



SIAMESE NEURAL NETWORKS FOR BIOISOSTERIC REPLACEMENT USING MATCHED MOLECULAR PAIRS AND BINDING ASSAY DATA

Ahmed Youssef^{1*}, Khaled Hassan¹, Mahmoud Elamin²

1. *Department of Intelligent Pharmaceutical Systems, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan.*
2. *Department of Computational Drug Analytics, Faculty of Medicine, Sudan University of Science and Technology, Khartoum, Sudan.*

ARTICLE INFO

Received:

29 November 2025

Received in revised form:

14 March 2026

Accepted:

14 March 2026

Available online:

28 April 2026

Keywords: Siamese neural network, Bioisosteric replacement, Matched molecular pairs, Binding affinity, Molecular graph learning, Medicinal chemistry

ABSTRACT

Bioisosteric replacement is a fundamental design principle in medicinal chemistry, yet identifying substituents that preserve biological activity remains largely empirical. A model-oriented approach frames this challenge as a paired molecular learning problem rather than a simple similarity search. Existing computational methods often rely on physicochemical similarity, historical replacement tables, or local transformation statistics, which may fail to fully leverage paired activity evidence from matched molecular pairs and binding assays. To address this, a Siamese neural network can be defined to learn from matched molecular pairs annotated with activity data, predicting whether a structural modification could serve as a bioisosteric replacement that maintains target affinity. In this twin-network architecture, shared weights encode both the original and modified molecules into comparable latent representations, with a contrastive or classification loss applied to the joint representation based on curated assay-derived activity labels. Conceptually, the model would distinguish activity-preserving substitutions from detrimental modifications while accommodating molecular graph or fingerprint inputs, and it could also highlight molecular features that inform replacement decisions. By providing a data-driven and interpretable framework, the Siamese approach has the potential to accelerate lead optimization by prioritizing high-probability substitutions for chemical synthesis and biological testing.

This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by/4.0/), which allows others to remix, and build upon the work non commercially.

To Cite This Article: Youssef A, Hassan K, Elamin M. Siamese Neural Networks for Bioisosteric Replacement Using Matched Molecular Pairs and Binding Assay Data. *Pharmacophore*. 2026;17(2):34-43. <https://doi.org/10.51847/Q89gHhQ6mf>

Introduction

Bioisosteric replacement is central to lead optimization because a carefully chosen substituent can preserve target engagement while altering solubility, permeability, metabolic stability, or selectivity. Classical bioisostere analysis has long relied on the medicinal chemist's judgment, but data-driven resources and structural analyses now make it possible to formalize substitution behavior across many ligands [1]. Structural databases and replacement-navigation tools further show that bioisosterism is not only a matter of atom counting, but also of spatial complementarity, hydrogen-bonding pattern, and binding-site context [2, 3]. Even with these resources, practical replacement selection remains time-consuming because each candidate substitution must be interpreted in relation to a specific scaffold, target, and assay context.

Large chemical and bioactivity databases create an opportunity to transform empirical replacement knowledge into a machine-learning task. ChEMBL-style assay deposition has made curated binding measurements more accessible for computational medicinal chemistry, while matched molecular pair platforms such as mmpdb provide systematic routes for extracting single-point structural changes [4, 5]. MMP analysis can connect a chemical transformation to an observed property shift, and later applications have extended this principle to ADME optimization and medicinal chemistry decision support [6, 7]. The remaining gap is a dedicated paired-learning framework that can treat the “before” and “after” molecules as linked evidence rather than as independent QSAR examples.

Siamese and twin-network architectures are well suited to paired molecular learning because both inputs are encoded through shared parameters before their relationship is modeled. In molecular machine learning, Siamese recurrent networks have

Corresponding Author: Ahmed Youssef; Department of Intelligent Pharmaceutical Systems, Faculty of Pharmacy, University of Khartoum, Khartoum, Sudan. E-mail: ahmed.youssef@gmail.com

already been used to compare compounds for bioactivity prediction, indicating that shared encoders can learn chemically meaningful similarity functions [8]. Similarity-based virtual screening with Siamese multilayer perceptrons and deep Siamese methods has further shown how pairwise architectures can prioritize molecules by learned resemblance rather than by hand-designed similarity alone [9, 10]. These studies suggest that a Siamese model for bioisosteric replacement could learn when two molecules remain functionally close despite a localized structural modification.

This MDL article formulates bioisosteric replacement prediction as a supervised paired-sample problem built from MMPs and binding assay data. The proposed model would compare an original molecule and a modified analogue, then estimate whether the transformation should preserve activity for a target or target family. Pairwise difference regression provides a related conceptual basis for learning changes rather than absolute properties, while activity-cliff analysis highlights why small structural changes can produce unexpectedly large activity shifts [11, 12]. A Siamese architecture trained on activity-labeled transformations could therefore bridge empirical MMP evidence with deep representation learning for medicinal chemistry design.

