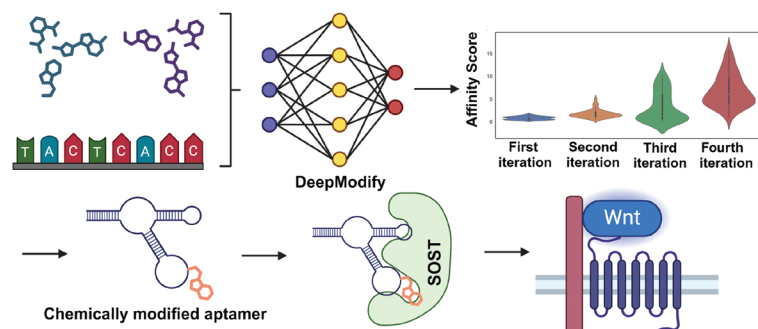


Machine learning-powered, high-affinity modification strategies for aptamers

Graphical abstract



Highlights

- Our developed machine learning model shows promising capabilities in accurately predicting the binding affinity of aptamers with post-SELEX modifications.
- The rationally designed aptamers with specific post-SELEX modification exhibit a 105-fold affinity enhancement (picomole level K_D value).
- The study offers valuable insights into the use of artificial intelligence (AI) for high-affinity prediction in aptamer with post-SELEX modification.

Authors

Gubu Amu, Xin Yang, Hang Luo, Sifan Yu, Huarui Zhang, Yuan Tian, Yuanyuan Yu, Shijian Ding, Yufei Pan, Zefeng Chen, Yixin He, Yuan Ma, Baoting Zhang and Ge Zhang

Correspondence

berryhe@aptacure.com (Y. He);
mayuan@hkbu.edu.hk (Y. Ma);
zhangbaoting@cuhk.edu.hk
(B. Zhang); zhangge@hkbu.edu.hk
(G. Zhang)

In brief

This study demonstrates the successful development of a machine learning model for accurately predicting the binding affinity of chemically modified aptamers, leading to rationally designed aptamers with 105-fold affinity enhancement and providing valuable insights into AI-driven high-affinity aptamer prediction.

Research Article

Machine learning-powered, high-affinity modification strategies for aptamers

Gubu Amu^{a,1}, Xin Yang^{a,b,1}, Hang Luo^{c,1}, Sifan Yu^{b,1}, Huarui Zhang^{c,1}, Yuan Tian^b, Yuanyuan Yu^b, Shijian Ding^b, Yufei Pan^a, Zefeng Chen^b, Yixin He^{a,*}, Yuan Ma^{b,c,*}, Baoting Zhang^{c,*} and Ge Zhang^{b,*}

^aAptacure Therapeutics Limited, Hong Kong SAR, China

^bLaw Sau Fai Institute for Advancing Translational Medicine in Bone & Joint Diseases, School of Chinese Medicine, Hong Kong Baptist University, Hong Kong SAR, 999077, China

^cSchool of Chinese Medicine, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, 999077, China

¹These authors contributed equally.

*Correspondence: berryhe@aptacure.com (Y. He); mayuan@hkbu.edu.hk (Y. Ma); zhangbaoting@cuhk.edu.hk (B. Zhang); zhangge@hkbu.edu.hk (G. Zhang)

Received: 8 October 2024; Revised: 25 December 2024; Accepted: 5 January 2025

Published online: 30 January 2025

DOI 10.15212/AMM-2024-0065

ABSTRACT

The binding affinity of aptamers to targets has a crucial role in the pharmaceutical and biosensing effects. Despite diverse post-systematic evolution of ligands by exponential enrichment (post-SELEX) modifications explored in aptamer optimization, accurate prediction of high-affinity modification strategies remains challenging. Sclerostin, which antagonizes the Wnt signaling pathway, negatively regulates bone formation. Our screened sclerostin aptamer was previously shown to exert bone anabolic potential. In the current study, an interactive methodology involving the exchange of mutual information between experimental endeavors and machine learning was initially proposed to design a high-affinity post-SELEX modification strategy for aptamers. After four rounds of interactive training (a total of 422 modified aptamer-target affinity datasets with diverse modification types and sites), an artificial intelligence model with high predictive accuracy with a correlation coefficient of 0.82 between the predicted and actual binding affinities was obtained. Notably, the machine learning-powered modified aptamer selected from this work exhibited 105-fold higher affinity (picomole level K_D value) and a 3.2-folds greater Wnt-signal re-activation effect compared to naturally unmodified aptamers. This approach harnessed the power of machine learning to predict the most promising high-affinity modification strategy for aptamers.

Keywords: Machine learning, Aptamer, High affinity, Post-SELEX, Sclerostin

1. INTRODUCTION

Nucleic acid aptamers, which are single-stranded RNAs or DNAs, exhibit the capability to bind to a wide array of targets with high specificity. Aptamers are usually generated using the systematic evolution of ligands by exponential enrichment (SELEX) technique [1]. Compared to monoclonal antibodies, the compact size and excellent solubility of aptamers facilitate the administration of high molar doses and enhanced tissue penetration, thereby optimizing bioavailability [2]. The modularity of oligonucleosides permits easy creation of aptamers in industrial scales [3]. Despite processing many advantages, naturally unmodified aptamers usually exhibit suboptimal binding affinity compared to monoclonal antibodies, limiting the pharmacodynamic

effects in drug discovery and recognition efficacy in probe exploitation [4, 5].

Chemical modification strategies of aptamers have been widely used to increase stability with lower immunogenicity and excellent safety profile [6]. Specifically, chemical modification strategies greatly expand the diversity of nucleoside composition, which have been commonly used for enhanced interactions between the modified aptamer and the target [7, 8]. For example, Davies et al. [9] and Gawande et al. [10] advanced the field of DNA chemistry by introducing hydrophobic amino acid-like side chains (indoyl, benzyl, pyridyl, isobutyl, naphthyl, and pyrrolyl) into aptamers using a modified SELEX (mod-SELEX) technique. These unique aptamers, known as SOMAmers, exhibit slow off-rate to proteins due to the expansion of hydrophobic interaction

interfaces [9, 10]. This feature holds promising prospects because the SOMAMers have demonstrated binding affinities to proteins that are comparable to antibodies [11]. However, only a small number of chemical modification types can be used in mod-SELEX due to the compatibility limitation of polymerases [12]. Additionally, a vast number of chemical modifications in aptamers introduced by mod-SELEX significantly enhance the synthetic costs and potential safety concerns. To address this issue, a polymerase-independent, post-SELEX modification approach was adopted to maintain flexibility, simplicity, and minimized selection bias compared to the mod-SELEX approach [12]. However, it is challenging to predict the effects of post-SELEX aptamer modifications on the binding affinity to proteins. Therefore, researchers must synthesize numerous aptamers with post-SELEX modifications to identify the promising aptamer candidate with the highest affinity. Notably, the potential for modification types and sites in aptamers is limitless. Manual characterization of the affinity between all post-SELEX modified aptamers and corresponding targets is impractical. Thus, it is desirable to explore virtual prediction strategies for optimizing the post-SELEX modification sites and types of aptamers for the highest binding affinity [3].

