

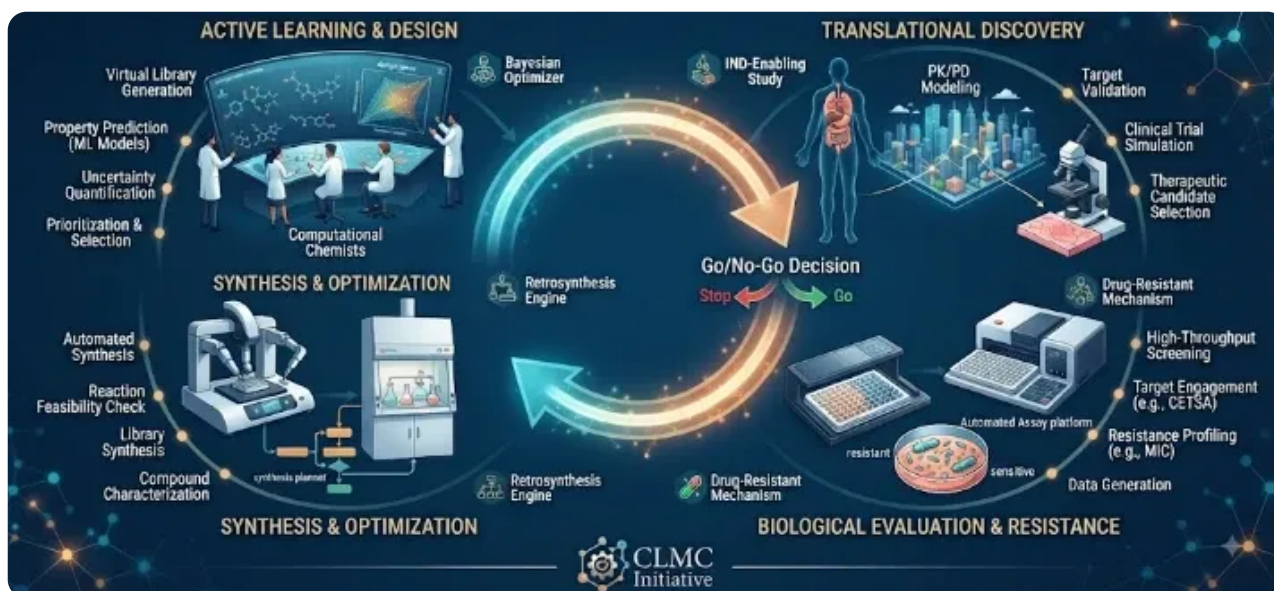
Closed-Loop Medicinal Chemistry: Active Learning with Synthesis and Resistance Constraints for Medicinal chemistry and translational discovery of new therapeutics

Md. Humayan Kabir

Mar 2, 2025

Abstract

This article presents closed-loop medicinal chemistry as a decision system in which active learning, synthesis, biological testing, resistance, and developability continuously inform the next experiment. Linear discovery cycles can waste synthesis when potency, mechanism, exposure, safety, and manufacturability are optimized separately. Active learning therefore selects compounds for information gain as well as expected performance, balancing exploitation of promising chemistry with exploration of uncertain regions. Scaffold, substituent, stereochemistry, synthetic route, potency, selectivity, permeability, metabolic stability, target occupancy, and resistance frequency are treated jointly. Failure routes include target mutation, bypass signaling, efflux, metabolic adaptation, polypharmacology, assay drift, and optimization toward nontranslatable endpoints. Free concentration, cell penetration, target occupancy, tissue distribution, active metabolites, clearance, and duration above a pharmacodynamic threshold define exposure. Bayesian optimization, graph models, uncertainty estimation, retrosynthesis, Pareto analysis, and information-gain acquisition support prioritization. Prospective cycles, held-out chemical series, independent synthesis, orthogonal assays, resistant models, and predeclared success criteria provide validation. The central idea is that every synthesized compound should answer a decision-relevant question rather than simply enlarge an analogue series.



