Conference Track Summary

AI/ML for Early Drug Discovery

Brought to you by Cambridge Healthtech Institute, a division of Cambridge Innovation Institute

250 1st Ave, Suite 300, Needham, MA 02494
T: 781.247.6262

Executive Overview

The Drug Discovery Chemistry Europe 2025 AI/ML for Early Drug Discovery track highlighted how artificial intelligence is becoming an increasingly valuable tool for improving target identification, molecular design, and lead optimization. Across presentations, speakers emphasized that AI delivers the greatest impact when combined with physics-based simulation, high-quality data, and experimental validation. Rather than replacing traditional discovery methods, the focus was on integrating complementary computational approaches into practical research workflows that improve decision-making, reduce uncertainty, and accelerate drug discovery.

Most Frequently Covered Issues

1. AI and physics-based discovery
Combining machine learning with molecular simulation and quantum mechanics to improve prediction and lead optimization.

2. High-quality data and validation
Trusted datasets and experimental validation remain essential for reliable AI-driven discovery.

3. AI as a decision-support tool
Using AI to prioritize targets, compounds, and experiments while keeping scientists central to the process.

4. Integrated discovery workflows
Embedding AI across target identification, molecular design, screening, and lead optimization.

5. Practical implementation
Focusing on workflow integration, benchmarking, and measurable scientific outcomes over model novelty.

Recurring Takeaways

- AI is most effective when combined with physics-based methods and experimental science.
- Better data quality consistently leads to better AI performance.
- Human expertise remains essential throughout the discovery process.
- Practical workflows matter more than increasingly complex algorithms.
- Success should be measured by improved scientific outcomes, not model performance.

Simulating Biologically Relevant Protein Motion for Challenging Disease Targets

Full Video Here

Woody Sherman, Founder & Chief Executive Officer, PsiThera

Understanding how proteins move, rather than relying only on static structures, can reveal opportunities for drug discovery that would otherwise remain hidden. By combining molecular dynamics simulations with experimental structural biology, researchers can identify transient conformations, cryptic binding pockets, and allosteric mechanisms that influence protein function. This integrated approach helps prioritize biologically meaningful targets before advancing into screening and optimization. (00:00:26-00:10:43)

The presentation focused on the TNF superfamily as a case study for discovering small-molecule allosteric modulators. Although these proteins share a common structural framework, each displays distinct motions and binding behaviors, making target-specific mechanistic understanding essential. Simulations, fragment screening, and iterative design cycles were used together to identify functional binding sites and rapidly improve early hits into biologically active compounds. (00:10:43-00:15:50)

A recurring theme was that computational predictions only create value when they fit naturally into experimental decision-making. An internally developed platform brings simulation results, compound properties, structural data, and medicinal chemistry feedback into a shared environment where multidisciplinary teams can review, prioritize, and refine candidate molecules. Human expertise remains central, with experimental validation guiding each design cycle and improving future computational models. (00:15:50-00:19:11)

The discussion concluded by examining the current strengths and limitations of AI-based protein structure prediction. While modern co-folding and generative models perform well on familiar systems, they often struggle to generalize to novel binding modes, protein motion, and energetic changes that are critical for drug discovery. Physics-based simulation and AI were presented as complementary approaches, with future progress likely coming from combining data-driven models with mechanistic understanding rather than relying on either method alone. (00:19:11-00:31:11)

Key Takeaways

- Protein dynamics provide mechanistic insight that static structures alone cannot capture.
- Combining simulation with experimental validation improves discovery of allosteric drug targets.
- Integrated computational platforms are most effective when they support collaborative human decision-making.
- AI models are advancing rapidly, but physics-based methods remain essential for predicting novel protein behavior and energetics.

Leveraging Multiomics Data to Identify and Prosecute Targets Implicated in Women's Health

Full Video Here

Petrina Kamya, PhD, Global Head of AI Platforms & Vice President, Insilico Medicine; President, Insilico Medicine Canada

Women's health conditions remain underfunded despite their substantial clinical and economic impact, creating major gaps in biological understanding and therapeutic development. Endometriosis illustrates this challenge, with limited knowledge of its underlying causes and no approved disease-modifying treatments. The presentation showed how multimodal AI can help uncover new therapeutic targets by integrating sparse public datasets with biological knowledge, creating hypotheses that can then be tested experimentally. (00:00:55-00:06:02)

The project combined transcriptomic datasets from hundreds of patients and controls with AI models that analyzed gene expression, pathway biology, published literature, grants, and druggability. Rather than relying on a single metric, multiple evidence sources were used to rank candidate genes according to biological relevance and therapeutic potential. This approach identified two previously unrecognized, non-hormonal targets associated with endometriosis that consistently showed abnormal expression across multiple datasets. (00:07:21-00:12:48)

