



Can Natural Products Drug the Dark Proteome by Exploiting Uncharacterized Binding Sites, Structural Disorder, and Previously Untapped Protein Biology?

Olivia Bennett^{1*}, Harry Collins², Jack Foster¹, Ethan Wright³

¹Department of Dark Proteome Drug Discovery, Faculty of Pharmacy, University of Nottingham, Nottingham, United Kingdom.

²Department of Uncharacterized Binding Sites and Structural Disorder, Faculty of Pharmacy, University of Southampton, Southampton, United Kingdom.

³Department of Previously Untapped Protein Biology, Faculty of Pharmacy, Newcastle University, Newcastle, United Kingdom.

ABSTRACT

Large regions of the human proteome remain poorly developed as therapeutic targets, but limited chemical annotation, sparse functional knowledge, missing structures, and conformational heterogeneity should not be conflated with demonstrated undruggability. This conceptual article asks whether natural products can expand access to such "dark" protein space by exploiting recognition modes that conventional target-first discovery may undersample. The analysis separates several forms of darkness: missing biological evidence, absent or incomplete structural information, intrinsic disorder, transient or cryptic pockets, and insufficient target-validation tools. It then considers natural products not as intrinsically privileged dark-proteome ligands, but as chemically and biologically distinctive perturbagens whose complex scaffolds, stereochemistry, and phenotype-first discovery history may broaden the search for conditional protein recognition. A proposed framework links evidence-state classification, ensemble-aware structural reasoning, cryptic-site detection, natural-product prioritization, phenotypic discovery, chemoproteomic target deconvolution, and staged credibility testing. The central argument is that a protein should become more credible as a dark-proteome target only when computational plausibility is progressively replaced by orthogonal evidence of binding, cellular engagement, functional modulation, specificity, and mechanism. Important boundaries remain: predicted pockets are not measured pockets, binding is not mechanism, phenotypic activity may be polypharmacological or artifactual, and unusual chemistry does not establish privileged recognition. Natural products may therefore be most valuable as structured probes of unknown biology, provided uncertainty and falsification remain explicit. This strategy remains hypothesis-generating and requires target-specific experimental validation.

Key Words: Dark proteome, Natural products, Intrinsically disordered proteins, Cryptic pockets, Chemoproteomics, Target deconvolution

eIJPPR 2026; 16(2):145-153

HOW TO CITE THIS ARTICLE: Bennett O, Collins H, Foster J, Wright E. Can Natural Products Drug the Dark Proteome by Exploiting Uncharacterized Binding Sites, Structural Disorder, and Previously Untapped Protein Biology? Int J Pharm Phytopharmacol Res. 2026;16(2):145-153. <https://doi.org/10.51847/dTtZPCgu9k>

INTRODUCTION

Drug discovery has historically concentrated on proteins for which biology, assays, chemical matter, and tractable mechanisms already reinforce one another. This creates a selection effect: well-studied targets accumulate ligands and disease evidence, whereas less-developed proteins can

appear intrinsically resistant to pharmacology simply because the tools needed to interrogate them are missing. Human target-development landscapes, the Target 2035 chemical-probe agenda, and Pharos all support a more cautious interpretation in which a substantial portion of protein space remains understudied or unprobed rather

Corresponding author: Olivia Bennett

Address: Department of Dark Proteome Drug Discovery, Faculty of Pharmacy, University of Nottingham, Nottingham, United Kingdom.

E-mail: olivia.bennett@nottingham.ac.uk

Received: 12 December 2025; **Revised:** 09 March 2026; **Accepted:** 11 March 2026

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



than experimentally proven inaccessible [1–3]. For this article, “dark” therefore denotes an evidence condition, not a molecular verdict. A protein may be dark because its function is poorly resolved, because no selective chemical probe exists, because structural states relevant to binding are unknown, or because these deficits coincide.

This distinction matters because improvements in proteome characterization do not automatically convert proteins into validated therapeutic targets. The Human Proteome Project continues to refine evidence for protein existence and characterization, yet protein detection, functional annotation, disease relevance, and medicinal-chemistry tractability remain different evidentiary questions [4]. The conceptual opportunity considered here is consequently narrower than claiming that unknown proteins are broadly druggable. It is whether natural products, coupled to modern structural computation, phenotypic screening, and chemical biology, can create experimentally testable entry points into proteins whose biology or recognition states have been inadequately explored. The corresponding burden of proof is high: novelty must arise from reproducible protein engagement and interpretable biology, not merely from unusual chemical structures, predicted pockets, or unexplained phenotypes.

What makes a protein “dark” to drug discovery?

