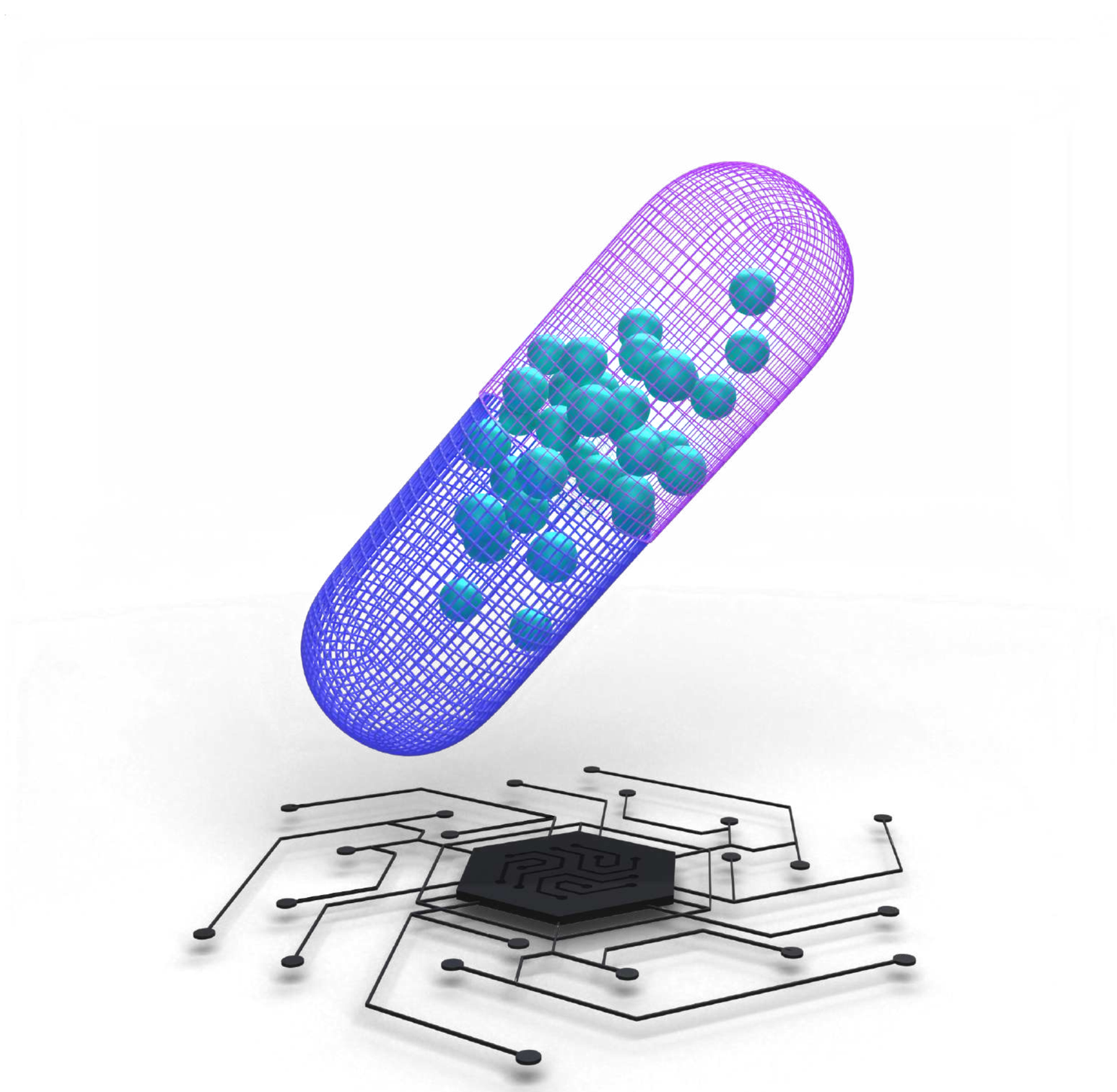
RECEPTOR.AI

Strategy and technology pipeline for drug candidate design



tel: +49 1517 283 7276 email: ai@receptor.ai website: receptor.ai

STRATEGY

The Receptor.AI in-house infrastructure is designed to facilitate three major drug discovery stages: hit discovery, lead discovery and lead optimization.