Background

Bioisosterism and Its Principles

Bioisosterism describes the replacement of one atom, group, or fragment by another that preserves essential biological recognition while potentially changing physicochemical or pharmacokinetic behavior. Classical and non-classical bioisosteres can involve steric mimicry, electronic resemblance, hydrogen-bonding equivalence, or scaffold-hopping relationships that retain interaction patterns in a binding pocket [2]. Deep-learning identification of bioisosteric substituents has shown that replacement knowledge can be represented computationally rather than only by expert-curated tables [1]. More recent tools for local structural replacement and property-aware assessment extend this principle by treating bioisosterism as a context-sensitive transformation rather than a universal fragment equivalence [13, 14].

Matched Molecular Pairs and Their Extraction

Matched molecular pairs formalize medicinal chemistry comparisons by identifying pairs of molecules that differ by a localized structural transformation. Open MMP platforms enable extraction of such pairs from large compound collections, while matched molecular series analysis extends the concept toward systematic structure-activity relationship interpretation [4, 15]. MMP-based machine learning has been used to support virtual compound optimization, and review work emphasizes the role of MMP analysis in translating observed transformation effects into medicinal chemistry guidance [6, 16]. Because these pairs explicitly connect a structural change to a measured response, they provide a natural training substrate for learning bioisosteric replacement rules.

Siamese and Twin Neural Networks in Molecular Machine Learning

Siamese neural networks process two related inputs through shared encoders so that the resulting embeddings can be compared by distance, interaction, or downstream classification. In chemistry, Siamese bioactivity models demonstrate that molecular pair learning can exploit shared representations to capture activity-relevant similarity [8]. Subsequent similarity-based virtual screening studies indicate that enhanced Siamese deep-learning methods can support compound prioritization when the learning objective is relational rather than purely single-molecule predictive [9, 10]. Pairwise sampling strategies for Siamese regression also show that the construction of informative pairs can influence uncertainty handling and the efficiency of learned molecular comparisons [17].

Data Sources for Activity-Labeled Molecular Pairs

Activity-labeled molecular pairs require assay evidence linking both members of a transformation to the same biological target under comparable measurement conditions. ChEMBL provides a major public source of bioassay annotations, but combining K_i , IC_{50} , or related endpoints from different sources can introduce substantial noise if assay context is ignored [5, 18]. Large-scale drug-target prediction work on ChEMBL illustrates the value of curated activity data for machine learning, while compound-protein interaction reviews emphasize the importance of representation choices for both molecules and targets [19]. In a bioisosteric Siamese model, assay curation would therefore be as important as architectural design because the label reflects a biological comparison rather than an intrinsic property of either molecule alone.

Prior Approaches to Predicting Substitution Outcomes

Earlier approaches to substitution outcome prediction include QSAR, Free-Wilson-style decomposition, MMP statistics, and transformation-level machine learning. MMP-derived models have been used to predict property improvements, and DeepDelta specifically frames derivative optimization as a learning problem over molecular changes [20]. Context-based MMP analysis for CYP inhibition further illustrates how local transformations can be analyzed in relation to a defined biological or ADME endpoint [21]. However, these approaches do not fully exploit a Siamese formulation in which both molecular members are encoded symmetrically and the learned comparison itself becomes the predictor of bioisosteric behavior.

Model Development Overview

High-Level Training and Inference Pipeline

The proposed pipeline begins with molecular pairs that differ by a single structural transformation and have comparable binding assay annotations for the same target. Each member of the pair would be passed through a shared Siamese encoder, following the paired-input logic established in molecular bioactivity and similarity-learning studies [8, 17]. The model would output a conceptual score representing whether the transformation is likely to preserve activity, rather than reporting an absolute potency value. During inference, the same architecture could rank candidate replacements for a lead compound by comparing the original structure with virtual analogues generated from admissible substitution rules.

Figure 1 illustrates the proposed Siamese paired-learning architecture in which matched molecular pairs, binding assay labels, shared molecular encoders, pair-interaction layers, interpretability outputs, and medicinal-chemistry decision support are integrated into a single bioisosteric replacement prediction workflow.