Protein darkness is better represented as a vector of missing evidence than as a single category. A target may have a high-confidence structure but weak disease linkage;

conversely, genetics or cell biology may implicate a protein for which no established ligand, binding site, or tractable assay exists. Resources that integrate target–disease, genetic, omic, literature, and drug evidence illustrate why structural availability alone cannot establish therapeutic relevance [5]. In the proposed terminology used here, biological darkness refers to unresolved function, causal disease relevance, or pathway position; chemical darkness to absent selective ligands or probes; structural/dynamic darkness to missing or misleadingly incomplete information about recognition states; and validation darkness to inadequate evidence connecting molecular engagement with biological response. These dimensions can coexist, but they should not be treated as interchangeable.

Artificial intelligence may reduce some of this sparsity by integrating heterogeneous target evidence, inferring structural hypotheses, or ranking candidate interactions. Yet contemporary target-assessment analyses also emphasize training-data bias, incomplete annotation, causal ambiguity, and the need for prospective biological validation [6]. Accordingly, a high model score does not make a dark protein less biologically dark unless the prediction leads to new evidence. The practical question is not whether a target is “dark” in the abstract, but which evidence layer is missing and which experiment could resolve it. **Table 1** organizes these distinctions as proposed evidence-state constructs so that progress in one dimension is not mistaken for resolution of another.

Table 1. Proposed evidence-state constructs defining different forms of protein darkness

Construct	Operational meaning	Principal unresolved evidence	What successful resolution adds	What it does not establish
Biological darkness	Function, pathway role, or disease relevance remains uncertain	Causal biology and context-specific relevance	A stronger biological target hypothesis	Ligandability or therapeutic efficacy
Chemical darkness	Selective chemical matter is absent or inadequate	Probe selectivity and reproducible engagement	An experimentally controllable perturbation	Exclusive mechanism or clinical usefulness
Structural/dynamic darkness	Relevant recognition states are missing, transient, or incompletely represented	Accessible conformations, interfaces, pockets, or ensembles	A state-aware structural hypothesis	Binding, affinity, or cellular accessibility
Validation darkness	A nominated interaction is insufficiently connected to function	Orthogonal engagement, specificity, and functional evidence	Greater confidence in target attribution	Universal mechanism or translational readiness

Note: These constructs are proposed here as evidence states rather than fixed protein classes. Their boundaries may shift as new biological, chemical, structural, or validation evidence becomes available.

Missing structural information versus missing biological knowledge

The distinction between missing structure and missing biology changes how a discovery program should proceed. Target-development resources classify proteins using chemical, biological, and clinical evidence that can remain sparse even when sequence or structural models are

available [3]. Structural darkness therefore concerns what conformations, pockets, interfaces, or assemblies are physically accessible; biological darkness concerns whether perturbing the protein matters for the phenotype of interest. A solved or predicted structure can reduce one uncertainty while leaving the other untouched. Conversely, strong disease genetics or pathway evidence can motivate

experimentation without revealing where a small molecule should bind.

This separation also prevents structural sophistication from outrunning therapeutic reasoning. Integrated target–disease platforms support hypothesis building from convergent biological evidence rather than from structure alone [5]. A visually convincing model of a pocket on a poorly annotated protein may therefore be medicinally interesting while remaining biologically underjustified. The reverse problem is equally important: a compelling disease target can be chemically inaccessible under current methods without becoming biologically unimportant. Dark-proteome discovery should thus ask two independent questions—whether there is a plausible controllable molecular state and whether controlling that state would answer a biologically meaningful question—before combining them into a target hypothesis.

AI-assisted integration does not erase this distinction. Models trained on known targets, literature, interaction networks, or molecular structures may reproduce the same historical asymmetries that made a target dark in the first place [6]. For that reason, the proposed framework treats computational ranking as an uncertainty-reduction step rather than as evidence closure. A prediction is most useful when it specifies what should be measured next: a conformational state, a binding event, a phenotype, or a disease-relevant consequence. Where biology remains annotation-limited, the model should expose that deficit rather than conceal it behind a unified score.

Intrinsic disorder and conditional structure

Intrinsic disorder challenges the assumption that ligandability requires a persistent, preorganized pocket. Reviews of intrinsically disordered proteins (IDPs) and biomolecular condensates document examples in which small molecules modulate dynamic proteins or assemblies, while also emphasizing difficulties in affinity measurement, selectivity, conformational interpretation, and mechanism [7]. The appropriate conceptual shift is therefore not from “undruggable” to “druggable,” but from static structure to conditional recognition. A disordered region may transiently favor conformations, interaction motifs, or local environments that permit ligand engagement under some conditions and not others. Such possibilities are target-specific and should not be generalized across the IDP proteome.