Closed-Loop Medicinal Chemistry is focused on one of the key challenges in closed-loop medicinal chemistry for new therapeutics: optimizing potency, developability, resistance, and mechanism independently leads to wasteful linear design. The more robust development approach is based on an actual quantifiable mechanism, not a single (wowza) assay result. With that change, chemistry, biology, exposure, resistance, and safety, become components of one translational argument, as does production.

The science core is closed-loop discovery via active learning to choose experiments that enhance the quality of the candidates and understanding of the chemical-biological system. There is no single measurement that can be used to support the entire mechanistic claim, as each one addresses a separate question. It minimizes the potential for misinterpretation of potency, binding, formulation performance and/or model fit as therapeutic understanding.

Active learning transforms medicinal chemistry by choosing the next experiment to perform based on the amount of information that will be gained. A good cycle involves exploring new, promising areas and testing new, unknown chemistry. Feasibility for synthesis, ADME and target occupancy can be all incorporated into the same objective function as well as resistance data. The true case of a truly translational loop is when a candidate is being promoted to the next step only after predictions are still valid after prospective synthesis and biologically independent testing.

Some design variables are inter-dependent and must be treated together as scaffold, substituent, stereochemistry, synthetic route, potency, selectivity, permeability, metabolic stability, target occupancy and resistance frequency can all change at the same time.

When optimizing a single factor is accompanied by an unseen penalty for another factor such as potency vs stability, selectivity vs delivery, or manufacturability, it may not be a good strategy. Multiparameter decision rules, then, offer a more realistic basis for candidate ranking than does a single numerical endpoint.

Failure routes are an important aspect of the mechanistic confidence, too. Some of the most important escape mechanisms are target mutation, bypass signaling, efflux, metabolic adaptation, polypharmacology, assay drift and optimization towards nontranslatable endpoints. These can be planned as deliberate stress tests to provide design information for the next chemistry, formulation, dosing, or model cycle – and it is no surprise.

The link between exposure and biological effect is the nominal dose. Important variables are free concentration, penetration into cells, target site occupancy, tissue distribution, metabolites that are active, clearance, and time above pharmacodynamic threshold. A simplified assay may score high on intrinsic activity but have a low score for exposure at the causal site, and thus may not be suitable for translation.

The concentrations for direct engagement or functional perturbation must be interpreted at realistic concentrations. At high doses, it can cause secondary stress responses, off-target effects, and/or assay artifacts that mimic the desired mechanism. If there is a causal process, the concentration-response relationship will help to establish the range in which the process is actually operating.

Biological interpretation also influences the interpretation of the results at the small-molecule discovery programs that are progressing from hit identification to translational candidate selection. The dependency and effective exposure may vary depending on genotype, cell state, tissue architecture, immune tone, metabolic environment and previous treatments. Representative model panels, then, offer more proof than repeated testing in a single easy-to-use system.

There is a need to measure resistance and adaptation in a prospective manner rather than to give them an ex post retroactive explanation. Mechanisms that survive repeated pressure can be identified by serial passage, long-term selection, genetic perturbation, repeated dosing or environmental stressors. A weaker, but more escape-proof contender may have more potential for translation than a weak screener.

Mechanism, when combined with safety assessment, makes safety assessment more informative. The points of concern are: concentration effect on normal cells, cardiac

channels, CYP interaction, concentration effect on immune signals, concentration effect on cardiac channels, concentration effect on genotoxicity, concentration effect off-target pharmacology and dose margin. The comparison of desired engagement versus unwanted biological responses allows discrimination between mechanism related pharmacology versus general cell injury or formulation related toxicity.

Manufacturing is not just a downstream logistics exercise as the steps taken, reaction yield, reagent availability, purification burden, solid form, scale up and robustness of the route will all have an impact on the product presented to the biological systems. Even though the same nominal product name is used, small changes in a process can have an impact on the composition, physical state, amount of impurities, stability, and/or dose reproducibility. Robustness of development is enhanced if attributes of the critical process are directly related to the mechanistic performance.