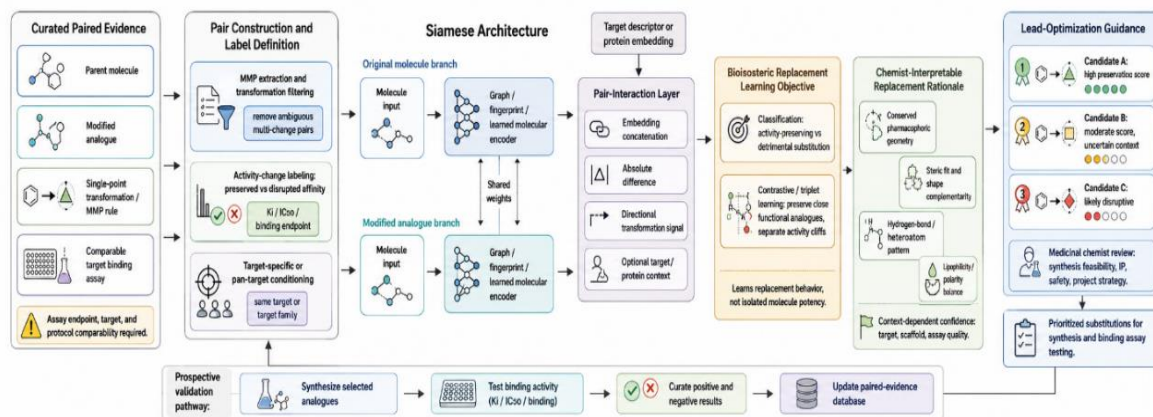


Figure 1. Siamese paired-learning architecture for bioisosteric replacement prediction from matched molecular pairs and binding assay evidence.

Core Input Representations

Each molecule could be represented as a graph, a fingerprint, or a learned sequence-derived embedding, depending on the desired balance between interpretability and expressiveness. Graph attention and graph neural network approaches have shown that atom-level message passing can support molecular property prediction by learning chemically relevant local environments [22]. Molecular contrastive learning and self-supervised graph transformers further suggest that pair-aware or augmentation-aware objectives can improve the representation of chemical similarity relationships [23, 24]. A Siamese bioisostere model would adapt these ideas by encoding two molecules with shared weights and then modeling the structural transformation through concatenation, absolute difference, or learned interaction features.

Design Principles

The design should emphasize activity-change specificity, target awareness, and interpretability. Activity-change specificity means that the model should learn to distinguish transformations expected to preserve binding from transformations likely to disrupt recognition, reflecting the importance of activity cliffs in molecular learning [12, 25]. Target awareness could be introduced through target descriptors or compound-protein interaction representations so that a replacement is interpreted in the context of a binding site rather than as a universal equivalence. Interpretability is also essential because medicinal chemists need to understand whether the model's judgment reflects shape, polarity, hydrogen-bonding, lipophilicity, or local steric effects.

*Data Sources and Pair Construction**Building a Database of Activity-Labeled MMPs*

An activity-labeled MMP database would be constructed by extracting single-point transformations from curated compound collections and linking both molecules in a pair to comparable binding measurements for the same target. MMP extraction platforms provide the structural pairing logic, while ChEMBL-style bioassay resources provide the activity annotations needed to label the transformation [4, 5]. Because assay values can vary by protocol, endpoint, and reporting source, the pair labels should be treated as curated relational evidence rather than as exact ground truth [18]. The resulting dataset would support a conceptual binary or ordinal label indicating whether the transformation could be considered activity-preserving under the chosen assay context.

Table 1 defines the paired-evidence structure required to convert matched molecular pairs and binding assay annotations into a supervised Siamese learning dataset for bioisosteric replacement prediction.

Table 1. Paired-evidence design matrix for training a siamese bioisosteric replacement model