Artificial intelligence (AI) has demonstrated notable successes in predicting target-ligand interactions in drug discovery [13, 14]. For example, Zhavoronkov et al. developed a machine learning approach for rapidly identifying promising DDR1 kinase inhibitors, which selectively inhibit DDR1 activity *in vitro* (IC_{50} : 10 nM) and *in vivo* [15]. Importantly, machine learning approaches have been successfully used in predicting the binding affinity of naturally unmodified aptamers to targets. Early in 2009 Knight et al. used machine learning techniques to construct a sequence-fitness landscape using aptamers and affinity data [16]. In 2021 Bashir et al. used particle display to segregate a library of aptamers based on affinity and subsequently trained machine learning models, which facilitated a truncated high-affinity DNA aptamer design (K_D : 1.5 nM) [2]. Recently, Iwano et al. reported a variational autoencoder for high-throughput SELEX data-driven RNA aptamer generation [17]. Although these computational approaches enable identification of aptamers with comparatively high affinity, the training data depend on high-throughput SELEX next-generation sequencing. More importantly, these AI models only generate *de novo*, naturally-unmodified aptamers or DNA fragments, which are ineffective in the affinity optimization of the selected aptamer. To date, there have been no machine learning-powered, high-affinity modification strategies in aptamer discovery.

In our previous work an aptamer (named aptsc156) was screened against sclerostin for the treatment of osteogenesis imperfecta and osteoporosis [18, 19]. The aliphatic acid-conjugated aptsc156 and stapled aptsc156 exhibited excellent pharmacokinetic characteristics and metabolic stability [20, 21]. However, the binding affinity was much lower compared to an anti-sclerostin

antibody and needed improvement. In the study, an interactive methodology was elucidated involving the exchange of mutual information between experimental endeavors and machine learning to a robust aptamer design with post-SELEX modification. Different types of modifications were synthesized and characterized in aptamers at different sites (a total of 422 modified aptamer sequences) to assess the effects on binding affinity. By utilizing acquired affinity data, a well-established evolved AI model was trained and a modification-sequence library was generated. A stacking learning approach was used combined with Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) [22] to train the model. This model was trained with the 422 modified aptamer sequences and the corresponding affinity data. The model was able to learn the complex patterns and relationships between the sequence modifications and binding affinities. The model showed a high predictive accuracy with a correlation coefficient of 0.82 between the predicted and actual binding affinities. Moreover, the model was able to generate a sequence library of potential aptamers with high-affinity post-SELEX modifications. The top 10% of the sequences predicted by the model were synthesized and tested, 80% of which had significantly improved binding affinities compared to the original sequences. This approach displays a bright potential for machine learning that can rapidly predict and determine the most promising post-SELEX high-affinity modification strategy for aptamers.

2. MATERIAL AND METHODS

2.1 Materials

All the chemical compounds used in this study were purchased from Aldrich Sigma (Burlington, Massachusetts, United States). Lipofectamine 2000, Dulbecco's modified Eagle's medium (DMEM), MEM- α medium, and fetal bovine serum (FBS) were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, United States). Serum-free cell freezing medium was purchased from Beijing T&L Biological Technology Co., Ltd. (Beijing, China). OptiMem medium was purchased from macgene (Beijing, China). A high-capacity cDNA reverse transcription kit was purchased from Applied Biosystems (Foster City, California, United States). The TransZol Up Plus RNA kit was purchased from TransGen Biotech (Beijing, China). The DNA primers and DNA alkynylation used in this study were purchased from Biosyntech (Suzhou, China). The Dual-Luciferase Reporter Assay System and passive lysis buffer (PLB) were purchased from Promega (Madison, Wisconsin, United States). The plasmids used in this study were purchased from GenScript Biotech Corporation (Burlington, New Jersey, United States). The HEK293 cells were purchased from BPS Bioscience Ltd. (San Diego, California, United States). The MC3T3-E1 cells were purchased from the American Type Culture Collection [ATCC] (Manassas, Virginia, United States).

Research Article

2.2 Aptamer synthesis with post-SELEX modifications

Alkynylated aptamers (1.0 equivalent), azide-linked modified molecules (100 equivalents), and CuSO_4 (50 equivalents) were combined in a reaction tube for aptamers with 2'-hydroxyl modifications. Subsequently, 200 mM NaHCO_3 (5 μL), acetonitrile/ H_2O (100 μL at an acetonitrile-to- H_2O ratio of 6%), and 20 mM tris(2-carboxyethyl) phosphine [TCEP] (3 μL) were added and the mixture was vortexed. The reaction tube was then briefly purged with nitrogen gas and placed on a shaker at 25°C for 4 h. The Fmoc groups in amino acids were removed using ammonium hydroxide (60°C for 1 h). Synthesis was performed on an H-8 DNA/RNA Oligo Synthesizer (K&A Labs, Schaafheim, Germany) following standard phosphoramidite chemistry for aptamers with base modifications or sugar ring scaffold changes with an increased coupling time for modified phosphoramidite monomers. Finally, the crude aptamers were purified using an Xbridge® Oligonucleotide BEH C18 OBD™ Prep Column (2.5 μm , 10 mm $\times$ 50 mm; Waters Corporation, Milford, Massachusetts, United States) and desalted using a Sephadex G25 column (Cytiva, Marlborough, Massachusetts, United States) with an Agilent 1260 Semi-Prep LC System (Santa Clara, California, United States). The samples were further lyophilized to dryness for storage and confirmed by electrospray ionization mass spectrometry analysis (Table S1, Figure S1).

2.3 Feature representation

The feature representation for small molecules was accomplished using extended connectivity fingerprints (ECFPs), a well-established method for encoding the structure of small molecules. ECFPs are a class of topologic fingerprints used in cheminformatics to represent the chemical structure of small molecules. An ECFP is a circular type of fingerprint in which the environment of each atom is encoded by considering the surrounding atoms in subsequent layers. The ECFP fingerprint of a molecule was generated and used as input for the machine learning models. One-hot vectors were used for feature representation of the small molecule modification sites. Using this encoding scheme, each modification site is represented by a binary vector containing one '1' and '0's elsewhere. This scheme allows the model to handle categorical data without making any assumptions about the relative importance of different modification sites.

To quantitatively characterize the possibility of high affinity, the affinity score (AS) value was adopted, which was measured as follows:

$$\text{AS} = \frac{Kd_{\text{initial}}}{Kd_{\text{modified}}}, \quad (1)$$

where Kd_{initial} and Kd_{modified} are the K_D values of the aptamer to the target protein before and after modification, respectively.

2.4 Model training

The model was trained using a stacking model that combines RF, XGBoost, and LightGBM algorithms. Stacking is a learning technique that leverages the collective power of regression models or multiple classification by using a meta-regressor or a classifier [23]. The base-level models undergo training on a comprehensive training set and the meta-model is subsequently trained using the outputs generated by the base-level models as informative features. The stacking model is often used to improve the model performance of multiple machine learning algorithms. RF is a bagging algorithm that generates multiple decision trees and aggregates the results. RF reduces overfitting by averaging the results of different trees. XGBoost is a powerful boosting algorithm that constructs robust predictive models by combining an ensemble of weak prediction models, commonly in the form of decision trees. LightGBM is a gradient boosting framework utilizing tree-based learning algorithms with the advantages of training speed and model performance. The training process was performed using a 10-fold cross-validation. The dataset was randomly partitioned into 10 sub-samples of equal size in this procedure. Of the 10 sub-samples, 1 sub-sample was kept aside as the validation data to assess the model, while the remaining 9 sub-samples were used as training data. This process was iterated 10 times with each sub-sample used once as the validated data. The results obtained from the 10-folds were then averaged to yield a consolidated estimation.