The researchers and engineers of Receptor.AI are working in close collaboration with the Partner to provide the best possible results for a given target using the whole repertoire of available in-house technologies. The results are uploaded to the SaaS platform, providing the client with a user-friendly interface to select the best compounds that are annotated using over 80 parameters.

The general principle of our platform is dynamic assembly of the whole drug discovery pipeline from reliable and self-consistent modules based on the constraints and requirements imposed by the target disease, protein of interest and their broad biomedical context.

Our platform is based on the string synergy between two computational and experimental approaches. All experimental techniques are provided by Receptor.AI partners, which are highly specialized on the corresponding methods and thus provide the best quality and pricing. All computational techniques are implemented in house.

Experimental part of the platform consists of the framework for generating proprietary data for AI model training and the set of techniques for experimental validation of the virtual screening and simulation results. The former includes 3D structure determination by the CryoEM, large-scale ultra-high-throughput DEL screening for generating binders and a set of high-throughput in vitro biochemical and biophysical techniques for target characterization and binding assessment.

The computational part consists of the set of predictive AI models and the infrastructure for molecular simulations. The AI models cover the whole drug discovery pipeline from target identification and pocket prediction to virtual screening, selectivity assessment, drug-membrane interactions prediction and evaluation of 80+ ADME-Tox, physico-chemical and drug likeness parameters. Molecular simulations are used to generate ensembles of representative protein conformations, to model the target protein in its native environment for confirming structural data and the mode of action and dedicated simulation for binding pocket prediction and attribution, including the pipeline of the allosteric and cryptic pockets identification.

Stage 0: Target exploration

In this stage, the Receptor.AI examines the target protein and determines the most appropriate binding pocket and the general strategy for subsequent development. The Receptor.AI also performs preliminary virtual screening against a small stock chemical space to get starting information about the binding propensities and activities of common chemical scaffolds against the target of interest.

Pocket prediction and molecular simulations

For proteins with no known binding pocket or several possible binding pockets, Receptor.AI performs pocket prediction by using two possible techniques:

- *Pocket prediction by AI model trained on all known pockets in PDB.* The model has an accuracy of ~75% and gives a good preliminary estimate of the pocket location.
- *Pocket prediction from the literature data.* If the functional residues are known from experiments, the pocket could be predicted in accordance with this data.

The final decision about the pocket, which will be used in subsequent stages of the pipeline, is based on the results of preliminary virtual screening and experimental validation.

At this stage, molecular simulations of the target proteins are performed to obtain the ensemble of representative conformations, which would be used to account for protein dynamics and conformational changes at this and later stages of the pipeline.

STRATEGY

Stage 1: Hit discovery

At this stage, the initial hit compounds are identified, and data is generated.

Chemical space

A diverse multi-billion chemical space composed of the databases of several major vendors (Enamine Real Space, WuXi Galaxi, Otava CHEMriya) is used (~30B compounds).

Virtual screening

At this stage, molecular simulations of the target proteins are performed to obtain the ensemble of representative conformations, which would be used to account for protein dynamics and conformational changes at this and later stages of the pipeline.

- The virtual screening is performed using the “level 1” stack of technologies:
- DTI and FB-DTI activity prediction models.
- Level 1 selectivity prediction model (proteome-wide DTI rank over ~9.3K proteins).
- Standard ADMET assessment and filtering (38 endpoints). The user can choose any set of ADMET parameters and set the filtering cutoffs for them prior to the screening.
- Automated docking with AI rescoring (docking to single protein conformation).