An ensemble view is essential because IDP function can depend on distributions of conformations and on environmental context rather than a single dominant structure [8]. In this setting, a ligand may shift an ensemble, stabilize a minority state, interfere with a partner interaction, or remain dynamically associated without generating a conventional rigid complex. The resulting

recognition problem is simultaneously structural and thermodynamic: identifying a plausible conformation is insufficient unless its population, interconversion, and response to ligand binding are compatible with measurable pharmacology. Disorder therefore creates both opportunity and uncertainty, and computational representations intended for these targets should preserve state heterogeneity rather than substitute a single predicted structure for an unresolved ensemble.

Direct studies illustrate that these recognition modes are heterogeneous rather than interchangeable. A small molecule can remain rapidly dynamic while bound to an IDP; the c-MYC transactivation domain can exhibit a ligand-addressable conformational switch; and enhanced sampling has been used to identify ligand-compatible states in the intrinsically disordered androgen-receptor N-terminal domain [9–11]. These examples support the existence of conditional ligand recognition in specific systems, not a general claim that disordered proteins are broadly tractable. For natural-product discovery, their importance is methodological: chemically unusual ligands should be tested against ensembles and functional state changes rather than judged solely against a static-pocket paradigm.

Uncharacterized pockets and ligand-induced ordering

Even folded proteins can conceal recognition sites when crystallographic or predicted structures capture only closed or apo-like conformations. PocketMiner demonstrates that machine learning can prioritize candidate cryptic-pocket locations from single protein structures, thereby identifying sites that are not obvious in the observed conformation [12]. Such predictions expand the structural search space, but they do not measure how often a pocket opens, whether a ligand can access it in a biological environment, or whether binding there alters function. A cryptic pocket is therefore best treated as a structural hypothesis whose value depends on subsequent evidence rather than as an automatically ligandable feature.

Feasibility becomes a separate question once a candidate site has been nominated. Analyses of cryptic sites emphasize that geometric detectability must be distinguished from accessibility, ligandability, state population, and pharmacological consequence [13]. This distinction also accommodates different mechanisms of pocket appearance: a ligand may select a pre-existing minority conformation, stabilize an intermediate, or contribute to local ordering after initial contact. For a dark target, these alternatives matter because they change what experimental result would count as confirmation. A bound pose alone cannot reveal whether the relevant state is sufficiently populated, kinetically accessible, or functionally coupled.

Computational strategies can broaden this search by combining molecular dynamics, enhanced sampling, structure prediction, and machine learning, yet their strengths are conditional on sampling depth, training coverage, and the conformational regime of the target [14]. The most defensible role for computation is therefore comparative prioritization: identifying states or regions worth testing, exposing alternative conformations, and highlighting uncertainty rather than certifying hidden ligandability. This distinction will become particularly important when unusual natural-product scaffolds are evaluated against structural states for which little or no ligand precedent exists.

Recent work increasingly frames cryptic-pocket discovery as an ensemble problem in which physics-based simulations and artificial intelligence provide complementary information [15]. Rare-state thermodynamics remain a central limitation. That boundary is especially important for the present thesis: natural products may be worth testing against unusual or transient recognition states, but computational plausibility alone cannot establish that such states are sufficiently populated, selectively engaged, or pharmacologically useful in cells. A credible dark-proteome program must therefore treat pocket detection as the beginning of an experiment, not the completion of target validation.

Why natural products may explore different recognition space

Natural products are relevant to dark-proteome discovery because they enlarge the kinds of molecular recognition worth testing, not because their origin guarantees privileged binding. Their scaffolds often combine stereochemical complexity, ring systems, hydrogen-bonding patterns, and functional-group arrangements that differ from simplified screening collections, while modern natural-product discovery has increasingly coupled these features to synthesis, analytics, and target-oriented biology [16]. The proposed advantage is therefore exploratory: unfamiliar chemistry may sample protein surfaces, dynamic conformations, or interaction topologies that are sparsely represented by established ligands. This proposition must remain comparative and testable. Structural diversity cannot by itself establish higher affinity, selectivity, cellular exposure, or preferential activity against dark proteins.

Contemporary natural-product chemistry makes this hypothesis experimentally actionable because complex

metabolites can increasingly be isolated, resynthesized, derivatized, and deployed as mechanistic probes rather than treated only as end-stage leads [17]. That capability is important when the initial value of a compound is the biology it reveals. However, chemical complexity can also complicate resupply, optimization, permeability, and causal interpretation. Labeled and label-free target-identification strategies offer complementary routes for connecting natural-product activity to protein engagement, while each carries biases related to probe modification, abundance, stabilization, or indirect interactions [18]. Thus, the relevant search advantage is not “naturalness”; it is the combination of differentiated chemistry with methods capable of determining whether an unusual phenotype actually arises from a reproducible protein interaction.