2.5 Model evaluation

The performance of DeepModify was evaluated using a 10-fold cross-validation method for each dataset to ensure the prevention of overfitting and maintain experimental accuracy. The proposed method was assessed by comparing the performance of DeepModify with three base models involved in the stacking process. The model efficacy in predicting binding affinity was gauged using the coefficient of determination (R^2) and the Pearson correlation coefficient (PCC) per the methodology used by Weirauch et al. We hypothesized that the closer these two evaluation indices are to 1, the more efficient the method. These two indicators were utilized for each dataset. The definitions for these performance calculations are provided by:

$$R^2 = 1 - \frac{\sum_i (y_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \quad (2)$$

and

$$\text{PCC}(y, Y) = \sum_i \frac{y_i - \bar{y}}{\sqrt{(y_i - \bar{y})^2}} \cdot \frac{Y_i - \bar{Y}}{\sqrt{(Y_i - \bar{Y})^2}}, \quad (3)$$

where y_i , Y_i , $\bar{y}$, and $\bar{Y}$ represent the observed, predicted, average observed, and average predicted binding affinity scores, respectively.

This evaluation metric was chosen because the metric effectively measures how well the predicted affinity values correlate with the actual values, which is crucial in testing model performance on predicting aptamer modification strategies. The data used for this study were obtained from a comprehensive database of small molecules and the corresponding modification sites. The database includes a wide array of small molecules with diverse structures, allowing for a robust and comprehensive analysis. The specific details of the database, including the source and data cleaning process and pre-processing, are beyond the scope of this section and will be discussed in the Data Preparation section. Python programming language was used for all data pre-processing, feature extraction, model training, and evaluation tasks. The specific machine learning libraries used include scikit-learn for the RF, XGBoost, and LightGBM algorithms. Model stacking was also implemented using the functionalities provided by these libraries. ECFP fingerprint generation and one-hot vectors (*RDKit*), a collection of machine learning and cheminformatics software, was used. The PCC was computed using the *scipy* library, a Python library used for scientific and technical computing. The computation was performed on a high-performance computing cluster equipped with multiple CPUs and a large amount of RAM to handle the heavy computational load of the machine learning algorithms and the large dataset. In conclusion, the current study used a combination of advanced feature representation techniques and a stacked ensemble of machine learning algorithms to predict high-affinity modification strategies in aptamers. The model performance was assessed using the PCC and the entire process was performed using commonly used open-source software and a high-performance computing infrastructure.

2.6 Bio-layer interferometry (BLI) assay

The interactions between selected aptamers and sclerostin were analyzed using the Octet 96/96e system (ForteBio) at 25°C. The running processes and conditions were as follows: baseline, 1X PBS buffer (10 mM phosphate buffer, 2.7 mM KCl, and 137 mM NaCl [pH 7.4]) for 120 s at 1000 rpm; loading, 100 nM 5-biotin aptamer in 1X PBST buffer (0.02% Tween20 in 1X PBS [pH 7.4]) for 300 s at 500 rpm; baseline 2, 1X PBST buffer for 300 s at 500 rpm; association, sclerostin (varied concentrations) in 1X PBST buffer for 300 s at 500 rpm; dissociation, 1X PBST buffer for 300 s at 500 rpm; regeneration, 5 M NaCl for 30 s at 500 rpm; and neutralization, 1X PBS buffer for 60 s at 500 rpm. Finally, the affinity constant K_D values were evaluated using fitting curves (1:1 binding model) with ForteBio Data analysis 11.0 [24].

2.7 TOP-Wnt-induced signal assay

The HEK293 cells were seeded on 24-well plates (1.5×10^5 per well) in DMEM medium. After the cells were cultured for 24 h, the Topflash plasmid (0.20 µg, firefly

luciferase reporter regulated by TCF/LEF transcription factors) and SV40 (0.02 µg, Renilla luciferase reporter, inner control) were co-transfected with Wnt-1 plasmid (0.25 µg per well) and sclerostin plasmid (0.20 µg per well) utilizing lipofectamine 2000. Following a 12-h incubation period, the transfection mixture was carefully aspirated and 0.5 mL of regular growth medium and varying concentrations of aptamers were introduced. After incubating at 37°C for 12 h, the medium was aspirated and the cells were lysed in 0.5 mL of 1X PLB by gently shaking the culture vessel for 15 min. Fifteen microliters of the lysate was transferred to a 96-well OptiPlate for luciferase activity determination and chemiluminescence was measured using the MD SpectraMax i3X Multi-Mode Microplate Reader system (molecular devices, Sunnyvale, California, United States) following the manufacturer's protocol for the Dual-Luciferase Reporter Assay System. Finally, the data were analyzed using Origin 2019b (description, non-linear curve fit; iteration algorithm, Levenberg-Marquardt; model, dose response).

2.8 Real-time qPCR assay

MC3T3-E1 cells (mouse preosteoblasts) were cultured in MEM-α medium containing 100 units/mL penicillin-streptomycin and 10% FBS. After seeding the cells in 24-well plates (2×10^5 cells per well) for 24 h, the cells were transfected with Wnt-1 plasmid (0.25 µg per well) and sclerostin plasmid (0.20 µg per well) using 2 µL of lipofectamine 2000. During the transfection process, the plasmids and lipofectamine 2000 were initially mixed in 0.1 mL of OptiMem medium for 20 min. Then, 0.4 mL of serum-free MEM-α medium without antibiotics was added and gently mixed. Finally, the transfection mixture was added to the cells after removing the old medium. The transfection mixture was removed after 6 h and 0.5 µM of Aptscl56 or Aptscl56-9b11s was added. Following incubation at 37 °C for 12 h, the medium was removed and total RNA from the cultured cells was isolated using the TransZol Up Plus RNA kit, according to the manufacturer's instructions. Following purification, reverse transcription of 1.0 µg of total RNA into cDNA was performed using a high-capacity cDNA reverse transcription kit [25]. The specific primers utilized for this study were as follows: *Runx2* (forward primer: CCT GAA CTC TGC ACC AAG TCCT, reverse primer: TCA TCT GGC TCA GAT AGG AGGG) [26]; *Alp* (forward primer: CCA GAA AGA CAC CTT GAC TGT GG, reverse primer: TCT TGT CCG TGT CGC TCA CCAT) [26]; *Ocn* (forward primer: GCA ATA AGG TAG TGA ACA GAC TCC, reverse primer: CCA TAG ATG CGT TTG TAG GCGG) [27]; and *GAPDH* (forward primer: CAT CAC TGC CAC CCA GAA GAC TG, reverse primer: ATG CCA GTG AGC TTC CCG TTC AG) [28]. Quantitative PCR reactions were performed using SYBR Green qPCR Mix on the 7900 HT Sequence Detection System (Applied Biosystems, Foster City, California, United States). The relative RNA expression of the target genes was determined using the $2^{-\Delta\Delta Ct}$ method with GAPDH serving as the endogenous normalizer.

