

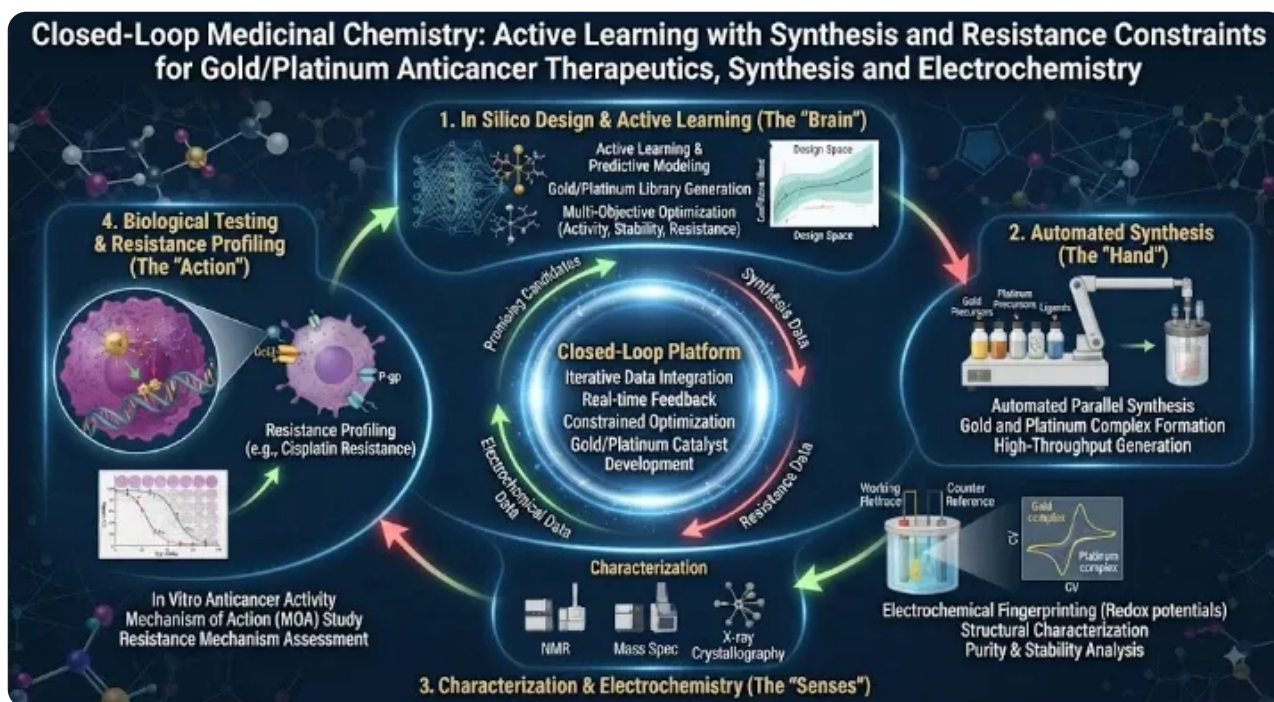
Closed-Loop Medicinal Chemistry: Active Learning with Synthesis and Resistance Constraints for Gold/Platinum Anticancer Therapeutics, Synthesis and Electrochemistry

Md. Humayan Kabir

Jul 19, 2026

Abstract

This article describes closed-loop medicinal chemistry as a learning system for optimizing gold- and platinum-based anticancer compounds. Instead of following a simple design–synthesize–test sequence, the proposed workflow uses active learning to choose compounds that are not only promising but also informative about chemical space, resistance, mechanism, stability, and synthesis feasibility. The piece stresses that metal complexes require descriptors beyond conventional organic medicinal chemistry because coordination geometry, oxidation state, ligand donor atoms, steric shielding, charge, and redox potential strongly affect biological behavior. Synthesis yield, route reliability, purification burden, reagent availability, and scale-up are treated as model constraints rather than afterthoughts. Electrochemical data, target engagement, mitochondrial effects, DNA damage, redox disruption, proteomic responses, and apoptosis timing add mechanistic depth. Resistance models are incorporated early so optimization does not favor fragile compounds. The article also emphasizes uncertainty calibration, Pareto-front trade-offs, standardized experimental data, provenance, adaptive assay allocation, and prospective evaluation to create a practical, experiment-driven discovery loop.