Dataset design component	Operational definition in this manuscript	Required curation decision	Why it matters for Siamese paired learning	Risk if poorly handled
Parent molecule	The original lead, analogue, or reference compound before replacement	Confirm chemical identity, salt handling, stereochemistry, and assay linkage	Provides one branch of the twin encoder and anchors the replacement comparison	Incorrect parent definition can create false structural differences or misleading activity changes
Modified analogue	The molecule after a localized substituent or fragment replacement	Ensure that the analogue differs mainly by the intended replacement	Provides the second Siamese branch for relational comparison	Multiple hidden changes may cause the model to learn confounded SAR patterns
Matched molecular pair rule	The localized transformation connecting parent and analogue	Extract only interpretable single-point or limited local transformations	Converts general molecular similarity into explicit replacement evidence	Overly broad transformations reduce the bioisosteric meaning of the pair
Binding assay endpoint	Ki, IC ₅₀ , Kd, or related target-binding measurement	Harmonize endpoint type, unit, assay format, and target annotation	Supplies the biological evidence used to label preservation or disruption	Mixed assay contexts may create noisy or contradictory labels
Target or target family	The protein, receptor, enzyme, transporter, or family against which both molecules are evaluated	Decide whether the pair is target-specific, family-specific, or pan-target with target descriptors	Allows the model to distinguish universal replacement tendencies from context-dependent bioisosterism	Ignoring target context may falsely treat conditional replacements as generally valid
Activity-preservation label	Binary, ordinal, or continuous indication of whether affinity is maintained after replacement	Define potency-change thresholds and uncertainty categories	Translates medicinal chemistry judgment into a learnable supervised signal	Arbitrary thresholds may misclassify borderline or assay-noisy pairs
Activity-cliff flag	Indicator that a small structural change produces a large affinity loss or gain	Identify near-neighbor pairs with substantial activity divergence	Helps the model learn difficult negative examples rather than only obvious dissimilar cases	Excluding cliffs may produce a model that overestimates structural similarity as functional similarity
Transformation frequency	Number of times a replacement appears across targets, scaffolds, or series	Control for overrepresented historical replacements	Prevents the model from memorizing common substitutions instead of learning transferable principles	Highly repeated transformations may dominate the latent representation
Scaffold context	Core chemical framework surrounding the replacement site	Track scaffold family, local environment, and substituent position	Clarifies whether a replacement works broadly or only within a specific chemical series	Scaffold leakage can inflate apparent generalization during validation
Assay-quality confidence	Confidence score reflecting comparability and reliability of the paired assay evidence	Assign high, moderate, low, or excluded confidence categories	Allows uncertain evidence to be down-weighted or analyzed separately	Low-quality labels can reduce model reliability and interpretability
Negative evidence retention	Inclusion of substitutions that failed to preserve activity	Preserve disruptive analogues, not only successful replacements	Essential for contrastive, classification, and triplet learning objectives	Success-only datasets make the model unable to distinguish risky replacements
Inference candidate rules	Replacement rules used to generate virtual analogues for scoring	Restrict candidates to chemically plausible and synthetically meaningful modifications	Aligns model predictions with realistic lead-optimization use	Unrealistic virtual replacements can create misleading prioritization outputs

Filtering and Quality Control

Filtering should remove ambiguous transformations, inconsistent assay annotations, poorly comparable endpoints, and pairs in which multiple structural changes confound the replacement signal. Matched molecular series analysis provides a framework for recognizing coherent local transformations, while ADME-oriented MMP modeling shows why transformation context and endpoint consistency are important for reliable learning [7, 15]. Activity-cliff-aware studies further warn that close structural similarity does not guarantee preserved potency, making careful separation of activity-preserving and activity-disrupting examples essential [12, 25]. Balancing should also avoid allowing trivial substitutions or highly repeated historical analogues to dominate the learned representation.

Target-Specific vs. Pan-Target Pair Construction

Target-specific pair construction would focus on transformations observed within a single protein or family, allowing the model to learn replacement behavior tied to a particular binding-site environment. Pan-target construction would combine multiple targets and introduce target information as an additional input, following the broader logic of drug-target prediction and compound-protein interaction modeling [19]. A target-aware Siamese design could therefore distinguish replacements that are generally conservative from those that only preserve activity in a specific binding context. This distinction is important because bioisosterism is often conditional: the same fragment exchange can behave differently across kinases, GPCRs, enzymes, or protein-protein interaction targets.

*Siamese Network Architecture**Twin Molecular Encoder*

The twin molecular encoder would use shared parameters so that the original and modified molecules are mapped into the same latent chemical space. A graph isomorphism network, graph attention network, or related molecular graph encoder could capture local atom environments, aromaticity, heteroatom placement, and substituent connectivity, consistent with graph-based molecular learning methods [22, 26]. Fingerprint or sequence-inspired encoders could also be used when computational simplicity or compatibility with historical MMP pipelines is preferred, drawing on molecular representation studies such as Mol2vec and learned property-prediction embeddings [27, 28]. Weight sharing is essential because the model should judge the relationship between two molecules rather than learn separate encoders for “parent” and “analogue” structures.

Pair-Interaction Layer and Output

After encoding, the two molecular embeddings would be combined through a pair-interaction layer that captures both similarity and directional transformation information. Concatenation, absolute difference, bilinear interaction, or a learned feed-forward fusion module could represent how the substitution changes the latent molecular profile, echoing the logic of pairwise difference regression for chemical search [11]. Transformation-oriented molecular generation and translation studies suggest that modification patterns can be learned as relationships between chemical states rather than as isolated molecule scores [29, 30]. The output would be interpreted conceptually as a bioisosteric replacement score, indicating whether the structural modification would be expected to preserve activity under the modeled context.

Training with Contrastive or Classification Loss

Training could use a binary classification loss when pairs are labeled as activity-preserving or activity-disrupting, or a contrastive loss when the objective is to bring activity-preserving transformations closer in representation space. Molecular contrastive learning provides a natural precedent for constructing objectives that encourage chemically meaningful similarity, while activity-cliff-informed contrastive learning shows how difficult near-neighbor comparisons can sharpen activity-relevant representations [23, 25]. A triplet-style formulation could compare an original molecule, an activity-preserving analogue, and a disruptive analogue, making the learning task explicitly relational. The loss design should avoid implying that all structurally close molecules are biologically close, because bioisosteric success depends on both molecular similarity and target-specific binding context.

