

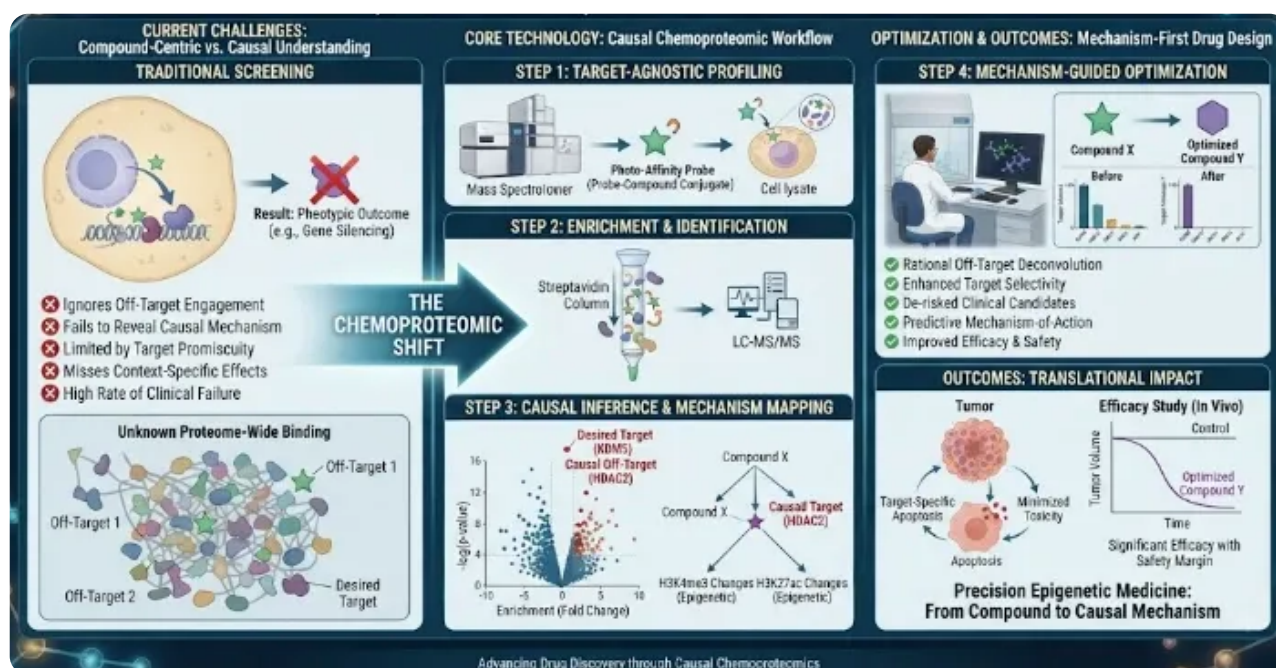
The Compound Is Not the Mechanism: Causal Chemoproteomic Optimization of Epigenetic Drug Discovery, Chemical Biology, Medicinal Chemistry and Proteomics

Md. Humayan Kabir

Oct 26, 2025

Abstract

This article explains why phenotypic activity from an epigenetic compound does not automatically reveal its therapeutic mechanism. It focuses on a causal chain connecting direct protein engagement, downstream pathway remodeling, chromatin effects, and phenotype. Competition-based profiling, thermal stability methods, covalent capture, genetic perturbation, rescue experiments, photoaffinity probes, and dose-resolved phenotyping are presented as complementary tools for separating causal targets from collateral interactions. The article stresses that no single assay is decisive; confidence grows when independent methods converge across dose, time, model system, and biological context. Molecular subtype, pathway dependence, target abundance, and treatment history are highlighted as important sources of context dependence. Probe artifacts, incomplete proteome coverage, off-target liabilities, and changes in target dependence are treated as key risks. The framework also integrates medicinal chemistry, scalable synthesis, permeability, metabolic stability, safety, prospective computational validation, and traceable analytical control. The central message is that mechanistic confidence should become an optimization endpoint alongside affinity and efficacy, with each candidate tested through falsifiable causal evidence.