The traditional method for the progress of medicinal chemistry is to build a molecule (design), make it (synthesis), test it (testing), interpret the results (interpretation), and then redesign it (redesign). The process is slow if thousands of possible analogues are generated and compete for limited synthetic resources and biological assays, particularly in metal-based anticancer research where the redox behaviour and ligand exchange introduces extra complexity. Closed-loop medicinal chemistry makes that a learning system. In addition to being chosen for predicted potency, the next compounds to be tested in active learning are chosen for the information needed to understand the resistance, mechanism, stability, and synthesis.

This is especially suitable for gold and platinum scaffolds, due to the fact that chemical space is structured by more than the usual kind of substituent changes. Ligand donor atoms, coordination geometry, chelation, oxidation state, charge and steric shielding and redox potential all influence biological behaviour. We have found that small changes in structure can alter the hydrolytic stability, protein reactivity, subcellular localization, and intracellular lifetime. Any useful optimization model must then be able to include chemical descriptors that are able to describe coordination chemistry, but not every chemical model can.

The problem of selection is solved when active learning is used. A compound that has the strongest predicted activity is not necessarily tested but rather an algorithm is used that balances exploitation and exploration. Some candidates may be selected due to predicted high performance and some due to high uncertainty, and some due to possibility of

improving the model by a poorly sampled chemical region. It is an appealing strategy if there are biological experiments, which take time, and only a small number of virtual candidates can be produced, and synthesis is expensive.

Closed loop starts with synthesis feasibility, which is a realistic closed loop. If the preparation is unreliable, the purification is impractical or sensitive intermediates break down repeatedly, then metal complexes predicted to have excellent anticancer activity are of little value. This means that reaction yield, number of steps, availability of reagent, burden involved in purification and behavior of the reaction on scale-up become model inputs or constraints. This is partly assisted by automated reaction planning, but is still a crucial area for experimental chemist knowledge for unusual coordination chemistry and for redox-sensitive systems.

An additional constraint layer is added in electrochemistry. Gold(III) complexes can be reduced in biological reactions and platinum complexes are frequently activated and/or substituted in a controlled manner. Properties obtained from cyclic voltammetry, stability studies, speciation analysis and reactivity measurements are useful to relate molecular structure to potential intracellular behavior. These measurements are better served as goals rather than being recorded after optimization of potency since they would be useful during active learning.

The B loop must have more than a viability readout. Cytotoxicity, which is useful for ranking, can be caused by many different mechanisms of growth inhibition. Mechanistic labels such as target engagement, mitochondrial effects, DNA damage, redox disruption, proteomic response and apoptosis timing serve as indicators for the identification of compounds with potential interest from the large number of compounds that may be broadly reactive and toxic. A multi-objective model can then be used to identify profiles with both potency and selectivity, an easily tractable mechanism, and reasonable chemical stability.

Routine resistance monitoring should not be a "late clinical surprise. The platinum resistance is often due to a change in transport, glutathione-dependent detoxification, DNA repair, tolerance of DNA lesions or apoptotic signaling. The adaptive routes for gold complexes are different for each target and redox chemistry. Structural changes that are found to work in resistant models identify whether there is something new in the compound or whether there's just a more potent version of the compound with a known liability.

In situations where there is limited resistance data, active learning is particularly useful. A

model can help prioritize experiments for a compound for which the difference between the sensitivity of the parental and resistant cells is most uncertain. The data thus obtained enhance the knowledge of structure-resistance relations. In successive rounds, chemical design progresses towards scaffolds that will remain active against adaptive pressure in a clinically relevant cell population, not towards fragile leads that are best fit for a single sensitive cell population.