*Learning Bioisosteric Rules and Generalization Across Targets**Target-Agnostic and Target-Specific Models*

A target-agnostic Siamese model would learn broad replacement patterns that recur across chemical series, while a target-specific model would condition the replacement decision on a particular protein environment. GraphBioisostere illustrates the value of graph neural modeling for general bioisostere prediction, suggesting that learned molecular representations can encode substitution patterns beyond hand-curated replacement tables [31]. Target-aware extensions could concatenate a protein embedding with the paired molecular embeddings so that the same fragment exchange is interpreted differently across binding sites, consistent with compound-protein interaction modeling principles. Such a formulation would be expected to recognize that a bioisostere is not an intrinsic property of a fragment pair alone, but a relationship among the original molecule, the replacement, and the biological target.

Transferring Knowledge across Target Families

A general Siamese model could be pre-trained on diverse activity-labeled transformations and then adapted conceptually to a new target family with limited paired evidence. Large-scale ChEMBL-based drug-target modeling shows why cross-target learning can be useful when assay information is unevenly distributed across proteins [19]. Molecular contrastive learning with attention-guided graph augmentation further supports the idea that transferable molecular representations can encode activity-relevant similarity in a way that may generalize beyond a single assay series [32]. For medicinal chemistry, this transfer setting would be especially relevant when a new target has only a small number of analogues but belongs to a broader family in which replacement behavior has already been observed.

Handling Sparse Activity Data

Sparse activity data are a central challenge because many potentially useful transformations are never synthesized, never tested, or tested only under non-comparable assay conditions. Semi-supervised molecular representation learning and self-supervised graph transformers provide a conceptual route for learning chemical structure regularities before supervised bioisosteric labels are introduced [24]. Controlled molecular generation can also support virtual analogue exploration by proposing chemically plausible modifications that can later be scored by the Siamese model [33]. In this setting, data augmentation should be treated cautiously because the model should expand chemical coverage without inventing activity evidence that has not been measured.

Model Interpretability and Substitution Guidance

Explaining Which Features Make a Replacement Bioisosteric

Interpretability should reveal whether the model's replacement decision is driven by conserved pharmacophoric geometry, heteroatom placement, steric fit, lipophilicity, or electronic similarity. Graph attention mechanisms provide one route for attributing importance to atoms and bonds, making them suitable for identifying local features that influence a learned molecular comparison [22]. MB-Isoster and structural bioisostere analyses similarly emphasize that replacement interpretation should consider three-dimensional and interaction-based equivalence rather than only two-dimensional fragment similarity [2, 34]. A useful Siamese model would therefore not merely rank substitutions, but also indicate why a candidate replacement appears compatible with the parent molecule and target context.

From Prediction to Design

For a given lead compound, the Siamese model could rank virtual single-point modifications by their expected ability to preserve activity while enabling improvements in developability. Bioisostere navigation tools and data-driven replacement assessments show how curated substitution knowledge can be transformed into practical medicinal chemistry guidance [3, 14]. Transformation-oriented neural models also suggest that chemical modification can be represented as a movement from one molecular state to another, which aligns naturally with replacement prioritization [29]. The output should be presented as a decision-support list for chemists rather than as an autonomous design verdict, because synthesis feasibility, safety, intellectual property, and project strategy remain essential filters.

Integration Into Lead-Optimization Workflows

Virtual Screening of Substitution Libraries

The Siamese model could be integrated into lead-optimization platforms by scoring virtual substitution libraries generated from allowed medicinal chemistry transformations. Similarity-based virtual screening with Siamese deep learning provides a precedent for ranking candidate compounds through learned pairwise relationships rather than through fixed similarity metrics alone [9, 10]. MMP analysis in drug discovery offers the complementary transformation vocabulary needed to generate interpretable analogue sets for scoring [16]. In practice, the model would serve as a prioritization layer that helps chemists focus on substitutions that are chemically plausible and biologically conservative.

Prospective Experimental Validation Loop

A prospective workflow would use the model to propose candidate replacements, synthesize a selected subset, test them in binding or functional assays, and return the resulting measurements to the curated MMP database. ChEMBL-style direct deposition of bioassay data provides a model for how activity evidence can be organized for repeated computational reuse [5]. Context-based MMP analysis demonstrates how endpoint-specific transformation evidence can guide decisions in a defined optimization problem [21]. This feedback loop would be expected to improve the reliability of the model over time, provided that new assay results are curated consistently and negative outcomes are retained rather than discarded.

