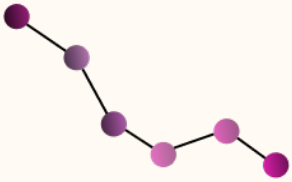
Aizen
THERAPEUTICS

Technical Review



DaX: A Novel Foundation Model for Modern Peptide Drug Design

Copyright © 2026 Aizen Therapeutics, Inc.

Key Highlights:

- Modern peptide drugs, characterized by their modifications with non-canonical amino acids (NCAAs), now rival small molecules and antibodies. While these peptide chemistry innovations have dramatically expanded the application landscape, today's experimental and computational tools are unable to systematically explore this modern design space at scale. The resulting data scarcity has left the most promising peptide chemical spaces untouched.
- Aizen Therapeutics has built a novel AI-based approach to unlock this hidden landscape. Our DaX™ foundation model abstracts the binding interface of receptor-ligand interactions, shifting the paradigm from a sequence-centric to interaction-centric architecture. By focusing on the interaction, DaX learns the universal physics and geometry of protein-ligand binding, rather than the memorization of conserved, sequence or structure motifs. Because the laws of physics governing interactions remain constant even when the amino acid sequences change, this interaction-centric approach solves the sparse NCAA training data problem, enabling the systematic exploration of NCAAs for modern peptide drugs.
- We show the validation of DaX™ on D-peptides, a class of NCAAs that have privileged drug-like properties of stability and immune tolerance. This validation was demonstrated across a diverse set of therapeutically relevant targets, including both target engagement and target activation, showing compelling hit rates, potency, and site-specific interactions.

The Requirements for Modern Peptide Drugs

Peptide therapeutics have evolved dramatically beyond the linear, canonical 20-amino acid sequences that characterized the early generations of peptide drugs, which typically involved the direct replacement of a missing endogenous peptide hormone. Today's clinically relevant peptides incorporate an expansive selection of medicinal chemistry innovations designed to address the inherent limitations of peptides such as poor protease stability, low permeability, and short *in vivo* half-lives. The introduction of NCAAs, such as D-amino acids and N-methylated amino acids, has become critical and widely used. They allow drug developers to tune a peptide's shape, permeability, protease resistance, and potency with unprecedented control.

Currently, the introduction of NCAAs into preclinical peptide hits is done via try-test-evaluate-learn cycles that involve making single NCAAs modifications to a candidate scaffold and measuring the effects. The unpredictability of these NCAAs modifications often means that one problem is solved while another liability is created. The ideal solution for drug developers would be to utilize AI models that have a granular understanding of NCAAs modifications, their 3D structures and characteristics, and how to incorporate them into *in silico* peptide designs. Standard display technologies like phage, mRNA and yeast display only efficiently operate in the canonical amino acid setting, where they can leverage biological production of binders. Exploring NCAA space forces a reliance on chemical synthesis techniques, effectively eliminating the display-based scale advantage and restricting discovery efforts to lower-throughput capacities. As a result, the field lacks systematic datasets of sufficient size and NCAA diversity to train the type of AI models that are common. This data sparsity bottleneck prevents experimental discovery of NCAA-augmented peptide drugs.

Why existing AI Models Fail on NCAA Designs

Current AI models for drug development activities typically fall into the following categories: structure-based prediction models (e.g. AlphaFold), sequence design tools (e.g. RFdiffusion), and workflow tools (e.g. BindCraft). Each of these current models faces the same challenges of designing modern peptide drugs with NCAAs: sparse training data and chemical diversity beyond the 20 canonical amino acids (**Figure 1**).

We can think of AI models as operating in either Memorization or Generalization modes. Sequence-

Copyright © 2026 Aizen Therapeutics, Inc.

centric tools rely on the statistical probability of amino acid order and therefore require large input training data sets. However, NCAAs disrupt these evolutionary patterns, rendering sequence-based predictions ineffective. In contrast, an interaction-centric view requires no prior examples of a specific sequence, only an understanding of how chemical surfaces engage in 3D space.

High generalization models, on the other hand, often work from first principles, and include molecular dynamics or quantum mechanical simulations. While accurate, these methods are computationally prohibitive for high throughput screening and lack the speed required for iterative design cycles in a drug development setting.

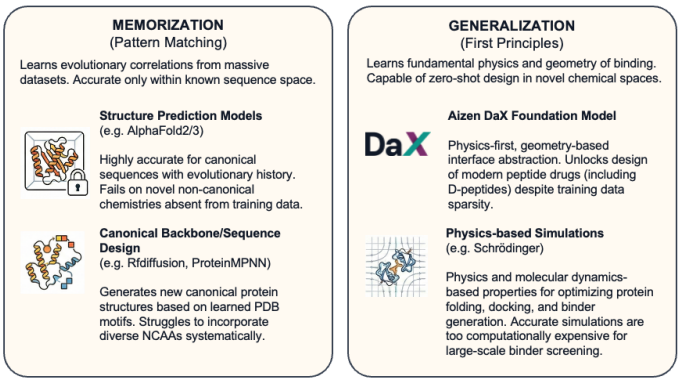


Figure 1: Evaluation of AI models for modern peptide design.