Research Article

2.9 Molecular docking

The secondary structure of the unmodified aptamer was predicted using the RNAfold web server [29] and the 3D structure was predicted using the RNAComposer server [30]. The resulting model with the lowest Gibbs free energy was imported into Molecular Operating Environment software. To obtain the aptamer structure (Aptsc156), all the hydroxyl groups at the 2'-position of the sugars were removed and corresponding modifications, such as 9b and 11s, were added. Energy minimization was then performed to obtain the final Protein Data Bank (PDB) file for docking. The HDock web server was subsequently used [31] with the receptor of sclerostin loop 3 (PDB ID: 2K8P) and the submitted aptamer files. The HDock server integrated template-based modeling and free docking capabilities to predict the architecture of protein-DNA/RNA or protein-protein complexes. The server automatically predicted the interaction using a hybrid algorithm of template-based and -free docking. The resulting complex structure was then analyzed using the Protein-Ligand Interaction Profiler web [32] to identify key interaction modes and refine the docking results. The visualization of the modeled interactions was performed using PyMOL, a software tool that facilitates a detailed examination of binding conformations.

2.10 Statistical analysis

The data are presented as the mean $\pm$ standard deviation. To assess inter-group differences in the study variables, including *in vitro* mRNA levels of bone formation biomarkers, a one-way ANOVA with Tukey's *post-hoc* test, was performed. Statistical analysis was performed using Origin 2019b and GraphPad Prism software. A $P < 0.05$ was considered a statistically significant difference.

3. RESULTS

3.1 Constructing a modified aptamer-affinity value dataset for training and verifying the DeepModify model

We implemented a comprehensive modification approach to optimize the binding affinity of the aptamer to its target. The major influencing factors, including the modification site in the aptamer, nucleobase type, structural characteristics of orientation, stereoisomerism, flexibility, weight, and hydrophilicity, have been taken into consideration. Liquid- and solid-phase synthesis approaches were applied to aptamer modification according to modification positions. To maximize the chemical diversity, 2'-hydroxyl modification, base modification, and nucleosugar scaffold changes were considered in this study. For the 2'-hydroxyl modification, 2'-propargylhydroxy guanosine was individually incorporated into aptamers at diverse sites, forming an alkyne-modified aptamer library A (12 sequences; Figure S2a). Then, 15 types of protein-like side chains (Figure S2b) were conjugated to the 2' position of the nucleosugar in the aptamer, forming a

larger modified aptamer library B (180 sequences, i.e., 12×15 sequences). Briefly, the functional azido groups were conjugated to alkenyl-aptamers via a highly efficient copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) approach under nitrogen protection. After purification using high-performance liquid chromatography (HPLC), the modified aptamer was immobilized on the surface of a streptavidin biosensor. The BLI assay was then used to perform association and dissociation procedures against sclerostin. The binding affinity of each modified aptamer against sclerostin was then calculated.

Phenyl, indolyl, a THF spacer, and a C3 spacer were individually incorporated into the nucleobase in aptamer at 22 diverse sites using phosphoramidite chemistry by standard DNA synthesis process for base modification and sugar ring scaffold changes, forming another modified aptamer library C (88 sequences, i.e., 4×22 sequences; Figure S3). After purification using HPLC, an enzyme-linked oligonucleotide assay (ELONA) was then used to perform association and dissociation procedures against sclerostin. The relative binding ability of each modified aptamer against sclerostin was then calculated. With the addition of previous data made using the quinoline modification (154 sequences, i.e., 7×22 sequence) [24], a library with 242 sequences for base modification was constructed. A comprehensive modified aptamer library D (422 sequences, i.e., $180 + 242$ sequences; Figure S3) comprised of major influencing factors was utilized as a modified aptamer-affinity value dataset for training and verifying the established machine learning model (DeepModify model).

3.2 Performance of corrected affinity score predicted by the DeepModify model

A cost-effective regression model was applied using an iteration method for predicting the affinity score of the modified aptamers with varied modification types and sites on the sclerostin aptamer. This iterative workflow is depicted in Figure 1, where each key step of the process is clearly outlined. The workflow began by generating an initial set of modified aptamers with diverse modification types and sites, which were experimentally tested for binding affinity. BLI and ELONA were used to evaluate the affinity of each modified aptamer against sclerostin. The affinity results were then used as training data for machine learning models, which included 20 different regression models. These models were evaluated using a 10-fold cross-validation approach to determine the best candidates for constructing an ensemble (Table S2). The selected base model (RF, LightGBM, and XGBoost) were used to create a stacking ensemble to leverage the complementary strengths for improved prediction accuracy. Extensive testing was performed to optimize the model, involving various training algorithms, feature representation methods, and dataset sizes. In each iteration, new rounds of data were generated from experimental results and added to the training set, thereby continuously improving the model. This