The computational tools, such as Bayesian optimization, graph model, uncertainty estimation, retrosynthesis prediction, Pareto optimization and information-gain selection, are used to enhance the leverage. The value of prediction is greatest when there is some uncertainty and applicability limitations and particularly when the structures are new, unusual biological states and/or the datasets are sparse. Models that reveal uncertainty can facilitate experiment selection than models that give confident rankings but do not reveal places where uncertainty exists.

Prospective validation calls for prospective cycles, cycles that are withheld, chemical series must be synthesized separately, assays must be orthogonal, resistant models, and success criteria must be predetermined. Orthogonal tests are important because if there are repeated measurements from the same family of assays then the same bias can be reproduced without reinforcing causality. If there is agreement between independent methods, then the design choice to the biological consequence is strengthened; if there is a disagreement, it is an indication of assumptions that need to be modified.

Data should be integrated so that the traceability of the raw data to the development decision is maintained. Chemical structures, process conditions, assay metadata, exposures data, omics features, resistance data, and safety signals should all be made identifiable and versioned. This discipline helps to avoid later misinterpretations of the results that are related to conditions other than those which actually caused the result.

The evaluation of decisions should not be laboratory spectacle, it should be about translation. Score cards that are useful include reproducibility, activity normalised by the exposure, mechanistic confidence, resistance frequency, safety margin, stability and

manufacturing robustness. This will incentivize balanced candidates, as well as provide transparency in making compromises, that can be costly without committing to further development.

The negative results have an exceptionally high value in this context. The limits of a hypothesis are set by the failure to synthesize, inability to achieve the desired target engagement, or instability of the exposure, failure to rescue, resistance emergence, poor batch reproducibility, or loss of activity in realistic models. Documenting those errors will enhance the design of future experiments and avoid re-investing in attractive but incorrect explanations.

The filter of reproducibility across laboratories and across batches is also a must. No matter how statistically significant a result is, a mechanism relying on just one operator, one preparation, one cell line or one modeling option is not free of fragility. When it comes to enduring biology, independent repetition under equivalent types of exposure differentiates the successes from the local and the long-term.

Another important aspect of translation is to have a practical perspective of the clinical and/or deployment context. That a technically complex product can have real-world value depends on factors such as route, administration burden, diagnostic needs, storage, monitoring, cost and timing of treatment. The impact of mechanistic excellence comes to the fore when product design is applied in the context in which biological benefit can only be effected.

Temporal design is an additional layer due to variation in biological system response to short pulses, long exposure, washout and multiple dosing. Using time resolved measurements can distinguish early activity from adaptation downstream and can indicate whether activity continues after the initiating signal has decreased. Sampling times must, as a consequence, be designed to reflect the expected closed-loop medicinal chemistry of new therapeutics and are not necessarily optimised for a single endpoint.

An integrated development record facilitates enhanced communication between medicinal chemistry, formulation science, pharmacology, computation, manufacturing, and clinical planning. The record is not intended to tell one single story, but rather to differentiate between known links, likely explanations, areas of unknown or unknown information, and the next experiment that could get the most relevant information. This transparency helps iteration to be faster, and decreases the pressure to justify early assumptions when there is conflicting data.

Novelty is put into perspective by benchmarking. Whether or not a new approach offers better prediction, mechanism, durability or manufacturability can be determined by established compounds (known inactive controls), standard formulations, baseline models, or accepted analytical methods. Head to head comparison also brings to light the gains that are lost once normalized for exposure, assay sensitivity, batch quality or experimental complexity.

A closed loop can be useful if all the compounds synthesized are designed to be valuable to the decision question, not merely to increase the size of an analogue library. The determining criteria is not if one assay is successful or not, but if the proposed mechanism is consistent in exposure, biological variation, resistance, safety, and production. That standard generates a more disciplined stream of discovery signal to a therapeutic concept that can weather the real-world complexities.