Phenotypic discovery as an entry point into dark protein biology

Phenotypic discovery is particularly useful when biological darkness precedes target identity. Instead of requiring a validated protein before screening, a compound-induced cellular state can become the starting observation from which target hypotheses are subsequently reconstructed. A recent study illustrates this logic by coupling phenotypic screening to chemoproteomic target deconvolution and additional mechanistic testing, moving from an unexplained response to a specific, experimentally interrogable target hypothesis [19]. The transferable lesson is procedural rather than target-specific: phenotype can open an otherwise inaccessible biological question, but the phenotype itself does not identify the responsible protein, distinguish direct from indirect effects, or establish that a newly implicated target is therapeutically useful.

For natural products, this inversion of the conventional target-first sequence is attractive because bioactivity may be discovered before a tractable molecular explanation exists. Yet unknown mechanism should be treated as an unresolved evidence state, not as proof of novelty. Complementary target-identification approaches can narrow that uncertainty, but probe perturbation, indirect stabilization, abundance effects, and nonspecific interactions remain possible [18]. The conceptual distinction is summarized in **Table 2**: a discovery feature may justify further investigation without licensing the stronger interpretation that often follows it.

Table 2. Proposed distinctions between natural-product discovery opportunities and stronger claims requiring validation

Discovery feature	Immediate contribution	Permitted interpretation	Stronger claim still requiring evidence
-------------------	------------------------	--------------------------	---

Chemical distinctiveness	Broadens candidate recognition chemistry	Alternative binding hypotheses are worth testing	Preferential dark-proteome ligandability
Conditional structural recognition	Allows testing of dynamic or transient states	Static-pocket absence may be insufficient	Functionally relevant binding in cells
Phenotype-first activity	Reveals perturbable biology before target identity	A target-deconvolution program is justified	The phenotype identifies its direct target
Target nomination	Connects activity to candidate proteins	Candidate engagement can be prioritized	Exclusive mechanism
Orthogonal target identification	Reduces method-specific ambiguity	Confidence can increase through convergence	Therapeutic relevance or translational readiness

The mature synthesis developed across the manuscript is shown in **Figure 1**: dark-proteome discovery becomes informative only when distinct sources of darkness are connected to conditional recognition and natural-product

exploration through explicit experimental bridges, while specificity and target credibility remain a boundary rather than an automatic endpoint.

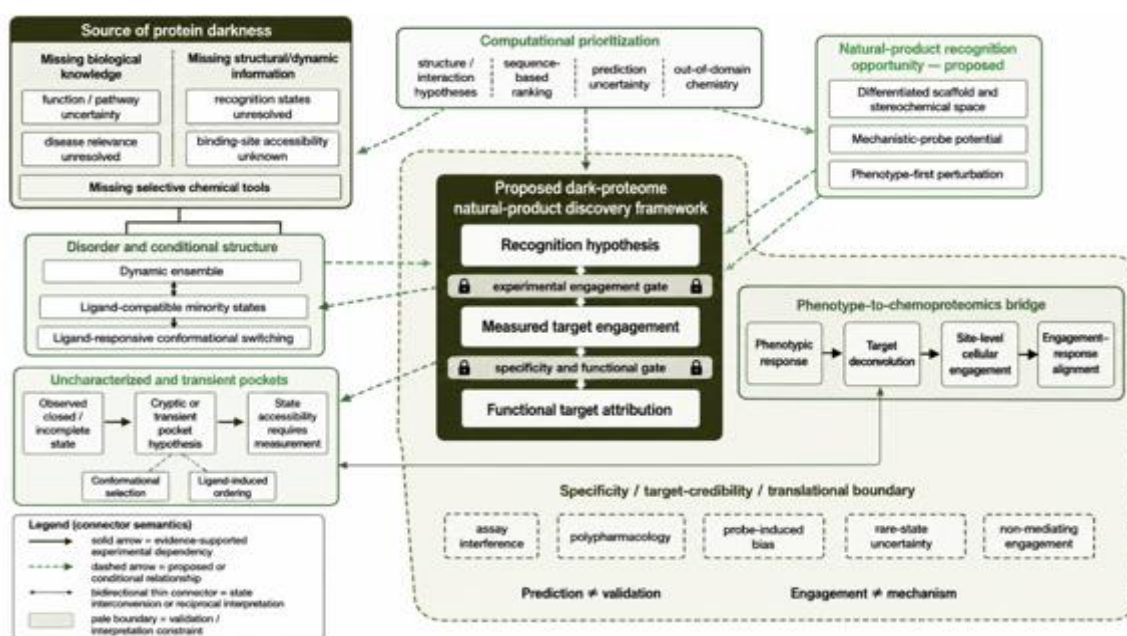


Figure 1. From dark-proteome uncertainty to a credible natural-product target hypothesis

Computational prioritization without established ligands

Computation is most valuable for dark targets when it converts missing information into ranked experiments rather than simulated certainty. Modern interaction-structure prediction can generate plausible biomolecular complex geometries even where experimentally solved complexes are absent, but predicted coordinates remain model outputs rather than measurements of binding, affinity, or function [20]. Sequence-based drug-design approaches provide an additional route when structural or ligand information is sparse and have demonstrated that protein sequence can support actionable compound prioritization in selected settings [21]. These capabilities expand the hypothesis space, yet neither supplies a substitute for biochemical or cellular evidence.