Computational predictions were followed by laboratory validation to determine whether the proposed targets had meaningful biological effects. Tissue studies confirmed elevated expression in endometriosis samples, while cellular and animal experiments showed that suppressing these targets reduced lesion growth, decreased cell proliferation, and increased apoptosis. These complementary studies strengthened confidence that the AI-generated hypotheses reflected disease biology rather than statistical artifacts. (00:12:48-00:18:58)

The presentation emphasized that AI accelerates hypothesis generation but does not replace experimental science. Public data, even when limited, can reveal promising starting points, yet questions around long-term safety, fertility, recurrence, and clinical efficacy still require extensive investigation. The greatest value comes from combining computational discovery with rigorous biological validation to shorten the path toward new therapeutic opportunities. (00:18:58-00:25:57)

Key Takeaways

- Multimodal AI can identify novel therapeutic targets even when disease data are limited.
- Integrating diverse biological evidence produces stronger target prioritization than any single dataset.
- Experimental validation remains essential for confirming AI-generated hypotheses.
- AI is most valuable when it accelerates discovery while complementing, not replacing, laboratory research.

Augmented Intelligence in Finding Druggable Targets: A Data-Driven Journey from Identification to Small-Molecule Synthesis

Full Video Here

Frederik van den Broek, Senior Director, Professional Services & Consulting, Corporate R&D, Elsevier

High-quality data remains the foundation of effective AI in drug discovery. While large language models can accelerate information retrieval and communication, they are vulnerable to hallucinations and unsupported conclusions when underlying evidence is weak. Reliable discovery therefore depends on trusted, well-curated datasets with clear provenance, allowing researchers to understand where evidence originates and maintain confidence in AI-assisted decisions. (00:00:00-00:05:44)

Using the transcription factor FOXM1 as an example, the presentation demonstrated how integrated chemistry and biology data can guide target exploration. Knowledge graphs, literature mining, bioactivity databases, and relationship mapping help researchers move beyond direct targets to identify related pathways, secondary targets, and existing compounds that provide practical starting points for drug discovery. This evidence-based approach supports informed decisions while preserving scientific context. (00:05:44-00:11:55)

Candidate molecules must also be practical to synthesize, making synthetic accessibility a critical part of lead selection. Rather than relying only on structural complexity, route-based assessment considers available building blocks and known reaction pathways to identify compounds that are realistically accessible. Retrosynthetic analysis can then suggest both published and predicted synthetic routes, providing chemists with additional options while keeping experimental expertise central to decision-making. (00:11:55-00:16:40)

The session concluded by looking toward agentic AI workflows that coordinate multiple trusted data sources through conversational interfaces. These systems have the potential to simplify complex discovery workflows, but their success depends on strong data foundations, transparent reasoning, and human oversight. AI was presented as an accelerator of scientific work, with data quality remaining the determining factor in the reliability of every downstream result. (00:16:40-00:19:53)

Key Takeaways

- Trusted, well-curated data is essential for reliable AI-assisted drug discovery.
- Knowledge graphs and integrated datasets strengthen target identification and prioritization.
- Synthetic accessibility should be evaluated alongside biological activity when selecting lead compounds.
- Agentic AI will be most effective when built on transparent, high-quality data and guided by human expertise.

AI and LQM-Driven Drug Discovery Applied to Neurodegeneration Drug Discovery

Full Video Here

Victor Sebastian Perez, PhD, Head of Computational Drug Design, EMEA, SandboxAQ

Advances in drug discovery increasingly depend on combining AI with physics-based simulation rather than treating either approach as a standalone solution. The presentation highlighted how benchmarking multiple computational methods helps determine which tools are best suited to a specific project, while collaborations across industry, academia, and computing partners provide the expertise and infrastructure needed to tackle complex therapeutic challenges, including neurodegenerative diseases. (00:00:36-00:05:36)

A major focus was expanding the data available for model development through large-scale, structure-based datasets. By combining experimental binding data with co-folded protein-ligand structures and applying rigorous quality filters, the team created a publicly available resource intended to improve model training and benchmarking. The work reflects the idea that stronger datasets are essential for advancing predictive performance across the broader research community. (00:05:36-00:09:29)

The presentation also explored how AI and physics-based methods complement one another during hit identification and lead optimization. Machine learning rapidly prioritizes compounds from vast chemical libraries, while free energy perturbation calculations provide more accurate estimates of binding affinity for selected candidates. Using AI to guide where computationally intensive simulations are applied allows researchers to explore chemical space more efficiently while maintaining scientific rigor. (00:09:29-00:23:36)