Integration of experimental feedback

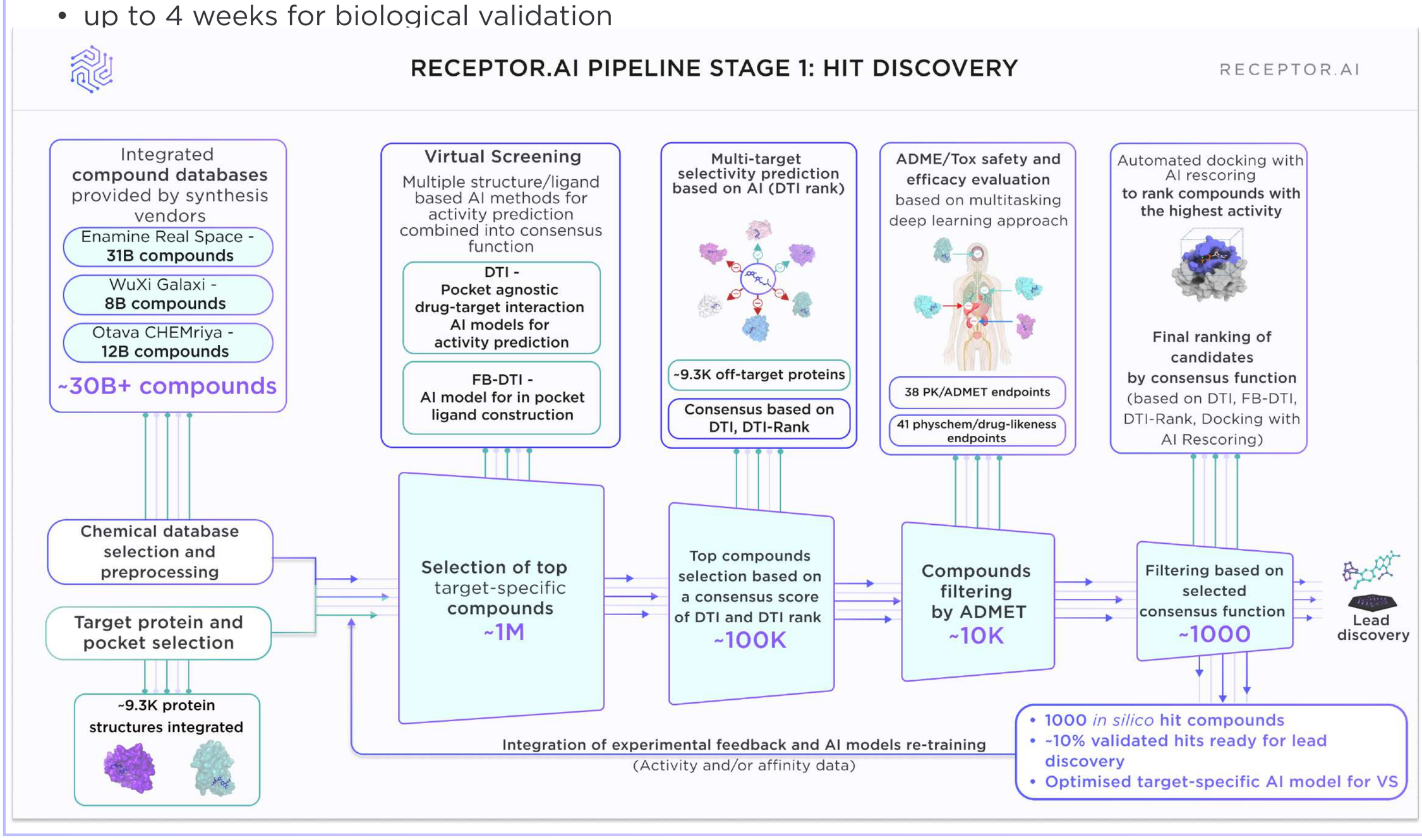
At this stage, the Receptor.AI advises using the activity assays, which allow for obtaining reliable data about the biological activity of at least ~100 potential hit candidates. The goal of validation is to prioritize active compounds and use the knowledge about their molecular scaffolds to train target-focused AI models for the following stages. The DTI model is re-trained based on experimental feedback to be used in the next stage.

Deliverables

- *In silico* hit candidates (1000).
- Experimentally validated hit compounds.
- Optimized target-specific AI model for the lead discovery stage.

Estimated timeline

- up to 3 weeks molecular simulations
- up to 4 weeks for the virtual screening stage
- up to 6 weeks for the synthesis and delivery of compounds
- up to 4 weeks for biological validation



STRATEGY

Stage 2: Lead discovery

At this stage, the hit compounds are developed towards the leads.