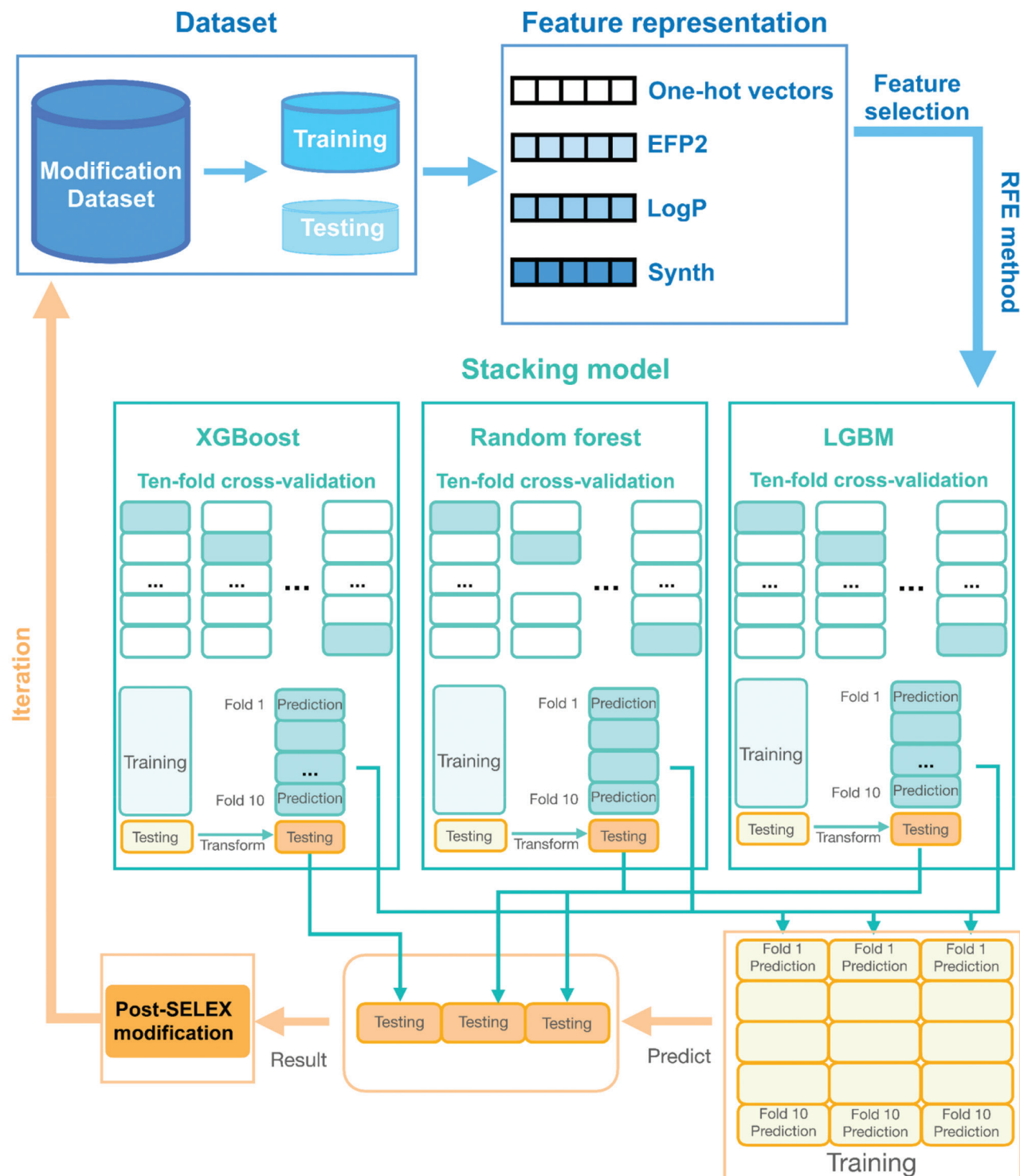


Figure 1 | Workflow of combined experimental measurement and machine learning for discovering high-affinity modification strategy in aptamers and the pharmaceutical effects.

The workflow consists of four main steps: first, the sites on each sequence are encoded as 40-dimensional vectors and the small molecules modified at the sites are represented as molecular fingerprints; second, a recursive feature elimination strategy is employed to perform feature selection for each individual base learner; third, a stacked integration classifier is trained using the training dataset, and the predicted labels for the test dataset are generated as the output; and fourth, experimental validation of the efficacy of the screened aptamers is performed.

Research Article

iterative process was repeated across multiple rounds to refine the aptamer modifications and the machine learning predictions. After establishing the model, SHapley Additive exPlanations (SHAP) values to interpret the model predictions and provide insights into which features (i.e., specific modifications and sites) had the most significant contributions to increased affinity. The final DeepModify model achieved robust performance metrics, including the PCC and R^2 , which were used to assess predictive capabilities.

The iterative nature of this workflow provided several key advantages. First, by continuously incorporating experimental results, the machine learning model was able to learn from newly synthesized aptamers and improve the predictive accuracy over time. This dynamic feedback loop enabled more precise identification of high-affinity modifications compared to a static model training approach. Second, the use of a stacking ensemble model, incorporating RF, XGBoost, and LightGBM, allowed leveraging the complementary strengths of these algorithms. This ensemble approach demonstrated improved predictive performance compared to any single model, as evidenced by superior PCC and R^2 values, which ensured a more reliable prediction of modified aptamer affinity. Finally, the incorporation of SHAP values for model interpretability ensured that the workflow was not only capable of making accurate predictions but also provided a transparent understanding of which modifications were most beneficial. This interpretability aspect was crucial for guiding future aptamer design by underlying relationships between aptamer modifications and the binding affinity. Collectively, the workflow detailed in [Figure 1](#) highlighted an efficient strategy for optimizing aptamer modifications through an interactive cycle of prediction and experimental validation, leading to the rapid development of high-affinity aptamers.

The performance of the stacking regression model was compared to RF, XGBoost, and LGBM regression models with respect to the PCC ([Figure 2a](#)). The stacking regression model outperformed the basic models with respect to prediction accuracy. The PCC values of the stacking regression model were consistently higher than the basic models, indicating superior performance in predicting the affinity score of modified aptamers. The performance of the same models was also evaluated with respect to the R^2 ([Figure 2b](#)). The stacking regression model demonstrated superior performance over the base models. The R^2 values of the stacking regression model were consistently higher, suggesting stronger robustness and reliability in predicting the affinity score. These results collectively suggested that the stacking regression model, as implemented in the DeepModify model, provided a more accurate and reliable prediction of the modified aptamer affinity score (AS, making the stacking regression model a powerful tool for developing high-affinity aptamer libraries.

ASs were calculated in the first round and 60 modified aptamers within 5 different modification types and 12

different modified sites were selected ([Figure S4](#)), which possibly improved the binding ability according to wet experiments. Several modifications helped improve the affinity of aptamers. With a total of 60 modified aptamers, a regression model was trained to predict the affinity score with an average PCC of 0.638 (averaged over 10 parallel machine learning experiments; [Figure 2c](#)). During the second round another 120 samples within 10 different modification types and 12 different modified sites were selected ([Figure S5](#)) from the prediction results of the model in the first round. Augmenting the data from the second batch to the first batch (a total of 180 samples), a regression model was retrained to update the AS with an average accuracy of 0.926 ([Figure 2d](#)). In the third round 88 modified aptamers within 4 different modification types and 22 different modification sites were selected ([Figure S6](#)). With a total of 268 experimental affinity results, a regression model was trained and predicted ASs with an average accuracy of 0.655 ([Figure 2e](#)). In the final round another 154 modified aptamers within 7 different modification types and 22 different modification sites were selected ([Figure S7](#)). A total of 422 aptamers with post-SELEX modifications were experimentally synthesized and the binding affinity was evaluated. Using these experimental affinity results, a regression model (DeepModify model) was finally achieved with an average accuracy of 0.822 ([Figure 2f](#)). The top 300 modified aptamers are listed for the convenience of selection and comparison. The difference in ASs of aptamers with post-SELEX modifications synthesized in the first batch of experiments and those predicted in the second, third, and final batches were compared. It is noteworthy that the average affinity scores significantly enhanced as the batches increased (0.82 for first batch, 1.77 for second batch, 3.27 for third batch, and 6.89 for last batch; [Figure 2g](#)). These findings indicated that the model was effectively learning from each round of experimental data and improved the predictive accuracy with each round. This continuous improvement in prediction performance underscores the power of machine learning models in aiding the development of high-affinity aptamer libraries.

Taken together, when using DeepModify trained with a dataset of > 422 samples represented by 40-bit one-hot vectors for modification sites and fingerprint for each small molecule ([Figure S8](#)), a robust stacking model with training/testing performance of 0.822 in PCC_{tr}/PCC_{te} was achieved. Furthermore, the prediction performance analysis of the RF model suggested that the PCC between the predicted AS (AS_{prd}) and the validated AS (AS_{val}) was > 80.0% because the AS_{val} was > 10.0. These results indicated the reliability of the selected model in accurately predicting the affinity of modified aptamers.