Aizen's Solution: A Foundation Model Built on a Binding-Interface Abstraction.

The Featurization Bottleneck in Current Models

Foundation models are defined by their featurization, the specific choices that control how training data is encoded into latent space and decoded into meaningful outputs. One standard approach treats protein sequences as text, featurizing them using the canonical 20-amino acid alphabet. While this approach powers sequence prediction, it creates a fundamental bottleneck: models trained on canonical, naturally occurring peptides can only output canonical amino acid binders. These models lack the capacity to generate the complex chemical diversity required for modern peptide design. To design peptide drugs that incorporate NCAAs, a foundation model requires a more generalizable input. This input must capture the detailed physics governing molecular interactions while remaining sufficiently abstract to generate chemical matter not represented in standard training data.

Aizen's Physics-First Approach

DaX™ solves both the sparse NCAA input training data problem and the chemical diversity output problem by abstracting the binding interface of the ligand-receptor interaction into a unique set of physics and geometry-based inputs. Rather than learning evolutionarily conserved sequence relationships via multiple sequence alignments, DaX™ learns the fundamental physical drivers of binding, as summarized in **Figure 2**. This interaction-centric featurization is modality-agnostic, allowing DaX™ to ingest and learn from large antibody and small-molecule binding datasets with high scale and diversity. By learning these physical insights and applying them to peptide design, we solve the data divide between abundant molecular data and sparse NCAA peptide data, creating a model that generalizes across the entire chemical spectrum.

Solving the Challenging Architecture Requirements

Choosing to build a model on physics rather than sequences generated significant engineering challenges. Because we could not rely on the well-validated architectures that operate on linear amino acid sequences, we had to either heavily modify or develop new modules in the overall model architecture at multiple levels:

- *High-Fidelity Feature Extraction:* The first challenge was fidelity. Standard models use simple one-shot encoding for amino acids, which is computationally inexpensive. DaX™, however, requires the encoding of multiple physical and geometric features to represent the interaction interface, which existing tools could not handle. Aizen developed a proprietary data processing pipeline that could operate with the required computational efficiency and maintain the required level of physical detail to unlock large training data.
- *Encoding Geometric Equivariance:* One mathematical hurdle in 3D modeling is maintaining equivariance such that the model understands that a molecule is the same object regardless of how it is rotated in space. Unlike

Copyright © 2026 Aizen Therapeutics, Inc.

text models that read linearly, DaX™ operates in a 3D coordinate system. We implemented and optimized specialized SE(3)-equivariant network layers that enable equivariance while maintaining computational efficiency beyond the performance of off-the-shelf solutions.
- *Novel Geometric Attention:* Finally, we had to innovate on the attention mechanism itself. Standard attention heads are designed to find relationships between individual amino acids in a sequence. We had to engineer geometric attention heads capable of learning relationships across irregular 3D interaction interfaces. This required developing new algorithms that allow the model to attend to complementary geometric and physical features; for example, a positively charged concave residue looking for a complementary negatively charged convex pocket.

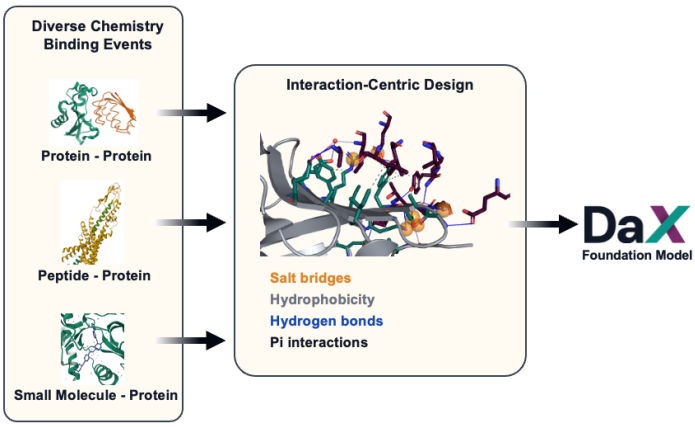


Figure 2: Graphical representation of the featurization of the binding interface via physical and geometric interactions that are used by the foundation model DaX.

Experimental Validation: *De novo* design of Fully D-amino acid Peptides

To test whether DaX had truly learned the fundamental physical rules controlling peptide binding, Aizen designed peptides from a chemical class absent from the training data: fully D-amino acid peptides. This represents a true test of zero-shot generalization: the ability to design functional peptides in a chemical space that the model has never seen before.

D-peptides are rarely found in nature and possess distinct stereochemistry. Their interactions with receptors differ from their L-peptide counterparts, and D-peptides are not represented in the PDB or other public database with sufficient scale to support memorization-based AI models. D-amino acids have been commonly used to augment endogenous peptides in therapeutic applications by decreasing the susceptibility of particular amino acids to recognition and cleavage by the body's proteases, decreasing the generation of antibodies and immunogenic outcomes. With such advantages, researchers have tried to develop lab-based methods for generating D-peptides. However, such tools do not scale for drug development activities due to limitations in the number of D-amino acids that can be incorporated into existing structures, and the throughput of testing.