Ranking should also communicate what the model does not know. Evidential drug–target interaction methods illustrate how uncertainty can be represented alongside predictions instead of reducing each candidate to an unqualified score [22]. For natural products, domain of applicability is especially important because chemically distinctive scaffolds may occupy regions poorly represented in training data; molecular deep-learning performance can degrade near the edge of familiar chemical space [23]. A dark-target workflow should therefore prioritize compounds through ensembles, sequence, predicted structures, or interaction models only when confidence, training-domain distance, and competing structural hypotheses remain visible. Computational prioritization is a resource-allocation device: its successful output is a better experiment, not a validated target.

A dark-proteome natural product discovery framework

The framework proposed here begins by diagnosing the source of darkness before choosing a discovery route. Biological darkness asks whether function or disease relevance is unresolved; structural/dynamic darkness asks whether ligand-responsive states are unknown; chemical darkness asks whether selective perturbagens are missing; and validation darkness asks whether engagement has been connected to function. Integrated target-evidence systems support the premise that therapeutic hypotheses require multiple evidence streams [5], but the ordering and connection of these darkness dimensions are proposed here. A target should not advance merely because one layer becomes information-rich while the others remain opaque. The second stage searches for conditional recognition opportunity. For ordered proteins this may involve conventional or cryptic pockets; for disordered regions it may involve state ensembles, conformational switches, or interaction motifs. Because cryptic-site feasibility depends on more than geometric detectability [13], natural-product chemical diversity is inserted as a proposed search expansion rather than a guaranteed solution. Candidate metabolites or derivatives can enter through target-first structural prioritization or phenotype-first perturbation. These routes are reciprocal: a predicted state can motivate compound testing, while an unexplained phenotype can motivate target deconvolution and retrospective structural interpretation.

The third stage imposes an evidence transition. Computational scores remain upstream of measured engagement, particularly when chemistry lies outside familiar training distributions [23]. Likewise, phenotype-first discovery creates a target hypothesis only after deconvolution and orthogonal testing [19]. The framework therefore advances candidates by replacing one uncertainty at a time: prediction with measurement, target nomination with direct engagement, engagement with functional modulation, and apparent selectivity with competing-explanation testing. Failure at any gate is informative; it can reject the candidate target, redirect the structural hypothesis, or reveal that the phenotype arises from polypharmacology rather than a single dark protein.

This staged architecture also prevents evidence gained in one domain from silently substituting for another: a clearer structure cannot compensate for uncertain disease biology, and a compelling phenotype cannot compensate for uncertain molecular attribution. Advancement therefore depends on cross-domain convergence rather than accumulation of evidence within only one familiar layer.

Chemoproteomic strategies for revealing unexpected engagement

Proteome-scale chemical biology provides the experimental bridge between unusual activity and unexpected protein engagement. Activity-based protein profiling can identify ligandable proteins or sites in native biological contexts without restricting discovery to a preselected target panel, although coverage remains determined by probe chemistry and residue accessibility [24]. High-resolution cellular chemoproteomic approaches can further localize small-molecule interaction sites, strengthening physical-engagement evidence while still stopping short of proving that the mapped site causes the phenotype [25]. These methods are especially useful for dark targets because they can generate target evidence rather than merely interrogate already annotated targets. No single chemoproteomic readout should be treated as definitive. Complementary target-deconvolution methods, stability-based proteomics, and approaches linking cellular engagement to downstream pharmacology can triangulate a candidate interaction more credibly than one assay alone [26–28]. Their failure modes are also complementary: affinity capture may be altered by probe derivatization, stability changes may be indirect, and engagement–response concordance may still coexist with network-level or polypharmacological effects. The proposed framework therefore values convergence across orthogonal measurement principles, ideally in the biological context where the phenotype occurs. Non-detection remains ambiguous when assay chemistry or proteome coverage is incomplete.