A subsequent analysis focused on the prediction of modified aptamers. A cluster analysis was performed to obtain a comprehensive view of the ASs for all modification types and sites. The distances between different classes were calculated and visualized in the cluster

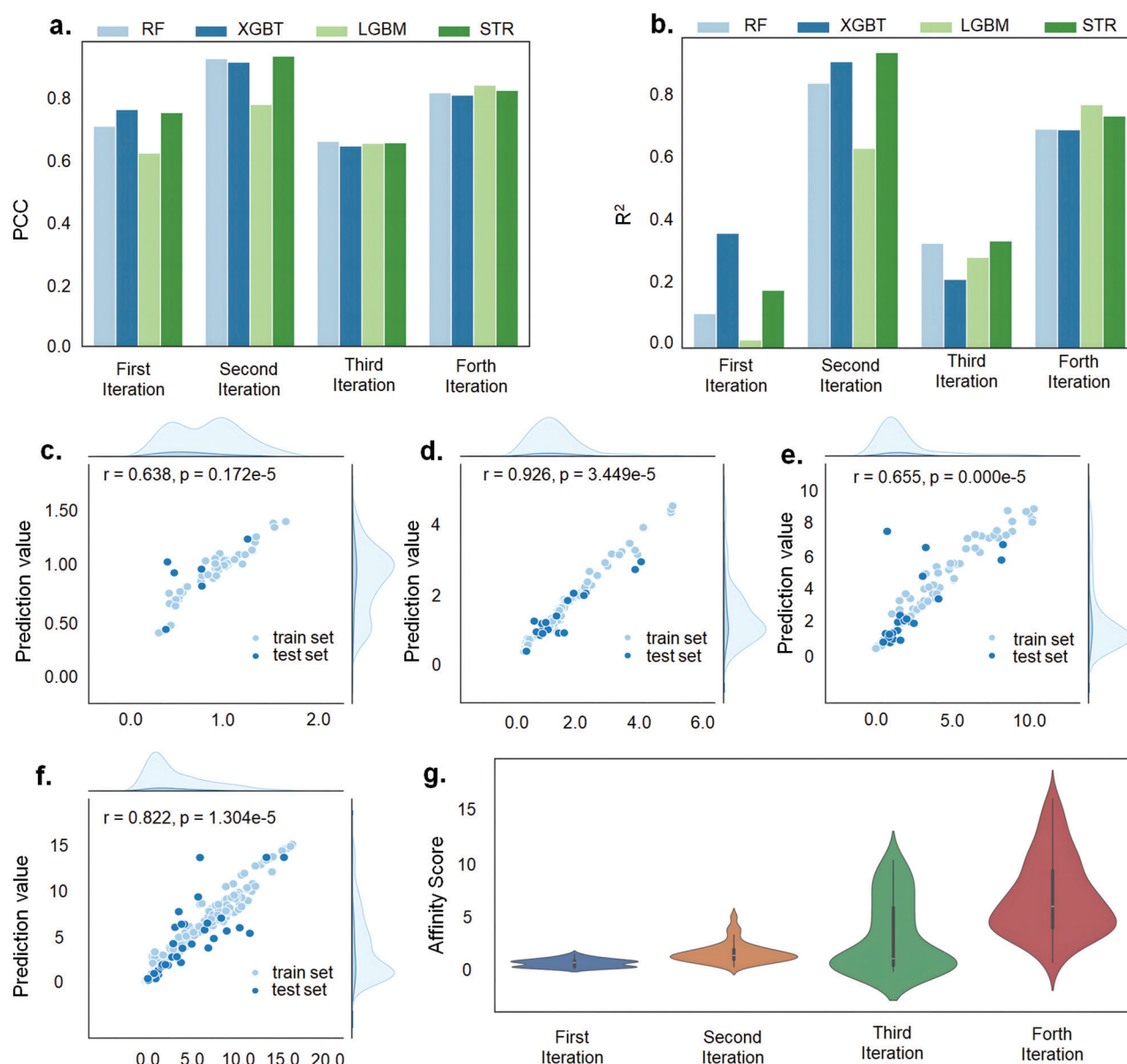


Figure 2 | Performance evaluation of DeepModify on affinity prediction.

(a-b). The performance of the stacking regression model with that of Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LGBM) regression models in terms of PCC (a) and R^2 (b). (c-f) Pearson's correlation coefficients between the predicted and the actual affinity scores of the models in the first (c), second (d), third (e), and last (f) batches. The correlation coefficient values and P -values labelled in the figure are the results from the test set to check the performance of the model. (g) The comparison of the difference in affinity scores of aptamers with post-SELEX modifications synthesized in the first batch of experiments and predicted in the second, third, and last batches.

analysis (Figure S9a). This finding provided a clear representation of the relationship between different modifications. The K-Means algorithm was applied for the clustering process. To determine the optimal number of clusters, the distortion score elbow plot for K-Means was generated (Figure S9b). This plot aided in identifying the best number of clusters. The silhouette score elbow for K-Means was also plotted (Figure S9c). The

silhouette score is a metric used to assess the similarity of a cluster object in comparison to other clusters. A higher silhouette score suggests that the object is matched to the assigned cluster and is not matched to the neighboring clusters. The silhouette score serves as a measure of the quality of clustering, with higher values indicating more distinct and well-separated clusters. All modification types and sites were divided into 5 clusters (clusters

Research Article

0–4; **Figure 3a**). Clusters of specific modifications and sites were associated with higher ASs, suggesting that these specific modifications and sites are important in increasing aptamer binding activity (**Figure 3b**). This cluster analysis not only provided a comprehensive overview of the ASs of all modifications and sites but also helped to select the most promising modifications and sites for further studies.

An in-depth interpretability analysis of the Deep-Modify model was performed using the SHAP values [33], which provide a measure of the importance of each feature in the model (**Figure S9d**). The SHAP values

were calculated for each modified site, allowing for an assessment of the contribution of each site to the model output. This analysis provided valuable insights into the relative importance of different modifications and sites in determining the affinity of modified aptamers. Furthermore, a 3D distribution visualization was performed to determine the affinity score distribution of different classes of modified aptamers. This analysis revealed the distribution of ASs for all modification types of aptamers on different nucleotide sites. As depicted in **Figure 3c and 3d**, both modification types and sites exhibited varying ASs, indicating the distinct

