukia Mistryukova, Vladimir Manuilov, Konstantin Avchaciov, and Peter O Fedichev. Protobind- diff: A structure-free diffusion language model for protein sequence-conditioned ligand design. *bioRxiv*, pp. 2025–06, 2025.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3d equiv- ariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*, 2023.

Arne Schneuing, Ilia Igashov, Adrian W Dobbelstein, Thomas Castiglione, Michael Bronstein, and Bruno Correia. Multi-domain distribution learning for de novo drug design. *arXiv preprint arXiv:2508.17815*, 2025.