

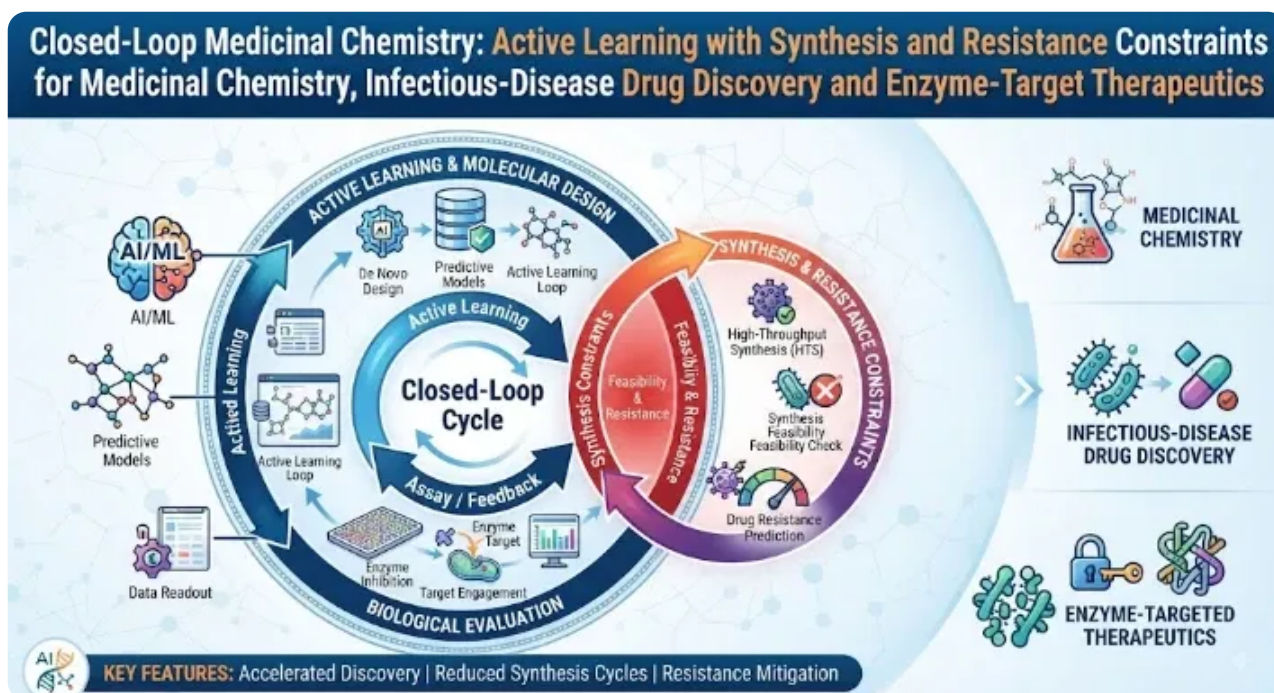
Closed-Loop Medicinal Chemistry: Active Learning with Synthesis and Resistance Constraints for Medicinal Chemistry, Infectious-Disease Drug Discovery and Enzyme-Targeted Therapeutics

Md. Humayan Kabir

Dec 7, 2025

Abstract

This article presents closed-loop medicinal chemistry for infectious-disease enzyme targets as a learning system that balances potency, pathogen penetration, synthesis feasibility, safety, and resistance. Structural information can guide inhibitor design, but whole-cell and intracellular assays are needed to determine whether compounds reach the target and disrupt the intended pathogen pathway. The article links enzyme potency, cellular uptake, target occupancy, metabolic collapse, pathogen killing, and resistance evolution through iterative design–make–test cycles. Active learning, graph models, docking, retrosynthesis, Bayesian optimization, and multi-objective acquisition are used to choose compounds that maximize both performance and information gain. Prospective synthesis, resistant-mutant testing, pathway rescue, exposure measurement, strain panels, mutation-frequency assays, serial passage, sequencing, and fitness analysis provide experimental feedback. The framework also includes tissue penetration, host-cell access, protein binding, metabolism, pharmacokinetics, host-enzyme selectivity, mitochondrial and genotoxicity testing, microbiome considerations, and drug combinations. Route robustness, purification, hazardous reagents, solid state, and scalable starting materials are treated as design constraints. Overall, the loop turns resistance into an optimization variable rather than a late-stage surprise.