Distinguishing new target biology from nonspecific binding

Dark-target discovery is unusually vulnerable to circular interpretation: a surprising phenotype suggests an unknown target, and the target appears important because the phenotype was surprising. Chemical-probe practice provides a warning against this logic. Systematic analysis of cell-based biomedical research has shown that suboptimal probe use can undermine target-mechanism interpretation [29]. A newly nominated dark protein should therefore face stronger controls, not weaker ones, because prior biological knowledge cannot rescue ambiguous pharmacology. Reproducibility across analogs or orthogonal perturbations, cellular exposure, concentration dependence, and target-relevant controls should be used to test whether the interaction survives alternative explanations.

Assay interference, aggregation, reactivity, and promiscuity are also plausible explanations for apparent novelty and must be actively falsified before new target biology is inferred [30]. Even a reproducible binding event remains only one layer of the argument. Cellular target-engagement measurements can be related to downstream

pharmacology, but concordance does not establish that the target is the exclusive mediator of response [28]. **Table 3** therefore treats credibility as a multidimensional evidence

problem in which binding, engagement, function, specificity, and mechanism remain analytically separate.

Table 3. Proposed framework dimensions for evaluating an unexpected dark-protein engagement

Dimension	Core question	Credibility-increasing observation	Principal competing explanation
Physical binding	Does the compound interact reproducibly with the protein?	Orthogonal binding evidence	Assay artifact or promiscuity
Cellular engagement	Is the target occupied in the relevant cellular context?	Exposure-consistent engagement	Indirect stabilization or probe bias
Site assignment	Is a specific interaction region supported?	Reproducible site-level evidence	Accessibility or labeling bias
Functional modulation	Does engagement alter target-relevant biology?	Concordant target-linked response	Network-level secondary effect
Specificity	Are plausible alternatives excluded?	Orthogonal perturbations and selectivity controls	Polypharmacology
Mechanism	Does target perturbation explain the phenotype?	Convergent causal evidence	Correlated but non-mediating engagement

What evidence would make a dark target credible?

A credible dark target requires an evidence chain rather than a single decisive assay. Site-resolved cellular chemoproteomics can strengthen the first link by demonstrating physical engagement in a biologically relevant context [25]. The next question is whether engagement occurs under exposures and conditions associated with the downstream response; approaches that couple cellular engagement to pharmacology can make that relationship directly testable [28]. The proposed hierarchy therefore progresses from reproducible binding to cellular engagement, then to functional modulation and target attribution. Each step answers a different question and should be allowed to fail independently.

Credibility also depends on falsification. Probe-quality problems show why nominal selectivity, convenient concentrations, or a single chemical tool cannot sustain a mechanistic claim [29]. Likewise, cryptic-site feasibility varies across targets, so evidence requirements should remain context-specific rather than being reduced to universal geometric or affinity thresholds [13]. For a disordered protein, structural evidence may need ensemble-sensitive measurements; for a cryptic pocket, state accessibility may be central; for a phenotype-first target, genetic or orthogonal chemical perturbation may be especially informative. The framework is therefore graded: greater convergence justifies greater confidence, but no fixed checklist converts a dark target into a clinically validated one.

Conceptual and translational boundaries

The largest conceptual risk is converting computational possibility into quantitative biological reality. A 2026 benchmark comparing AI- and physics-based approaches with experimentally characterized cryptic-pocket populations found that relative trends could sometimes be captured while absolute opening probabilities, particularly for rare states, remained unreliable [31]. That result does not invalidate cryptic-pocket computation; it limits the claim it can support. Ranking a state for testing is different from measuring its population, just as predicting a ligand-compatible pose is different from demonstrating cellular engagement.

The same restraint applies to the central natural-product thesis. Natural products provide diverse chemistry and powerful discovery precedents, but the available evidence does not establish that they are systematically enriched for dark-proteome ligands [16]. The proposed recognition-space advantage must therefore remain falsifiable. If natural products do not outperform appropriate comparator chemistry for a given conditional state, the framework should accommodate that negative result rather than reinterpret it as hidden success. More broadly, poor probe practice can produce convincing but unstable biological narratives [29]. Translation should begin only after target-specific evidence survives these alternative explanations.

Table 4 makes the final boundary explicit by separating plausible discovery mechanisms from the observations still required before stronger interpretation.