Chemical space:

The chemical space for lead discovery is built based on the molecular scaffolds of the hit compounds from the previous stage. For this, we can use several approaches:

- Scaffold-based AI molecular generator for series expansion of the hits. The model can produce billions of chemically correct compounds by modifying substituents of the given scaffold.
- Pharmacophore search in combined combinatorial space from several vendors could be performed to find the most similar synthesizable molecules. The above is preferable if you want to get easily synthesizable leads.
- Incorporation of any derivatives designed by human chemists. Receptor.AI can incorporate any molecules designed by your chemists into the virtual screening database and/or make a series expansion for them.

Generating an ensemble of target conformations

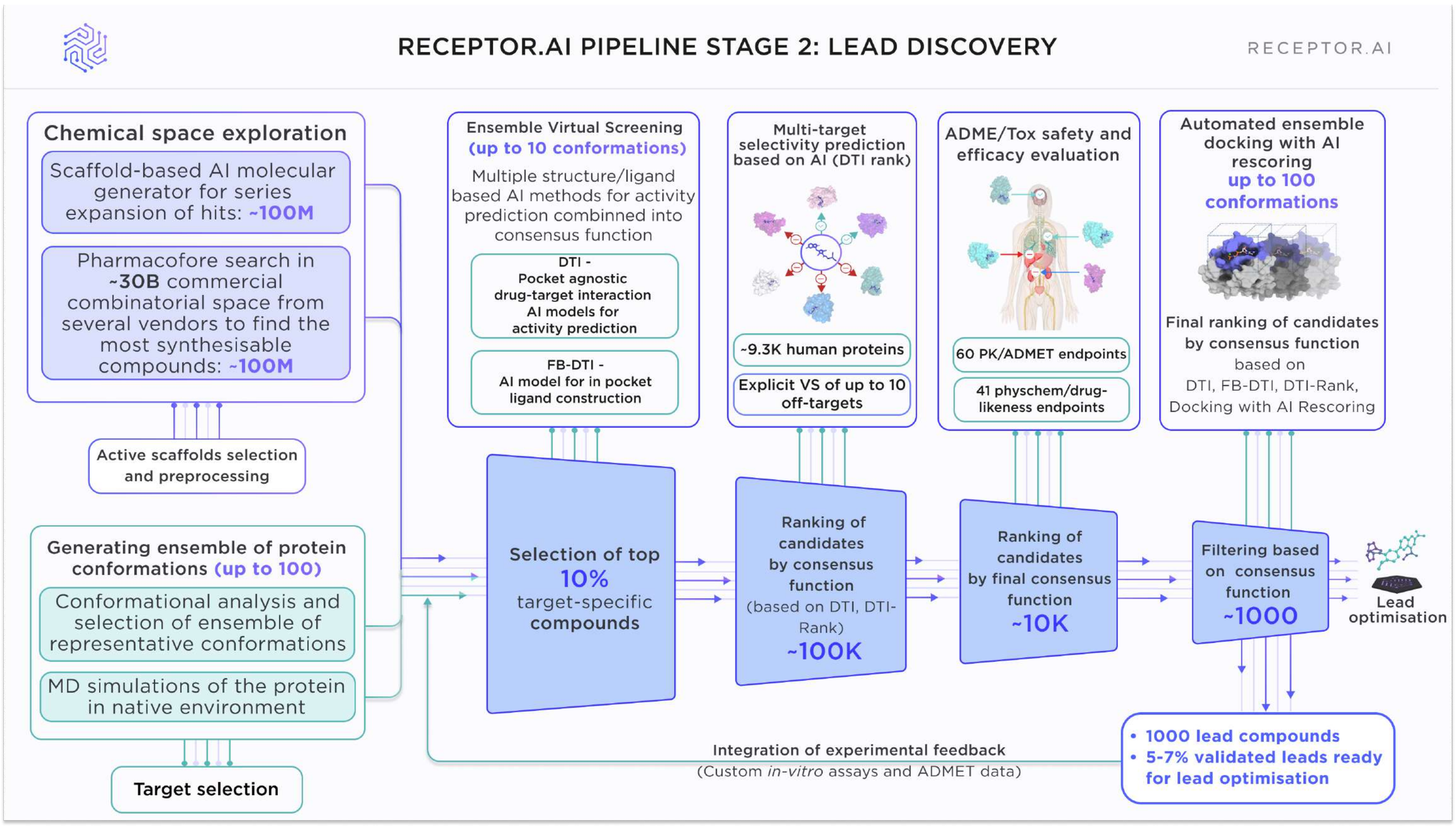
At this stage, molecular simulations of the target proteins are performed to obtain the ensemble of representative conformations, which would be used to account for protein dynamics and conformational changes at this and later stages of the pipeline.