Evaluation Strategy

Predictive Performance

The model should be evaluated by its ability to distinguish activity-preserving from activity-disrupting substitutions under a prospective-like validation design. Activity-cliff studies show that this task is difficult because small structural changes can produce large biological effects, so evaluation should emphasize challenging near-neighbor comparisons rather than only easy dissimilar pairs [12, 25]. Pairwise difference regression also indicates that models trained on molecular changes should be assessed in a way that reflects the relational nature of the prediction [11]. Performance discussion should focus on whether the model is useful for prioritization and error analysis, not on isolated numerical claims.

Generalization across Targets and Chemical Series

Generalization should be assessed on targets, scaffolds, and chemical series that were not used during training. Matched molecular series analysis provides a basis for evaluating whether transformation knowledge transfers beyond a single analogue series, while MMP-based ADME prediction illustrates how learned transformation effects can depend on local chemical context [7, 15]. Bioisosteric replacement resources further show that substituent equivalence can vary according to binding environment and structural presentation [1, 13]. A robust Siamese model should therefore be expected to capture recurring

principles of conservative replacement while still flagging cases where the target or scaffold context makes extrapolation uncertain.

Prospective Enrichment in a Real-World Lead-Optimization Exercise

A realistic evaluation would retrospectively mimic a lead-optimization campaign by asking whether the model's top-ranked replacements align with modifications that medicinal chemists later found worth advancing. Virtual compound optimization with MMP-guided machine learning provides a conceptual template for using historical analogue decisions as evidence for prospective prioritization [6]. Bioisostere identification by neural networks and graph-based bioisostere prediction further support evaluating whether learned models recover plausible replacement patterns without relying entirely on manually curated rules [1, 31]. This strategy would test the practical role of the Siamese model as a ranking and triage tool rather than as a source of definitive activity predictions.

Table 2 presents an evaluation and deployment readiness framework that links predictive performance, generalization, interpretability, assay validation, and medicinal-chemistry decision use.

Table 2. Model evaluation and deployment readiness framework for bioisosteric siamese networks

Evaluation domain	Core question	Recommended assessment approach	Evidence of readiness	Failure mode the assessment should detect	Practical implication for lead optimization
Pair-level discrimination	Can the model distinguish activity-preserving replacements from disruptive substitutions?	Evaluate binary or ordinal prediction on held-out matched molecular pairs	Strong separation of preserving versus disrupting pairs under comparable assay contexts	Model treats all close structural analogues as biologically conservative	Chemists can use the model as an early triage filter for analogue selection
Activity-cliff sensitivity	Does the model recognize small structural changes that produce large affinity shifts?	Test performance on near-neighbor pairs with substantial potency differences	Correctly flags risky substitutions despite high structural similarity	Model over-relies on similarity and misses sharp SAR discontinuities	Reduces the chance of prioritizing deceptively similar but inactive analogues
Scaffold holdout generalization	Does learned replacement behavior transfer beyond familiar chemical series?	Split validation by scaffold or chemical series rather than random pairs	Maintains useful ranking performance on unseen scaffolds	Performance collapses when historical analogue series are removed	Supports broader use across medicinal chemistry programs
Target holdout generalization	Can the model generalize to new targets or target families?	Evaluate on targets excluded from training, with and without target descriptors	Identifies conservative replacements while expressing uncertainty for unfamiliar targets	Model memorizes target-specific historical transformations	Determines whether the system should be target-specific or pan-target
Transformation holdout generalization	Can the model reason about underrepresented replacement types?	Hold out selected transformation classes during validation	Provides calibrated uncertainty or partial transfer for rare replacements	Model fails on uncommon or synthetically novel substitutions	Helps define the boundary between reliable scoring and exploratory use
Ranking enrichment	Are top-ranked replacements more likely to preserve affinity than unranked candidates?	Compare enrichment among top-scored virtual analogues against historical or prospective assay outcomes	Top-ranked substitutions show higher preservation rates than baseline MMP or similarity ranking	Scores do not improve prioritization over simple structural similarity	Justifies use as a synthesis-prioritization tool
Interpretability alignment	Do model explanations correspond to medicinal chemistry reasoning?	Review atom, fragment, and interaction-level explanations with expert chemists	Explanations highlight plausible steric, electronic, pharmacophoric, or polarity features	Explanations are unstable, chemically implausible, or dominated by artifacts	Builds trust and supports design discussion rather than black-box scoring
Calibration and uncertainty	Are predicted preservation scores reliable enough for decision support?	Assess score calibration across assay-quality, target, scaffold, and transformation strata	High-confidence predictions are more accurate than low-confidence predictions	Model expresses unjustified certainty in sparse or noisy regions	Enables chemists to distinguish actionable suggestions from speculative ones