Table 4. Proposed synthesis of mechanistic opportunities and boundary conditions in natural-product exploration of dark protein space

Discovery mechanism	Plausible opportunity	Boundary condition	Evidence needed before stronger interpretation
---------------------	-----------------------	--------------------	--

Dynamic ensemble recognition	Ligand engages a conditional IDP state	Ensemble population and function are context-dependent	State-sensitive binding plus functional perturbation
Cryptic-pocket recognition	Ligand selects or stabilizes a transient cavity	Detected geometry may be rarely accessible	Experimental accessibility, engagement, and functional coupling
Natural-product search expansion	Distinct chemistry samples underexplored recognition space	Diversity does not imply privileged binding	Comparative target-specific testing
Computational prioritization	Models rank states or compound–target hypotheses	OOD chemistry and rare states reduce confidence	Prospective experimental validation
Phenotype-first discovery	Bioactivity exposes previously untapped biology	Phenotype may be indirect or polypharmacological	Target deconvolution and orthogonal perturbation
Chemoproteomic engagement	Proteome-scale assays reveal unexpected interactions	Engagement may be indirect or non-mediating	Site, specificity, and engagement–response evidence

CONCLUSION

The dark proteome is not a single reservoir of hidden drug targets. It is a collection of different evidence deficits, some biological, some chemical, some structural, and some created by insufficient validation. The conceptual contribution developed here is a proposed discovery logic in which these deficits are diagnosed separately and then connected experimentally. Natural products may be useful because differentiated chemistry and phenotype-first discovery can expose recognition states or biological relationships that conventional target-first workflows undersample, but this remains a testable advantage rather than an intrinsic property of natural products.

The practical standard is progressive replacement of uncertainty with evidence. Dynamic or cryptic structures should motivate experiments, not claims of ligandability; phenotypes should motivate deconvolution, not immediate target assignment; binding should motivate functional testing, not automatic mechanism. Computational models can prioritize this process, and chemoproteomics can reveal unexpected engagement, but neither removes the need to exclude nonspecificity, polypharmacology, assay artifacts, and annotation bias. Unknown biology should therefore be investigated aggressively but interpreted conservatively. A dark protein becomes scientifically valuable not when it is newly labeled “druggable,” but when reproducible perturbation makes its previously inaccessible biology testable.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Oprea TI, Bologa CG, Brunak S, Campbell A, Gan GN, Gaulton A, et al. Unexplored therapeutic opportunities in the human genome. *Nat Rev Drug Discov.* 2018;17(5):317-32. doi:10.1038/nrd.2018.14
- [2] Carter AJ, Kraemer O, Zwick M, Mueller-Fahrnow A, Arrowsmith CH, Edwards AM. Target 2035: Probing the human proteome. *Drug Discov Today.* 2019;24(11):2111-5. doi:10.1016/j.drudis.2019.06.020
- [3] Kelleher KJ, Sheils TK, Mathias SL, Yang JJ, Metzger VT, Siramshetty VB, et al. Pharos 2023: An integrated resource for the understudied human proteome. *Nucleic Acids Res.* 2023;51(D1). doi:10.1093/nar/gkac1033
- [4] Omenn GS, Orchard S, Lane L, Lindskog C, Pineau C, Overall CM, et al. The 2024 report on the human proteome from the HUPO Human Proteome Project. *J Proteome Res.* 2024;23(12):5296-311. doi:10.1021/acs.jproteome.4c00776
- [5] Buniello A, Suveges D, Cruz-Castillo C, Bernal Llinares M, Cornu H, Lopez I, et al. Open Targets Platform: Facilitating therapeutic hypotheses building in drug discovery. *Nucleic Acids Res.* 2025;53(D1). doi:10.1093/nar/gkae1128
- [6] Pun FW, Podolskiy D, Izumchenko E, Mortlock A, Oprea TI, Scheibye-Knudsen M, et al. Target identification and assessment in the era of AI. *Nat Rev Drug Discov.* 2026;25(7):534-52. doi:10.1038/s41573-026-01412-8
- [7] Qin C, Wang YL, Zheng J, Wan XB, Fan XJ. Current perspectives in drug targeting intrinsically disordered proteins and biomolecular condensates. *BMC Biol.* 2025;23(1):118. doi:10.1186/s12915-025-02214-x
- [8] Abyzov A, Blackledge M, Zweckstetter M. Conformational dynamics of intrinsically disordered proteins regulate biomolecular condensate chemistry. *Chem Rev.* 2022;122(6):6719-48. doi:10.1021/acs.chemrev.1c00774
- [9] Heller GT, Shukla VK, Figueiredo AM, Hansen DF. Picosecond dynamics of a small molecule in its bound