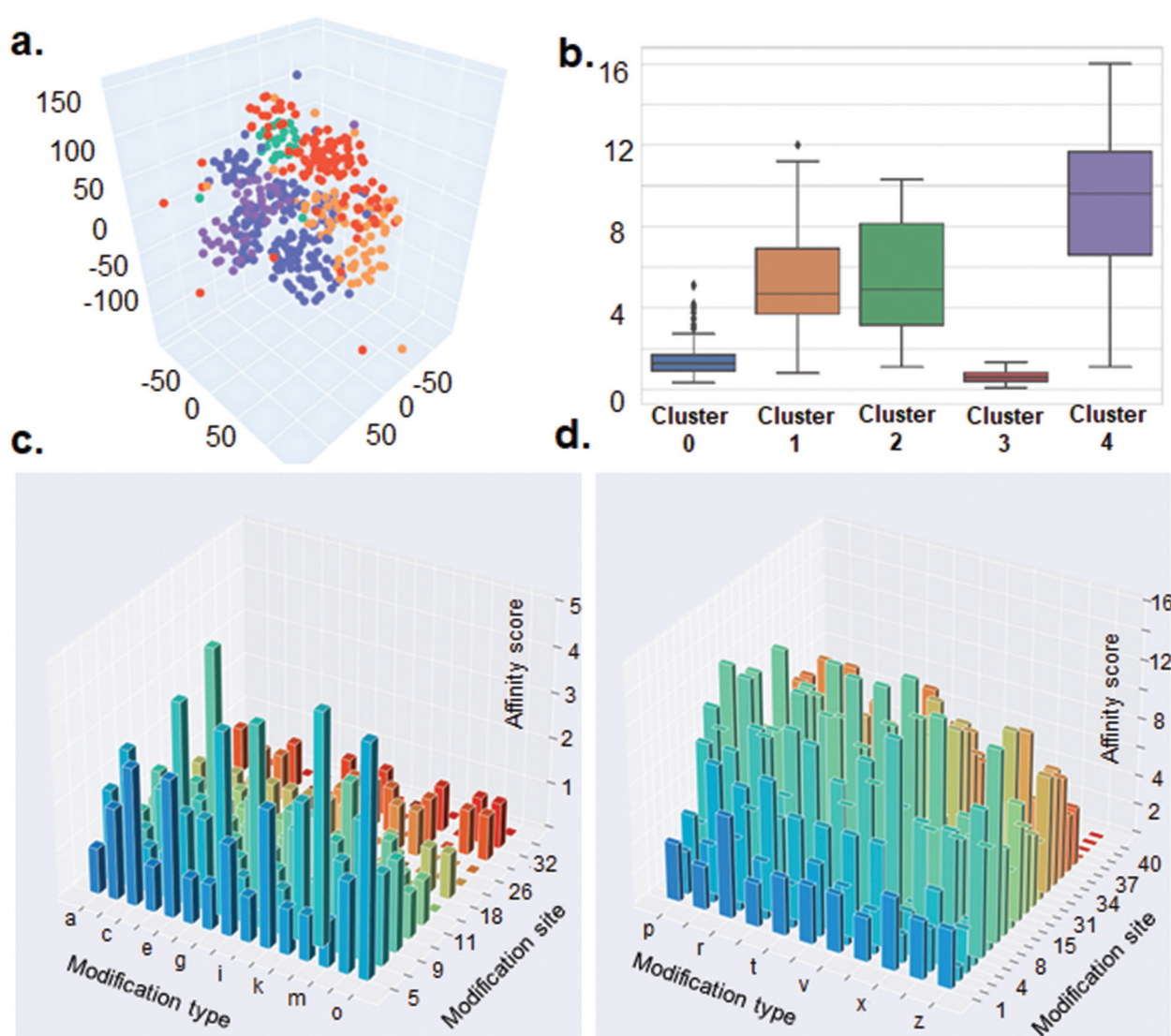


Figure 3 | Differential analysis of feature distribution and affinity scores of different classes of modified aptamers in three-dimensional space.

(a) Feature distribution of modified aptamers between different classes in three-dimensional space. (b) Comparison of the differences in affinity scores of different classes of modified aptamers. (c-d) Distribution of affinity scores for all modification types of aptamers after modification at different sites.

roles in aptamer binding activity. These results demonstrated that specific modifications and sites contributed to higher predicted affinity, thereby potentially increasing aptamer binding activity. The model showed that modifications incorporating hydrophobic groups, such as quinoline derivate or aromatic amino acids at G4, G9, and T11, significantly enhanced ASs. The analysis indicated that introducing a modification at G38 led to a notable decrease in ASs. The identification of such patterns suggests that the model has effectively learned that while certain modifications enhance aptamer functionality, the benefits are highly dependent on the nature of the modification types and modification sites within the aptamer sequence.

3.3 Discovery and affinity characterization of DeepModify-predicting aptamers with post-SELEX modifications

Higher ASs using the DeepModify model revealed significantly increased aptamer binding affinity with post-SELEX modifications, such as Aptscl56-9b ($K_D = 12.6$ nM), Aptscl56-4s ($K_D = 13.5$ nM), Aptscl56-6s ($K_D = 5.88$ nM), Aptscl56-8s ($K_D = 5.62$ nM), and Aptscl56-11s ($K_D = 5.28$ nM). By contrast, lower ASs revealed significantly decreased aptamer binding affinity with post-SELEX modifications, such as Aptscl56-7i ($K_D = 121.6$ nM), Aptscl56-5k ($K_D = 160.3$ nM), Aptscl56-32o ($K_D = 196.5$ nM), Aptscl56-35o ($K_D = 205.7$ nM; Table 1, Figure S10). The binding modes between aptamers and sclerostin were investigated using the molecular docking assay (Figure 4a). Compared to the docking binary complex formed by unmodified aptamer (Aptscl56) and

sclerostin (Table S3), there were more interactions in the docking binary complex formed by modified aptamer (Aptscl56-9b and Aptscl56-11s) and sclerostin (Tables S4 and S5). The binding affinity of aptamers with dual modifications (Aptscl56-9b11s [$K_D = 0.634$ nM]) was superior to Aptscl56-9b4s ($K_D = 18.4$ nM), Aptscl56-9b6s ($K_D = 1.20$ nM), or Aptscl56-9b8s ($K_D = 1.22$ nM). No binding was observed between Aptscl56-9b11s and control protein (human serum albumin [HSA]), suggesting that this chemical modification would not enhance non-specific binding (Figure S11). Furthermore, the docking results showed that the binary complex formed by a dual-modified aptamer (Aptscl56-9b11s) and sclerostin had the lowest root mean square deviation [RMSD] (41.57 Å), the largest hydrogen bond number ($n = 9$), and the shortest hydrogen bond length (2.51 Å), as well as the largest interaction diversity (hydrophobic interactions, hydrogen bond, and salt bridges; Table S6) compared to unmodified aptamer (Aptscl56) and single-modified aptamers (Aptscl56-9b and Aptscl56-11s). Notably, Aptscl56-9b11s exhibited a 105-fold affinity enhancement compared to naturally unmodified aptamer (Aptscl56 [$K_D = 62.8$ nM]), indicating that the DeepModify model had enormous potential in the selected aptamer affinity optimization.

3.4 DeepModify-predicting aptamers with post-SELEX modifications could attenuate the inhibitory effect of the transfected sclerostin on Wnt signaling and promote mRNA expression of bone formation markers *in vitro*

The binding affinity of aptamer drug candidates to the targets has a crucial role in pharmacodynamics. To

Table 1 | The measured K_D values of the aptamers with post-SELEX modifications against sclerostin.