The new challenges of modern therapeutic development are to understand the reasons for a result and to determine when it is reliable, and causal chemoproteomics is emerging as a way to go beyond mere questions of potency, delivery, or association. A mechanistic core is at the heart of this shift, comprising direct protein interactions, downstream pathway modification, chromatin modifications and phenotype. That makes the network significant since a compound can have an attractive phenotype and have several proteins that it works through, and the actual mechanism of action remains to be determined. A more robust development strategy then considers mechanism, context, and uncertainty as interwoven design variables and not separate design activities occurring at different phases of development.

However, chemical proteomics has proven to be particularly useful when the epigenetic compound is broadly phenotypically active with uncertain target attribution. Competition based profiling, thermal stability methods, covalent capture strategies, and genetic rescue experiments aid in the distinction between direct engagement from downstream chromatin remodeling. The proteomic data also suggest additional interactions that are not captured by traditional target panels, especially if a scaffold is targeting multiple enzymes or protein complexes that share a similar function. The establishment of a causal workflow thus provides a much firmer foundation than activity alone to optimize for medicinal chemistry.

This has obvious implications for a new way of discovery based on photoaffinity or clickable probes, competition proteomics, thermal stability profiling, genetic

perturbation, rescue experiments, and dose-resolved phenotyping. None of these tools can be used for a specific decision in isolation – each only measures a portion of a larger biological or clinical process. Confidence increases with consistent conclusions drawn from independent measurements across the dose, time, model system and relevant biological context. It is important to note that discordant results do not necessarily mean that they are of no value, as the lack of consistency can sometimes be due to hidden mechanisms, assay artefacts, population differences or exposure constraints, and should be explored before further development of the assay becomes too costly.

The key question for decisions is what engaged protein is required to provide the desired response and which interactions occur just in the presence of the protein. In order to answer that question, it is necessary to make explicit comparisons to one's intended mechanism and plausible alternatives to it, such as off-target activities, parallel pathways, differential exposure, or measurement bias. Perturbation experiments are particularly useful since the removal, blocking or restoration of a proposed component can determine whether the observed effect is dependent on that component. Reliability or the "feel good factor" is gained, therefore, when successful challenge tests are conducted, not when repeated testing is performed in a single assay that is favorable.

Almost all the results in causal chemoproteomics for epigenetic drug discovery are context dependent. Relevant variations include the molecular subtype, pathway dependence, abundance of the target, and treatment history. An intervention that is effective under one set of these conditions may have a less potent effect, be more toxic, or cause a different effect in another set of conditions. When using stratification, it is useful to relate variables to mechanism or exposure to the changes in performance of treatment.

Risk assessment is also very important, since the probe-induced artifact, lack of coverage of the proteome, off-target liabilities and changing target dependence due to cellular context. It is most powerful to support safety work with the same mechanistic approach employed for efficacy to pinpoint unwanted effects to target biology, carrier behavior, exposure, metabolites, or analytical bias. This helps minimise the risk that safety is seen as an afterthought in the checklist unrelated from discovery logic. It also allows for more transparent trade-offs for more effective efficacy when either more exposure is needed, more pathway modulation is needed, or a longer persistence in sensitive tissue is required.

The extra layer added to this is the one of translational development, a process focused on optimizing binding to the target protein, instead of binding to its receptor. Experiments that recapitulate clinically relevant timing, range of concentrations, barriers, treatment

history, and sources of heterogeneity are more informative for preclinical success. It is not especially helpful to obtain a "complex" model from a "simple" model, but it is helpful to obtain a series of models that will provide answers to specific uncertainty questions. For each step thus, a specific decision has to be made: continue, redesign, restrict the target population, adjust doses, sample a different kind of mechanistic information.