Assay-context robustness	Are predictions stable across endpoint and protocol variation?	Stratify validation by K_i , IC_{50} , K_d , assay format, and source reliability	Performance remains interpretable within comparable assay categories	Apparent accuracy is driven by inconsistent or mixed assay labels	Encourages disciplined assay curation before model deployment
Prospective experimental utility	Does the model improve real analogue selection in a lead-optimization setting?	Use the model to rank candidate substitutions, synthesize a selected subset, and test binding activity	Prospective hits are enriched among model-prioritized candidates	Retrospective performance fails to translate into experimental gain	Determines whether the model has practical medicinal-chemistry value
Human-review compatibility	Does the output support expert judgment rather than replace it?	Evaluate whether ranked candidates include rationale, uncertainty, and feasibility notes	Chemists can interpret, accept, reject, or modify suggestions based on project constraints	Output is presented as an autonomous verdict without context	Keeps the model aligned with synthesis feasibility, safety, IP, and project strategy
Continuous learning readiness	Can new assay results be incorporated without degrading reliability?	Establish update rules for adding positive and negative prospective results	New curated results improve future ranking without label drift	Feedback loop retains only successes or mixes non-comparable assays	Supports long-term improvement of the paired-evidence database

Limitations

Bias in MMP Data

MMP-derived training data are biased toward chemical changes that historical project teams chose to synthesize, register, and assay. Open MMP platforms make transformation extraction scalable, but the resulting pairs still reflect available chemistry, corporate priorities, and publication patterns rather than an unbiased map of replacement possibilities [4]. Data-driven assessment of bioisosteric replacements can expose recurring trends, yet it also inherits the limitations of the underlying compound collections and assay records [14]. As a result, the Siamese model may be more reliable for familiar medicinal chemistry transformations than for unusual, unexplored, or synthetically challenging substitutions.

Context-Dependence of Bioisosterism

Bioisosterism is inherently context-dependent because a replacement that preserves recognition in one binding pocket may disrupt shape complementarity, water networks, or electrostatic interactions in another. Structural analyses of phosphate isosteres and local replacement tools show that equivalent fragments must be interpreted in relation to their binding environment and molecular presentation [2, 13]. Assay curation concerns also matter because combining measurements from different sources can blur whether a transformation truly preserves activity or merely appears comparable under noisy experimental conditions [18]. The model’s generalizability is therefore bounded by the diversity, consistency, and biological relevance of its training data.

Conclusion

A Siamese neural network for bioisosteric replacement prediction provides a conceptual framework for learning from matched molecular pairs and binding assay data. By encoding the parent molecule and its modified analogue through shared molecular encoders, the model can focus directly on the transformation rather than treating the two structures as unrelated compounds. This paired formulation aligns naturally with medicinal chemistry reasoning, where design decisions often depend on whether a specific substitution preserves biological recognition.

The main strength of the approach is that it links empirical analogue evidence with deep molecular representation learning. A target-aware version could condition replacement decisions on protein context, while interpretable attribution could help chemists understand which features support a predicted activity-preserving substitution. Because the model can be integrated with existing MMP pipelines, it could fit into lead-optimization workflows without replacing expert judgment.

Important challenges remain. Historical MMP data are incomplete and biased toward previously explored chemical space, while bioisosteric behavior can change across targets, scaffolds, and assay formats. Prospective validation in active drug-discovery projects would therefore be necessary before such a model could be trusted as a practical prioritization tool.