Several case studies illustrated how integrated computational workflows shortened discovery timelines and improved hit rates for challenging targets, including neurodegeneration and oncology programs. The session concluded with a look at emerging agent-based platforms that automate routine computational tasks, enabling scientists to focus more on interpretation and decision-making while keeping human expertise central to the discovery process. (00:23:36-00:27:56)

Key Takeaways

- AI and physics-based simulation are most effective when used together in complementary workflows.
- High-quality structural datasets improve model training and benchmarking across drug discovery.
- Machine learning can efficiently prioritize compounds for more rigorous physics-based evaluation.
- Agent-based workflows can automate repetitive computational tasks while preserving human oversight.

Harnessing Co-Folding for Drug Discovery: Identifying a Selective Cryptic Pocket in a Synthetic Lethality Oncology Target

Full Video Here

Jose Carlos Gómez-Tamayo, Principal Scientist, CADD, Johnson & Johnson Innovative Medicine

Co-folding is attracting growing interest in drug discovery, but its greatest value may lie in solving difficult problems where structural information is incomplete and conventional approaches struggle. Using a synthetic lethality oncology target as a case study, the presentation explored how co-folding and physics-based simulations can be combined to identify cryptic binding pockets that are absent from available structural data. The goal was not simply to predict binding, but to uncover selective opportunities that could support safer, more targeted therapies. (00:00:02-00:08:18)

The workflow began with molecular dynamics and enhanced sampling methods to explore protein conformations and reveal hidden pockets that static structures could not capture. Retrospective virtual screening then compared docking and co-folding approaches, showing that each contributed different strengths. While neither method was sufficient on its own, combining them with careful post-processing substantially improved the ability to recover known active compounds and identify physically realistic binding modes. (00:08:18-00:16:26)

The presentation also examined how generative AI can extend this process by proposing new compounds while co-folding evaluates their interaction with potential binding sites. Reinforcement learning was used to iteratively improve generated molecules, helping distinguish pockets that are merely accessible from those that are genuinely druggable. Physics-based free energy calculations then provided an independent assessment of the most promising poses and candidates. (00:16:26-00:20:21)

The broader message was that challenging discovery problems benefit from complementary computational approaches rather than a single predictive model. Co-folding, molecular dynamics, docking, and free energy methods each address different sources of uncertainty, and together provide greater confidence than any individual technique. Future progress will depend on improving model training with larger, higher-quality datasets while continuing to validate predictions with rigorous physical methods. (00:20:21-00:24:05)

Key Takeaways

- Co-folding can help uncover cryptic binding pockets that are difficult to identify with static structures alone.
- Combining co-folding with physics-based simulations improves virtual screening performance.
- Generative AI and reinforcement learning can support exploration of novel, druggable chemical space.
- Multiple complementary computational methods provide more reliable predictions than any single approach.

Generative Design in Drug Discovery: Are We Truly Innovating or Merely Complicating?

Full Video Here

Anthony Bradley, D.Phil, Assistant Professor, Department of Chemistry, University of Liverpool

The rapid growth of AI tools in drug discovery has been remarkable, with thousands of new computational methods emerging over the past decade. Despite major advances in areas such as protein structure prediction, free energy calculations, and open-source software, these technical achievements have not yet translated into meaningful improvements in overall research productivity. The presentation argued that better individual tools do not automatically create a more efficient drug discovery process. (00:01:12-00:09:06)

A central theme was that current investment priorities do not align with the largest drivers of drug discovery cost. While significant effort has gone into generating new molecules and targets, comparatively less attention has been given to technologies that improve validation, lead optimization, and clinical success rates. Because downstream failures dominate overall costs, reducing uncertainty and improving decision quality were presented as more valuable objectives than simply accelerating molecule generation. (00:09:06-00:20:56)

The discussion also examined where AI is genuinely well suited to drug discovery. Modern machine learning excels at identifying patterns, prioritizing experiments, and guiding efficient learning from available data, but it remains fundamentally limited when asked to extrapolate beyond its training domain. Models can enrich decision-making and reduce experimental burden, yet they should be viewed as decision-support tools rather than engines capable of replacing scientific reasoning or overcoming sparse biological knowledge. (00:20:56-00:28:06)

The presentation concluded by calling for a shift in how progress is measured. Rather than emphasizing predictive accuracy or the number of generated compounds, the field should evaluate technologies based on their ability to improve learning, reduce experimental cost, and increase the probability of clinical success. Better metrics, stronger validation, and thoughtful adoption strategies were presented as essential for ensuring AI delivers measurable impact on drug discovery. (00:28:06-00:32:23)

Key Takeaways

- Technical advances alone do not guarantee improvements in drug discovery productivity.
- Validation and lead optimization have greater impact on cost than faster molecule generation.
- AI is most valuable when it guides learning and prioritizes experiments within its domain of confidence.
- Meaningful evaluation requires metrics tied to scientific and clinical outcomes rather than model performance alone.