Therefore, D-peptides represent the ideal chemical class for the evaluation of DaX™.

Results

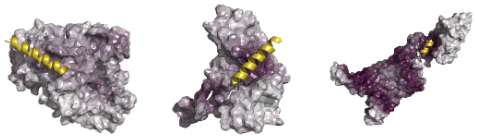
DaX™ was trained by applying the new foundation model's featurization to natural protein structures and their paired ligand from the Protein Data Bank. This training set did not include examples of D-peptides binding our selected benchmarking target receptors (CD98, TROP-2, GLP-1R). The structure for each benchmarking target was downloaded from the PDB and uploaded into DaX™, where the physics-based interaction featurization was performed to identify key factors for peptide binding. A guided diffusion model was applied to generate linear, alpha-helical peptides composed of D-amino acids with high predicted binding to the target sites.

DaX successfully:
- Developed D-peptide binders across a diverse panel of therapeutic targets (CD98, TROP-2, GLP-1R)
- Achieved high hit confirmations across a diverse set of therapeutic targets
- Activated receptors in *in vitro* assays
- Demonstrated site-specific binding
- Produced hits with 97% sequence novelty vs. public databases

The fact that Aizen-designed D-peptides bind and activate interactive therapeutic targets strongly

Copyright © 2026 Aizen Therapeutics, Inc.

supports the thesis that the DaX foundation model has learned the underlying physics of binding interactions to produce its result, because D-peptides were absent from the training data, our Aizen peptides had novel amino acid sequences.



	BBB Shuttle: CD98hc	Cancer Targeting: TROP-2	Receptor Modulation: GLP-1R
Tested	82	69	66
Binding Confirmation	25%	16%	12%
Initial *In Vitro* Activity	9.1 μM	9.7 μM	0.635 μM

Figure 3: Validation results of the DaX foundation model on three therapeutic targets.

Target Activation

Aizen's initial fully D-peptide designs can bind to a diverse set of oncology, metabolic, and immunology targets. While binding to a therapeutic target has applications for protein drug conjugates and blood brain barrier shuttles, some therapeutic applications require target engagement in controlled orientations to enable cellular signaling. GLP-1 and analogs are highly studied, gold-standard signaling peptides in this regard due to their commercial impact and central role in human metabolic control. Aizen evaluated several model functions for GLP-1R *de novo* peptide design. Results showed a 12% hit confirmation and ~635 nM EC50 cAMP activation. Predicted computational docking orientations aligned with endogenous GLP-1 and exendin-4, a highly validated positive control. Sequence analysis showed that 100% of these GLP-1 cAMP activating sequences were novel and not found in public databases.

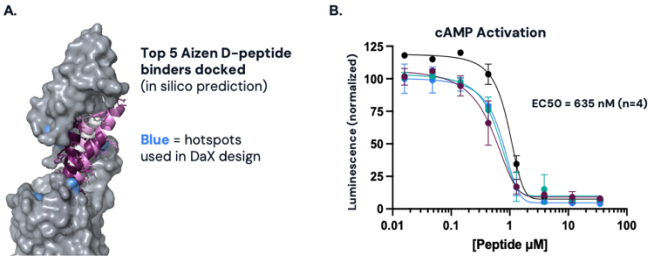


Figure 4: Aizen-designed GLP-1R activators' (a) predicted docking and (b) cAMP activation.

Value for Drug Development

By forcing DaX™ to learn the underlying physics and geometry of binding, rather than memorizing sequences, we enable the direct design of highly modified, non-canonical peptides that current AI models cannot produce. This approach unlocks the sparse NCAA training data, providing direct access to the diverse chemical spaces essential for modern peptide drugs with advanced drug-like properties. For example, DaX™ specifically empowers the development of oral biologics by designing fully D-peptides to best-in-class proteolytic stability while maintaining high affinities. Furthermore, our model supports multi-parameter optimizations, allowing drug developers to simultaneously solve for competing molecular properties, such as potency, specificity, permeability, and solubility, greatly reducing the traditional lead optimization cycle.

With DaX™, drug developers can systematically explore the modern peptide chemical space, including regions inaccessible to experimental screening, to generate candidates with improved selectivity, potency, permeability, metabolic stability, and manufacturability.

Conclusion

Modern peptide drugs depend increasingly on non-canonical chemical diversity to achieve drug-like properties. However, the lack of high-throughput experimental methods and the near-total absence of NCAA training data have made it impossible for existing AI models to systematically explore this space. Aizen's DaX™ foundation model has solved the sparse data problem via a focus on binding interface featurizations that enable the design of diverse peptides.

For more information, please visit www.aizentx.com

Copyright © 2026 Aizen Therapeutics, Inc.