Open-source implementations, transparent curation workflows, and shared benchmarks would accelerate the development of reliable MMP-Siamese models. Collaborative evaluation across targets and chemical series would help clarify when deep-learning-guided bioisosteric design is most useful. With careful validation, Siamese models could become a practical bridge between accumulated medicinal chemistry knowledge and modern molecular design systems.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Ertl P. Identification of bioisosteric substituents by a deep neural network. *J Chem Inf Model.* 2020;60(7):3369-75.
2. Zhang Y, Borrel A, Ghemti L, Regad L, Boije af Gennäs G, Camproux AC, et al. Structural isosteres of phosphate groups in the protein data bank. *J Chem Inf Model.* 2017;57(3):499-516.
3. Ertl P. Craig plot 2.0: an interactive navigation in the substituent bioisosteric space. *J Cheminform.* 2020;12(1):8.
4. Dalke A, Hert J, Kramer C. mmpdb: An open-source matched molecular pair platform for large multiproperty data sets. *J Chem Inf Model.* 2018;58(5):902-10.
5. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* 2019;47(D1):D930-40.
6. Turk S, Merget B, Rippmann F, Fulle S. Coupling matched molecular pairs with machine learning for virtual compound optimization. *J Chem Inf Model.* 2017;57(12):3079-85.
7. Awale M, Riniker S, Kramer C. Matched molecular series analysis for ADME property prediction. *J Chem Inf Model.* 2020;60(6):2903-14.
8. Fernandez-Llaneza D, Ulander S, Gogishvili D, Nittinger E, Zhao H, Tyrchan C. Siamese recurrent neural network with a self-attention mechanism for bioactivity prediction. *ACS Omega.* 2021;6(16):11086-94.
9. Altalib MK, Salim N. Similarity-based virtual screen using enhanced Siamese multi-layer perceptron. *Molecules.* 2021;26(21):6669.
10. Altalib MK, Salim N. Similarity-based virtual screen using enhanced siamese deep learning methods. *ACS Omega.* 2022;7(6):4769-86.
11. Tynes M, Gao W, Burrill DJ, Batista ER, Perez D, Yang P, et al. Pairwise difference regression: a machine learning meta-algorithm for improved prediction and uncertainty quantification in chemical search. *J Chem Inf Model.* 2021;61(8):3846-57.
12. Van Tilborg D, Alenicheva A, Grisoni F. Exposing the limitations of molecular machine learning with activity cliffs. *J Chem Inf Model.* 2022;62(23):5938-51.
13. Zhang T, Sun S, Wang R, Li T, Gan B, Zhang Y. BioisoIdentifier: an online free tool to investigate local structural replacements from PDB. *J Cheminform.* 2024;16(1):7.
14. Helmke PS, Kandler J, Ilie S, Gaskin L, Ecker GF. Data-driven assessment of bioisosteric replacements and their influence on off-target activity profiles. *RSC Med Chem.* 2025;16(12):6048-58.
15. Ehmki ES, Kramer C. Matched molecular series: measuring SAR similarity. *J Chem Inf Model.* 2017;57(5):1187-96.
16. Yang Z, Shi S, Fu L, Lu A, Hou T, Cao D. Matched molecular pair analysis in drug discovery: methods and recent applications. *J Med Chem.* 2023;66(7):4361-77.
17. Zhang Y, Menke J, He J, Nittinger E, Tyrchan C, Koch O, et al. Similarity-based pairing improves efficiency of siamese neural networks for regression tasks and uncertainty quantification. *J Cheminform.* 2023;15(1):75.
18. Landrum GA, Riniker S. Combining IC50 or Ki values from different sources is a source of significant noise. *J Chem Inf Model.* 2024;64(5):1560-7.
19. Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chem Sci.* 2018;9(24):5441-51.
20. Fralish Z, Chen A, Skaluba P, Reker D. DeepDelta: predicting ADMET improvements of molecular derivatives with deep learning. *J Cheminform.* 2023;15(1):101.
21. Raut JA, Dixit VA. A context-based matched molecular pair analysis identifies structural transformations that reduce CYP1A2 inhibition. *RSC Med Chem.* 2025;16(7):3281-90.
22. Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J Med Chem.* 2019;63(16):8749-60.
23. Wang Y, Wang J, Cao Z, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell.* 2022;4(3):279-87.
24. Rong Y, Bian Y, Xu T, Xie W, Wei Y, Huang W, et al. Self-supervised graph transformer on large-scale molecular data. *Adv Neural Inf Process Syst.* 2020;33:12559-71.
25. Shen WX, Cui C, Su X, Zhang Z, Velez-Arce A, Wang J, et al. Activity cliff-informed contrastive learning for molecular property prediction. *Res Sq.* 2024:rs-3.
26. Li M, Zhou J, Hu J, Fan W, Zhang Y, Gu Y, et al. Dgl-lifesci: An open-source toolkit for deep learning on graphs in life science. *ACS Omega.* 2021;6(41):27233-8.
27. Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J Chem Inf Model.* 2018;58(1):27-35.

28. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model.* 2019;59(8):3370-88.
29. Tysinger EP, Rai BK, Sinitskiy AV. Can we quickly learn to “translate” bioactive molecules with transformer models? *J Chem Inf Model.* 2023;63(6):1734-44.
30. Xu Z, Lei X, Ma M, Pan Y. Molecular generation and optimization of molecular properties using a transformer model. *Big Data Min Anal.* 2023;7(1):142-55.
31. Masunaga S, Furui K, Kengkanna A, Ohue M. GraphBioisostere: general bioisostere prediction model with deep graph neural network. *J Supercomput.* 2026;82(3):132.
32. Wang J, Guan J, Zhou S. Molecular property prediction by contrastive learning with attention-guided positive sample selection. *Bioinformatics.* 2023;39(5):btad258.
33. Shin B, Park S, Bak J, Ho JC. Controlled molecule generator for optimizing multiple chemical properties. In: *Proceedings of the Conference on Health, Inference, and Learning; 2021.* p. 146-53.
34. Elias TC, de Oliveira HC, da Silveira NJ. MB-Isoster: A software for bioisosterism simulation. *J Comput Chem.* 2018;39(29):2481-7.