The Receptor.AI is going to perform MD simulations of the target protein in its native environment. The conformational analysis and selection of the ensemble of representative conformations will follow.

Virtual screening

The virtual screening is performed using the “level 2” stack of technologies:

- DTI and FB-DTI target-focused activity prediction models will be used in the ensemble screening mode (up to 100 protein conformations). The results from all conformations will be merged and ranked.
- Level 2 selectivity prediction is used. The above includes proteome-wide DTI rank calculations and an explicit virtual screening of up to 10 unwanted off-target proteins. The top compounds for each off-target would be removed from the list of candidate compounds if found there.



STRATEGY

- Advanced level 2 ADMET assessment and filtering (~60 endpoints). The user can choose any set of ADMET parameters and set the filtering cutoffs for them prior to the screening.
- Assessment of 40 phys-chem parameters and drug-likeness metrics with the possibility of filtering over them within the cutoff provided by the user.
- Ensemble docking with AI rescoring will be performed. Up to 100 protein conformations are evaluated. The ranking of compounds is performed by the docking score distributions, which provides better accuracy.

Integration of experimental feedback:

At this stage, it is advised to use the activity assays, which allow for obtaining reliable data about the biological activity of at least ~100 potential lead candidates and, if possible, specific cellular or in vivo assays. The goal of validation is to prioritize active molecules with favourable ADMET profiles.

At this stage, we can also incorporate experimental data from any custom assays specific to the target protein and any custom ADMET endpoints.

Based on experimental feedback, the DTI model and the multi-parametric ADMET model will be re-trained to be used at the next stage.

Deliverables:

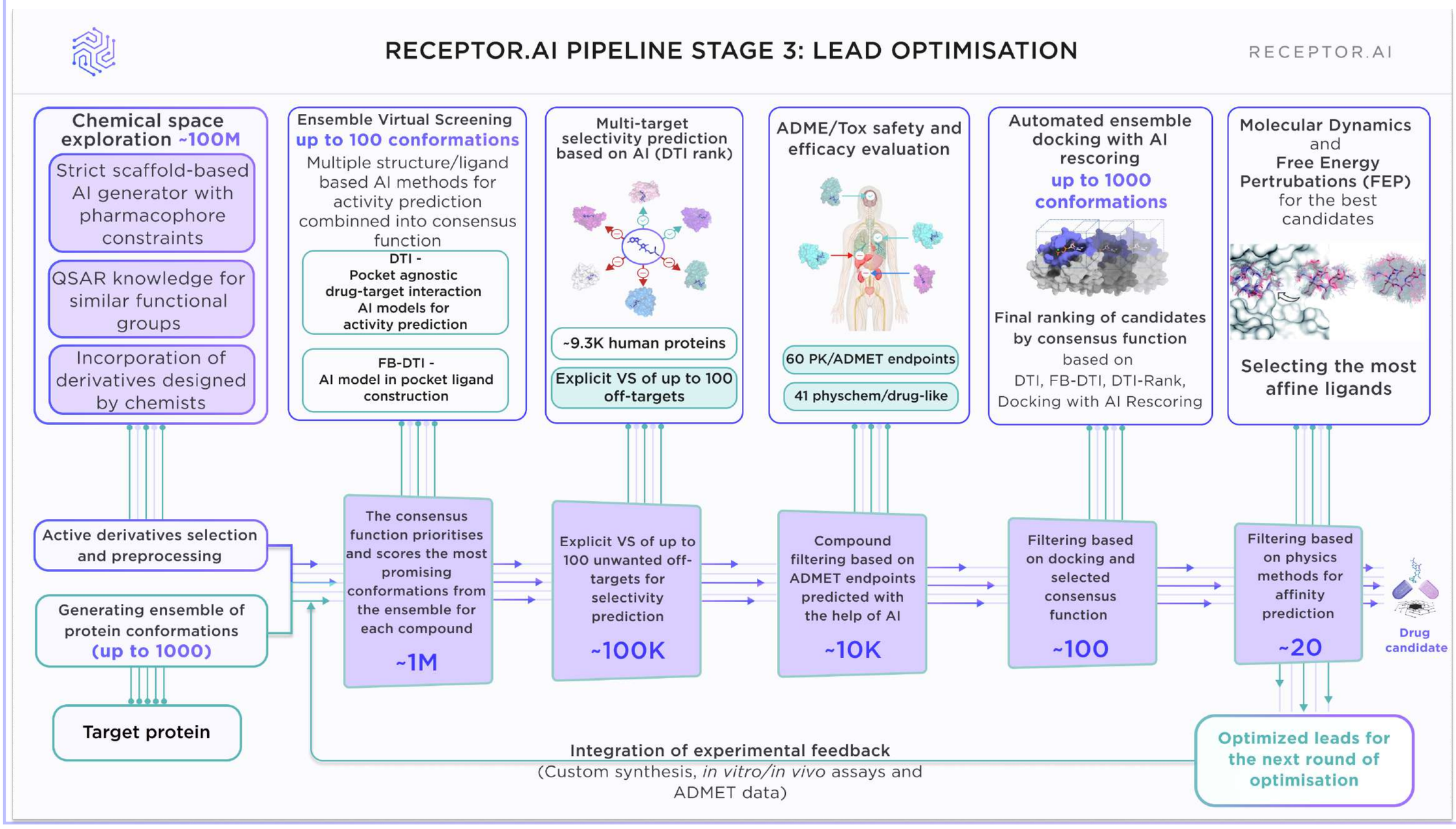
- *In silico* lead candidates (1000).
- Experimentally validated leads.
- Optimized target-specific AI model and ADMET models for the lead optimization stage.

Estimated timeline

- up to 3 weeks molecular simulations
- up to 4 weeks for the virtual screening stage
- up to 8 weeks for the synthesis and delivery of compounds
- up to 6 weeks for biological validation
- up to 3 iterations with experimental laboratory

Stage 3: Lead optimization

This is the final stage, which leads to the selection of one or several drug candidates, which are ready for IND and preclinical phases. The most precise techniques are used at this stage in combination with the most focused chemical space.



STRATEGY

Chemical space:

The chemical space for lead optimisation is built based on the high-precision modification of the lead's scaffolds. For this, we can use several approaches:

- Strict scaffold-based AI molecular generator with enforced pharmacophore constraints.
- Incorporation of QSAR knowledge for similar functional groups (if available) based on the literature data and the expertise of chemists.
- Incorporation of any custom derivatives designed by human chemists. We can incorporate any molecules designed by your chemists into the virtual screening database and/or make a series expansion for them.

Virtual screening:

The virtual screening is performed using the highest "level 3" stack of technologies:

- DTI and FB-DTI target-focused activity prediction models will be used in the advanced ensemble screening mode (up to 1000 protein conformations). The results from all conformations will be merged and ranked.
- Level 3 selectivity prediction is used. This includes proteome-wide DTI rank calculations and an advanced explicit virtual screening of up to 100 unwanted off-target proteins. The top compounds for each of the off-target will be removed from the list of candidate compounds if they are found there.
- "Ultimate" level 3 ADMET assessment and filtering for all ~60 endpoints, plus custom ADMET endpoints and custom assays endpoints (if any). Strict cutoffs are applied to the chosen set of the most critical endpoints.
- Assessment of 40 phys-chem parameters and drug-likeness metrics with the possibility of filtering over them within the cutoff provided by the user.
- Advanced ensemble docking with AI rescoring will be performed. Up to 1000 protein conformations are evaluated for each candidate molecule. The ranking of compounds is performed by the docking score distributions, which provides better accuracy.
- MD+FEP calculations will be performed for the best candidate molecules (up to 10) to get the most precise estimate of the binding mode and affinity (optional).