In addition to manufacturing, the analytical control is also essential because a variety of factors such as chemical handles, stereochemistry, permeability, metabolic stability, and scalable synthesis affect interpretable chemoproteomic data. If the assay inputs are not well defined or if the material tested is different from batch to batch, a biological explanation becomes less useful. Therefore, critical attributes must be set to a range, measured with a high level of sensitivity and documented to link process conditions to biological activity. This discipline is particularly critical for complex molecules, for nanoparticles, biomaterials, and for studies with a large amount of data, where the change in exposure, signaling, or apparent treatment response can be subtle.

A better distinction between prediction and confirmation in the analysis of data also plays a role in causal chemoproteomics in epigenetic drug discovery. While computational models, machine learning, causal estimators and mechanistic simulations can rank hypotheses, combine high-dimensional measurements, and quantify uncertainties, the output of models is not a form of truth. Useful models include the statements of assumptions, the quantification of the calibration, flagging poorly supported regions, and the suggestions of an experiment for the next most informative experiment that will influence a decision. Prospective validation is still crucial as retrospective fit can be beneficial for a pattern that doesn't take hold in new populations, new batches, or new biotic conditions.

The other significant change is in regards to time as several processes that are essential to the causal chemoproteomics approach in epigenetic drug discovery occur following the initial exposure. There is not only a wide variety of possible pathways to early response, adaptation, clearance, immune programming, tissue remodeling, adherence, or resistance, but even if a given endpoint is perceived as favorable, the pathway can be quite different. Longitudinal measurement indicates if an intervention results in a stable state, a brief disturbance or a liability in the future. Time resolved data also facilitate causal interpretation because they allow us to determine if the proposed mechanism occurs prior to the downstream phenotype or after the appearance of the phenotype.

If each study is linked to a specific uncertainty budget, then experimental design can be

enhanced and become more efficient. Novelty is not necessarily the deciding factor; questions about direct protein engagement, remodeling of downstream pathways, chromatin effects, and phenotype can be prioritized by the ability to shift what protein engages what to produce the desired response and which interactions are merely 'along for the ride' when exposure to the protein occurs. Replication across complementary models is then used to determine the robustness of the conclusion, in the sense that it does not rely on one assay, one laboratory condition or one population. Such discipline also avoids wasted optimization and informs designs of the next cycle because for every failed prediction, the mechanism space is reduced and hence the next cycle is improved.

Traceability between claims and measurements is also necessary for the regulatory and clinical credibility. In the context of epigenetic drug discovery, a causal chemoproteomics development package is more robust when causing, processing, model assumptions, protocol deviations and attributes of the material are tied together from early experiments to the translation. The traceability helps to differentiate between true biological variation and technical noise and makes it easier to review safety or effectiveness claims later. It also provides a more solid foundation for revising conclusions with new data when they affect an earlier assumption.

More precision is also needed in communicating around this field. The usage of the terms targeted, selective, durable, safe, predictive, or personalized, without a measurement, comparator, population, and uncertainty range is of little value. Clear reporting helps to make negative results more useful and makes exploratory findings look like they are not mechanisms. Precision provides scientists, scientists and developers, clinicians and regulators with a common metric to evaluate if a claim is based on direct measurement, calibrated inference or an untested assumption.

Thus, the best direction for new technologies in the field of causal chemoproteomics for epigenetic drug discovery is a closed loop, rather than a single new chemistry technology, between chemistry or formulation, mechanistic measurement, quantitative modeling, safety, manufacturing control and clinically relevant context. The failure of assumptions speeds up the process of progress, and the success of mechanisms survives planned difficult tests. The logic of mechanistic confidence becomes an optimisation end-point along with affinity and efficacy. At the level of the field, the most important innovation is a development system that is able to learn from each experiment, while maintaining the interpretation of cause and effect, the ability to be reproduced, and to be relevant to the real world as central concerns of all decisions it makes.