Explicit uncertainty calibration is another benefit to closed-loop design. Many neural networks and graph-based models are capable of making accurate predictions in uncharted chemical territory and produce accurate results that look like predictions. The identification of cases where the confidence is low requires the use of ensembles, Bayesian approximations, conformal prediction and distance based applicability measures. If a domain-awareness is necessary, it is for metal complexes, as the training sets are much smaller than many of the datasets of organic molecules, and the danger is that there may be confident extrapolations beyond the represented coordination environments.

The questions for every synthesis in an effective loop are thus: What compound would be most promising to synthesize? and What compound would be most informative to synthesize? Those questions yield contrasting responses. A candidate at an established optimum could have an immediate effect on the performance, whereas if the candidate is uncertain, they could help provide a direction for a new class of ligands, redox window, or geometry. Convenience can't be the only driver of deciding the next analogue as things can be budgeted alongside exploitation and exploration.

Data quality becomes an important design concern. The learning algorithm can be misled by batch effects, inconsistent exposure times, variable assay conditions, varying purities of the metals, solubility artifacts and hidden differences in metal species. The medicinal chemistry platform will thus include standardized protocols, replicate measurements, negative controls, reference complexes and metadata capture. A model trained with noisy labels furthers confusion instead of discovery.

Interpretation of humans is still essential within a closed loop. While algorithms can identify statistical relationships, there is still a need for judgment in the context of plausible coordination, redox pathways, reactive intermediates, and biology in mechanistic chemistry. Chemists are also aware of when a predicted improvement will rely on an unstable motif or an impractical synthetic pathway. The ultimate architecture is a machine-driven prioritization system that is backed by expert-defined constraints and that allows for the open examination of every learning cycle.

The outcome is an alternative culture of optimization. Rather than focusing on individual accomplishments in the area of potency, a program follows the progression of knowledge about chemical space, synthesis, electrochemistry, target engagement, selectivity and resistance. The compounds turn into experiments to eliminate uncertainty and move forward a therapeutic goal. Negative results are still very useful as they reduce the number of possibilities of either mechanism and structure.

Of special interest is the case of multi-objective optimization where properties that are desirable compete. Increased lipophilicity could promote membrane penetration but could also increase the amount of nonspecific binding; increased reactivity could increase target engagement, but also reduce the specificity or selectivity; increased kinetic stability could increase the intactness of the complex, but could also decrease activation. These trade-offs can be useful to visualize without reducing all the properties into an arbitrary score by using Pareto-front analysis. Chemists can then select candidates that are examples of different balanced solutions.

Closed-loop platforms also provide the chance for adaptive assay allocation. Mechanistic experiments are costly, and if they can be predicted on the basis of earlier measurements, they need not be carried out on each of the candidates. There are rapid potency, stability, redox and exposure assays, which can provide input to a triage model followed by chemoproteomics or resistant-cell studies of selected candidates. This hierarchy ensures that information is maintained and costs and experimental throughput are kept in check.

Provenance is also recorded on a successful platform. There is a need for traceable metadata for every compound structure, synthetic route, analytical characterization, assay batch, model version, and for every decision rationale. When algorithms are determining which chemistry will be followed next, it becomes significantly more important to know about the reproducibility. Having these records cleared can help future teams determine whether a candidate was chosen due to the predicted potency, uncertainty reduction, resistance behavior or manufacturing constraint.

The loop can learn from discontinued compounds over time, as well as successful compounds. Boundaries in chemical space are established by a multitude of factors such as poor solubility, rapid resistance, unstable oxidation state, difficult purification, weak target engagement, and narrow selectivity. If those boundaries are captured, then re-investment in areas of structural appeal and lack of productivity is avoided. The accumulated data turns into a strategic asset that links the intuition of chemicals with experimentally proven limits.

The complexity of gold and platinum anticancer chemistry is sufficient to reveal the limitations of the optimization approach, but also provides an excellent testing ground for closed loop discovery. The active learning, synthesis constraints, electrochemical characterization, mechanistic assays, and resistance models are interconnected and constitute a decision system. This is not a system to supplant the intuition of medicinal chemists; rather, it's an experimental engine for the intuition. The outcome is a path towards biologically relevant metal-based leads, not just for their potency.