- state with an intrinsically disordered protein. *J Am Chem Soc.* 2024;146(4):2319-24. doi:10.1021/jacs.3c11614
- [10] Lama D, Vosselman T, Sahin C, Liaño-Pons J, Cerrato CP, Nilsson L, et al. A druggable conformational switch in the c-MYC transactivation domain. *Nat Commun.* 2024;15(1):1865. doi:10.1038/s41467-024-45826-7
- [11] Zhu K, Wang H, Zhang J, Hu R, Jiang L, Zhang H, et al. Targeting the intrinsically disordered AR-NTD through a machine learning-based enhanced sampling workflow. *Nat Commun.* 2026;17(1):7206. doi:10.1038/s41467-026-73863-x
- [12] Meller A, Ward M, Borowsky J, Kshirsagar M, Lotthammer JM, Oviedo F, et al. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat Commun.* 2023;14(1):1177. doi:10.1038/s41467-023-36699-3
- [13] Lazou M, Kozakov D, Joseph-McCarthy D, Vajda S. Which cryptic sites are feasible drug targets? *Drug Discov Today.* 2024;29(11):104197. doi:10.1016/j.drudis.2024.104197
- [14] Bemelmans MP, Cournia Z, Damm-Ganamet KL, Gervasio FL, Pande V. Computational advances in discovering cryptic pockets for drug discovery. *Curr Opin Struct Biol.* 2025;90:102975. doi:10.1016/j.sbi.2024.102975
- [15] Zhang S, Bowman GR. Decrypting cryptic pockets with physics-based simulations and artificial intelligence. *Curr Opin Struct Biol.* 2026;96:103215. doi:10.1016/j.sbi.2025.103215
- [16] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: Advances and opportunities. *Nat Rev Drug Discov.* 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z
- [17] Butler MS, La Clair JJ. The role of natural product chemistry in drug discovery: Two decades of progress and perspectives. *J Nat Prod.* 2026;89(1):3-28. doi:10.1021/acs.jnatprod.5c01368
- [18] Wang S, Zhang Y, Yu R, Chai Y, Liu R, Yu J, et al. Labeled and label-free target identifications of natural products. *J Med Chem.* 2024;67(20):17980-96. doi:10.1021/acs.jmedchem.4c01576
- [19] Zhao J, Jiang H, Zhao K, Liu T, Zhang Q, Zhao Z, et al. Discovery of an allosteric 14-3-3 inhibitor for suppressing NRF2-driven cancer via phenotypic screening and chemoproteomic-based target deconvolution. *Chem Sci.* 2025;16(37):17248-60. doi:10.1039/d5sc04324g
- [20] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493-500. doi:10.1038/s41586-024-07487-w
- [21] Chen L, Fan Z, Chang J, Yang R, Hou H, Guo H, et al. Sequence-based drug design as a concept in computational drug design. *Nat Commun.* 2023;14(1):4217. doi:10.1038/s41467-023-39856-w
- [22] Zhao Y, Xing Y, Zhang Y, Wang Y, Wan M, Yi D, et al. Evidential deep learning-based drug-target interaction prediction. *Nat Commun.* 2025;16(1):6915. doi:10.1038/s41467-025-62235-6
- [23] van Tilborg D, Rossen L, Grisoni F. Molecular deep learning at the edge of chemical space. *Nat Mach Intell.* 2026;8(4):575-87. doi:10.1038/s42256-026-01216-w
- [24] Niphakis MJ, Cravatt BF. Ligand discovery by activity-based protein profiling. *Cell Chem Biol.* 2024;31(9):1636-51. doi:10.1016/j.chembiol.2024.08.006
- [25] Wozniak JM, Li W, Governa P, Chen LY, Jadhav A, Dongre A, et al. Enhanced mapping of small-molecule binding sites in cells. *Nat Chem Biol.* 2024;20(7):823-34. doi:10.1038/s41589-023-01514-z
- [26] Gao Y, Ma M, Li W, Lei X. Chemoproteomics, a broad avenue to target deconvolution. *Adv Sci (Weinh).* 2024;11(8). doi:10.1002/advs.202305608
- [27] George AL, Dueñas ME, Marín-Rubio JL, Trost M. Stability-based approaches in chemoproteomics. *Expert Rev Mol Med.* 2024;26. doi:10.1017/erm.2024.6
- [28] Valerie NCK, Sanjiv K, Mortusewicz O, Zhang SM, Alam S, Pires MJ, et al. Coupling cellular drug-target engagement to downstream pharmacology with CeTEAM. *Nat Commun.* 2024;15(1):10347. doi:10.1038/s41467-024-54415-7
- [29] Sterling J, Baker JR, McCluskey A, Munoz L. Systematic literature review reveals suboptimal use of chemical probes in cell-based biomedical research. *Nat Commun.* 2023;14(1):3228. doi:10.1038/s41467-023-38952-1
- [30] Aldrich C, Bertozzi C, Georg GI, Kiessling L, Lindsley C, Liotta D, et al. The ecstasy and agony of assay interference compounds. *J Med Chem.* 2017;60(6):2165-68. doi:10.1021/acs.jmedchem.7b00229
- [31] Zhang S, Miller JJ, Bowman GR. How well can AI and physics-based simulations predict the probability a cryptic pocket is open? *J Chem Theory Comput.* 2026;22(8):3839-50. doi:10.1021/acs.jctc.6c00135