The advantage of rapid learning cycles in infectious-disease medicinal chemistry is that resistance to a drug can compensate for any improvements in potency – and therefore be the determining factor. Closed-loop medicinal chemistry for enzyme targeted infectious disease therapeutics has come a long way from just searching for hits due to the size of the design space, the lack of access to target in the pathogen, and the ability to make only a few analogs, as well as resistance mechanisms evolving over time. The more interesting question is whether the active-learning cycle of choosing compounds that are potent, penetrate the cell, can be synthesized, are safe, and can withstand resistance is stable in realistic biological and development environments. This change is from a ranking activity to a discovery program based on exposure, cause and repeatability. Linking molecular design to experiments, which will reveal why a candidate is successful, where it fails and which constraints must be given priority is increasingly vital to progress in closed-loop anti-infective discovery by the scheduled publication time.

Structural clarity is provided by enzyme-targeted programs, and whole-cell and intracellular assay can provide information on whether designed inhibitors will penetrate biological barriers, and disrupt the target pathogen pathway. Past experience and recent development demonstrates that as biological complexity (neurons, etc.) increases, it is hard to find a nice laboratory signal that remains as is. A helpful research methodology is then to maintain a fairly strong activity signal while adding a more realistic context to which the signal must be effective.

Mechanistic depth begins with understanding links between enzyme inhibition, cellular

uptake, target occupancy, collapse of metabolic pathways, killing of the pathogen and resistance route. This structure can be masked by simple endpoint assays as similar outcomes can come from diverse causal pathways. Enzyme assays, whole-cell potency, models of infection inside cells, binding to the target, metabolomic analysis of cell responses, kill curves, and selection of resistant cells gives a more comprehensive view, as this allows for the dissociation of direct effects from downstream adaptation, nonspecific stress and delivery problems. These measurements allow for the creation of an attractive response that can be tested as a chain of chemistry, biological state and functional outcome.

Also design variables must be considered as interacting aspects of the design rather than individual variables. These factors need to be considered together, and a gain in one may result in a loss in another. These parameters work together, and a gain in one can lead to a loss in another. Therefore, it is more practical to try to optimize trade-offs, rather than a single maximum. Multi-parameter profiles are particularly informative when potency, selectivity, stability, tissue exposure and manufacturability all impact the translational value.

When uncertainty is visible, prediction is useful, but when there is uncertainty, there is computational leverage via active learning, graph models, docking, retrosynthesis, Bayesian optimization, and multi-objective acquisition. When assessing a novel scaffold or material, or an underrepresented patient state, models trained on existing chemical or biological regions may be unreliable. With the help of calibration, forward testing, DOA checks, and disagreements within model families, an early warning of fragile predictions is provided before costly experiments are performed.

Orthogonal validation is a valuable tool in experimental design. Prospective synthesis, resistant-mutant testing, pathway rescue, biochemical confirmation, exposure measurement and external strain panels address various aspects of the causal question and minimize reliance on a single assay technology. Consensus in the field of chemistry, imaging, omics, genetics and functional biology leads to a more solid mechanistic argument than multiple repeat measurements of the same endpoint. Disagreement is also informative in that it will often uncover subtle limits of exposure, assay artifacts, compensatory pathways, or context-dependent pharmacology.

Another complexity arises from the diversity of life. Target abundance, barrier properties, metabolic state, host-cell niche, or resistance history can change target abundance, barrier properties, metabolic state, immune tone, or pathway dependence without changing the

nominal treatment, and this can alter the effectiveness of the treatment. Average responses thus may not reflect subgroups that may have different exposure-response relationships. Biological variation is incorporated into the design and is not a late justification for inconsistent results – it is analyzed on a stratified basis, there are representative model panels, and there are explicit checks for biological variation.

Resistance and adaptation needs to be taken into consideration early as every optimization cycle can focus on an escape-prone interaction. Activity is known to survive evolutionary or compensatory pressure by mutation-frequency assays, serial passage, mutant libraries, sequencing, cross-resistance, fitness cost, and combination testing. A low potency, but good blockage of common escape route, candidate can be more valuable than a weak lead. Resistance-aware development also can hone mechanism as recurrent escapes often will point to the proteins, pathways or exposure processes with the most causal influence.