Unlocking the Future of AI: How QUELO Powers Next-Generation Drug Discovery with First-Principles Physics

Full Video Here

David Pearlman, VP Product, QSimulate

AI has transformed many areas of biology, but drug discovery presents a fundamentally different challenge because relevant experimental data are limited, inconsistent, and often noisy. Unlike large language models that benefit from enormous training datasets, predictive models for molecular binding must learn from comparatively sparse information. The presentation argued that physics-based simulation can help close this gap by generating reliable computational data that complements experimental measurements and strengthens AI-driven discovery. (00:00:01-00:06:26)

Advances in GPU computing have made quantum mechanical calculations practical for routine drug discovery workflows. By combining quantum mechanics for the binding site with molecular mechanics for the surrounding protein, the QUELO platform performs free energy calculations that were previously too computationally expensive for broad application. This enables more accurate modeling of systems where classical approaches often struggle, including compounds with changing charge states, complex metal coordination, and challenging noncovalent interactions. (00:06:26-00:13:42)

Several case studies demonstrated how quantum-enhanced free energy perturbation improves prediction for difficult targets. Applications included metalloproteins, halogen-dependent binding, and covalent inhibitors, where conventional force fields frequently lose accuracy. By capturing electronic effects that classical molecular mechanics cannot represent directly, the approach expanded the range of systems that can be modeled with confidence while maintaining practical computational performance. (00:13:42-00:21:46)

The presentation concluded by proposing a workflow that combines AI and physics rather than treating them as competing technologies. AI rapidly narrows enormous chemical libraries, while quantum-based calculations validate the most promising candidates and generate high-quality training data that can be fed back into future models. This iterative cycle positions simulation as a source of trustworthy data that continuously improves AI predictions while keeping experimental validation at the center of drug discovery. (00:21:46-00:26:27)

Key Takeaways

- Sparse and noisy experimental data limit the predictive power of AI in drug discovery.
- GPU-enabled quantum mechanics makes high-fidelity simulation practical for pharmaceutical research.
- Quantum-enhanced free energy methods improve prediction for challenging chemical systems.
- Combining AI with physics-based simulation creates a feedback loop that strengthens future predictive models.

From Fragment Seed to Small Molecule Leads

Full Video Here

Jordi Mestres, PhD, Founder & CSO, Chemotargets

Structure-based generative modeling offers an alternative to screening ever-larger chemical libraries by allowing protein binding sites to guide molecule design directly. The presentation described an approach in which protein surface features are mapped, matched with compatible molecular fragments, and used to generate compounds that reflect the local characteristics of the binding cavity. Rather than searching vast libraries, the objective is to grow molecules within the protein environment itself. (00:01:00-00:06:37)

The workflow begins by identifying fragments that are compatible with specific protein surface features and then evaluating how well those fragments are retained within the binding site. Fragments are ranked according to their interaction with the protein before being expanded through chemically feasible transformations derived from known reactions. This iterative process narrows the search space while maintaining a strong emphasis on synthetic accessibility and medicinal chemistry practicality. (00:06:37-00:13:17)

As molecules evolve, additional filtering evaluates physicochemical properties, ADME characteristics, synthetic accessibility, and novelty to guide progression toward drug-like candidates. These constraints help balance potency with developability, reducing the tendency for generated compounds to drift into unrealistic regions of chemical space. The platform also records information from each design cycle, allowing accumulated knowledge to influence subsequent rounds of molecular generation. (00:13:17-00:18:33)

A case study following the evolution of a published fragment-derived inhibitor demonstrated that the workflow could reproduce a plausible optimization pathway while generating only a limited number of high-priority candidates. The presentation concluded that combining protein-guided fragment selection with iterative learning provides a practical framework for generating synthetically feasible lead molecules tailored to specific binding sites rather than relying solely on exhaustive virtual screening. (00:18:33-00:24:01)

Key Takeaways

- Protein-guided fragment generation provides an alternative to ultra-large virtual screening.
- Fragment ranking and iterative growth focus design on chemically feasible molecules.
- Physicochemical and developability filters help maintain realistic drug-like properties.
- Iterative learning enables each design cycle to refine and improve subsequent molecular generation.