Integration of experimental feedback

At this stage, Receptor.AI advises using the most precise activity and in vivo assays. The goal of validation is to select hit candidates with excellent ADMET profiles, which are ready for IND and preclinical trials.

Based on experimental feedback, the DTI model and the multi-parametric ADMET model will be re-trained to be used on the next iteration of lead optimization (if needed).

Deliverables

- Optimized leads for the next round of optimization.
- Drug candidate ready for IND and preclinical phase.

Estimated timeline

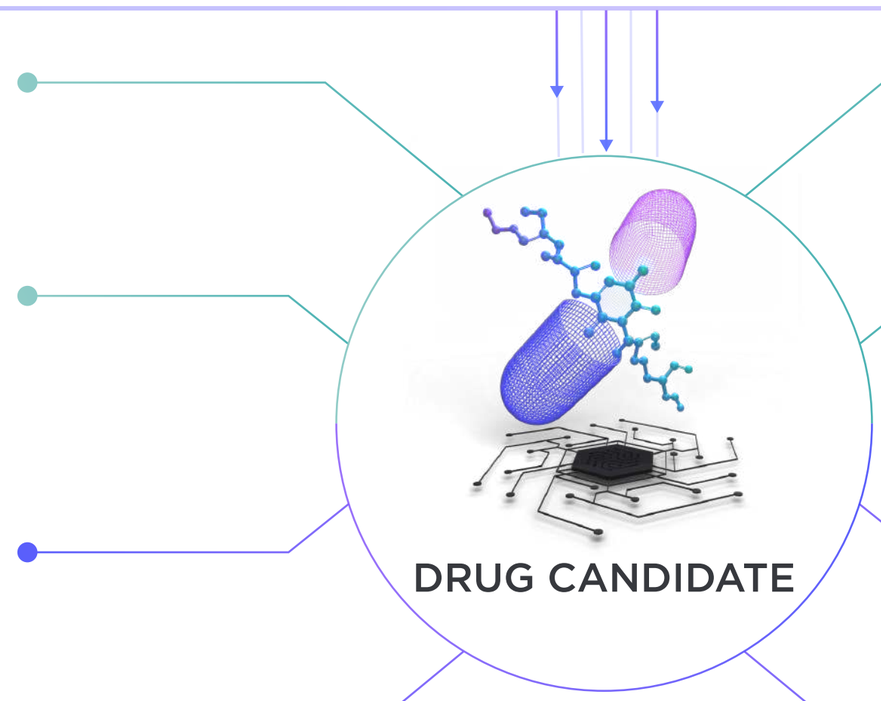
- up to 3 weeks molecular simulations
- up to 4 weeks for the virtual screening stage
- up to 9 weeks for the synthesis of compounds
- up to 2 months of biological validation
- up to 4 iterations with experimental laboratory

Basic characteristics of different pipeline stages

Stage	Chemical space	Virtual screening	Molecular modeling	Experimental feedback
Exploratory	Stock	ADMET: L0	Pocket prediction	L1: affinity/ activity
Hit Discovery	Large enumerated/ combinatorial	Selectivity: L1 ADMET: L1 Docking: L1	-	L1: affinity/ activity
Lead Discovery	Generated series/ pharmacophore search	Selectivity: L2 ADMET: L2 Docking: L2 Ensemble: L2	MD for conformational ensemble generation	L2: affinity/activity custom ADMET custom assays
Lead Optimization	Restrained generated series	Selectivity: L3 ADMET: L3 Docking: L3 Ensemble: L3	MD+FEP for affinity prediction	L3: affinity/activity custom ADMET custom assays *in vivo*

IMPACT

Optimized leads (drug candidates) could be obtained in 2.5 years.



Drug likeness
40 physchem and druglike-ness filters

Novelty
x200 higher chance to find novel compounds

Selectivity
Against **~9K+** human proteins

Speed of screening
~30B enumerated chemical space in less than a day

Unlocking undruggable targets
PPI, unknown pockets

x7 reduced cost and time
for experimental validation

Safety
40+ PK/ADME-Tox AI endpoints

High success rate
proven by experimental validation