Molecular promise to reachable biology through delivery and pharmacokinetics. The active form's ability to accumulate in the tissue, penetrate host cells, bind proteins, be metabolized, reach intended tissue and have a dose-response effect that is compatible with the proposed mechanism determines whether or not the active form reaches the intended site at the necessary concentration and duration. The Nominal dose approach cannot capture this relationship because absorption, distribution, binding, transformation, clearance and local barriers will alter the effective exposure. Mechanistic exposure assessments should, therefore, be placed along with efficacy data from the earliest development stages.

Mechanism is a powerful tool that can enhance the safety assessment. Host-enzyme selectivity, cytotoxicity, mitochondrial testing, genotoxicity, drug-drug interactions, and microbiome considerations help differentiate target linked pharmacology from widespread cellular damage/immune activation/reactive chemistry/drug formulation driven toxicity. Comparative testing of therapeutic activity and undesirable effects is possible with disease-relevant and healthy systems.

The manufacturing is not a downstream logistics issue, it's a scientific variable. Despite having the same nominal composition, changes in the route of synthesis, yield of reaction, hazardous reagents, purification effort, solid-state form, formulation and scalable starting materials can all have an impact on product quality, biological activity and exposure. Quality-by-design thinking is not just end-stage specification testing, but it's connecting critical material attributes and process parameters with measurable performance.

Industry success that's independent of raw material variation, larger equipment, storage and transport, and independent production runs makes for more believable scalability.

Realistic models of the clinical use of translation are also needed. Treatment timing, dose frequency, route, diagnostic requirements and acceptable complexity depends on acute versus chronic infection, tissue site, diagnostic availability, combination standard, adherence and resistance epidemiology. A technically beautiful platform, which is not useful in the setting where treatment is required, because its administration requirements, storage conditions or monitoring requirements are too high, or because of the cost, is not of practical value. Thus, the judicious choice of development decisions can be optimized in early stage commitments when it is easier to change than when we are near the end of the process.

Data integration is the link between these layers. Causally-understandable chemical structures, reaction outcomes, enzyme potency, cell exposure, pathogen kill, resistance mutations, safety, and pharmacokinetics can be represented and stored instead of independently with respect to each assay result. Iterative learning is reproducible thanks to traceable metadata, versioned models, standardized controls and prospective decision rules. The end result is a development record that reveals not only which one candidate came out on top, but what the assumptions were on which that candidate was made, and what the uncertainties were that were left unanswered.

Closed-loop anti-infective programs like the ones developed at Pfizer appreciate information gain, whole-cell translation, resistance retention, a prospective hit-rate, a success rate in synthesis, tissue exposure, and breadth of strains. The higher the quality of a decision, the more success criteria will be met: durability, reproducibility, exposure margin, biological breadth, manufacturing robustness, as well as potency. These metrics incentivize candidates to optimize for a specific assay, instead of a general one, and make it more likely that a candidate will be retained if they are found to be credible by an independent assay. The rationale of a less dramatic outcome being the development option is also made clear by a transparent scorecard.

Also, a strong program specifies a constructive way of failure. Negative outcomes from computationally tractable compounds that cannot be synthesized; enzyme hits that do not allow entry of the pathogen; mutants that can be quickly selected; changes in exposure that cause instability; and limited range of strains covered by the exposure all denote limits to the mechanism and thereby limit repeat investment in less successful hypotheses. These factors can lead to a lack of synthesis, to a lack of target engagement, to

unstable exposure, to the emergence of resistance, or to the loss of effect in subgroups. The absence of synthesis, a lack of target engagement, unstable exposure, emergence of resistance or lack of effect in subgroups can all give a future model a more realistic view of the search space. Successive failures help to learn as much as successful experiments.

A closed loop allows medicinal chemistry to learn from the synthesis and pathogen evolution, and enables resistance to be a factor of design, not a design fail. The main advantage of the central advantage is that it allows to connect isolated performance claims with the mechanism, exposure, biological diversity, safety and production aspects of the explanation. This organization makes it easier to select candidates since there is a test for every major development claim. Progress is now not as much based on the "wow" factor of single experiments, but on a logical string of experiments that holds in the face of growing complexity in laboratory systems and then in real clinical settings.