Name	Sequence	K_D (nM)	AS_{val}
Aptscl56	CGGGGTGTGGGTTCTGCTTAGCTTGATTGGCAGCTGCC	66.4	1.0
Aptscl56-9b	CGGGGTGT b GGTTCGTCGTTAGCTTGATTGGCAGCTGCC	12.6	5.3
Aptscl56-4s	CGG s GTGTGGGTTCTGCTTAGCTTGATTGGCAGCTGCC	13.5	4.9
Aptscl56-6s	CGGGG s GTGGGTTCTGCTTAGCTTGATTGGCAGCTGCC	5.88	11.3
Aptscl56-8s	CGGGGTG s GGGTTCTGCTTAGCTTGATTGGCAGCTGCC	5.62	11.8
Aptscl56-11s	CGGGGTGTGG s TTCGTCGTTAGCTTGATTGGCAGCTGCC	5.28	12.6
Aptscl56-7i	CGGGGT i TGGGTTCTGCTTAGCTTGATTGGCAGCTGCCT	121.6	0.5
Aptscl56-5k	CGGG k TGTGGGTTCTGCTTAGCTTGATTGGCAGCTGCC	160.3	0.4
Aptscl56-32o	CGGGGTGTGGGTTCTGCTTAGCTTGATT o CAGCTGCC	196.5	0.3
Aptscl56-35o	CGGGGTGTGGGTTCTGCTTAGCTTGATTGGC o CTGCC	205.7	0.3
Aptscl56-9b4s	CGG s GTGT b GGTTCGTCGTTAGCTTGATTGGCAGCTGCC	18.4	3.6
Aptscl56-9b6s	CGGGG s GT b GGTTCGTCGTTAGCTTGATTGGCAGCTGCC	1.20	55.3
Aptscl56-9b8s	CGGGGTG sb GGTTCGTCGTTAGCTTGATTGGCAGCTGCC	1.22	54.4
Aptscl56-9b11s	CGGGGTGT bGs TTCGTCGTTAGCTTGATTGGCAGCTGCC	0.634	104.7

Research Article

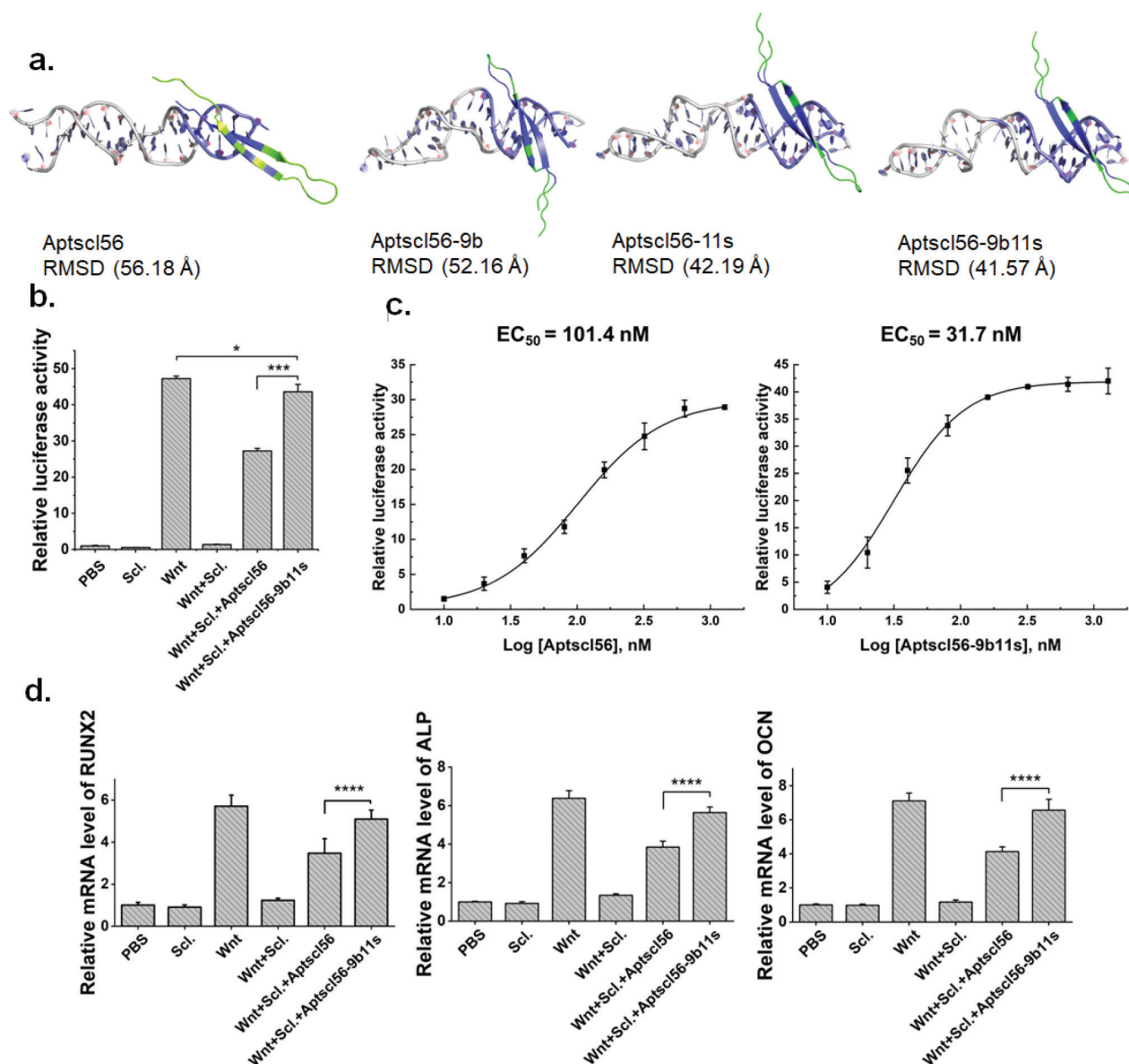


Figure 4 | The dynamic simulation and therapeutic efficacy of that DeepModify model-predicted Aptscl56-9b11s.

(a) The molecular dynamic simulation of naturally unmodified aptamer (Aptscl56), aptamers with post-SELEX modifications (Aptscl56-9b, Aptscl56-11s, and Aptscl56-9b11s). (b) The effects of Aptscl56 and Aptscl56-9b11s on the transfected sclerostin-induced suppressed Wnt signaling in HEK293 ($n = 3$). (c) EC₅₀ measurements of Aptscl56 and Aptscl56-9b11s on the transfected sclerostin-induced suppressed Wnt signaling in HEK293 cells ($n = 3$). (d) The effects of Aptscl56 and Aptscl56-9b11s on downregulation of mRNA expression of bone formation markers (*Runx2*, *Alp*, and *Ocn*) in MC3T3-E1 cells ($n = 3$). * $P < 0.05$, *** $P < 0.001$, **** $P < 0.0001$.

evaluate the therapeutic efficacy of the aptamer with the highest affinity, the inhibitory effect of Aptscl56-9b11s on sclerostin-suppressed Wnt signaling was measured. For this purpose, HEK293 cells were co-transfected with the Topflash (Firefly luciferase reporter regulated by TCF/LEF transcription factors) and SV40 (Renilla luciferase reporter [inner control]), Wnt-1, and sclerostin plasmids. After sclerostin inhibitor treatment, the Wnt signaling that was suppressed by the transfected

sclerostin could be reactivated. The *in vitro* data showed that the luciferase signaling induced by Wnt-1 (47.2-fold) was significantly attenuated by sclerostin (1.4-fold), suggesting that a model of sclerostin-suppressed Wnt signaling was successfully constructed. The luciferase signaling in the Aptscl56-9b11s treatment group was greater than the Aptscl56 treatment group (43.6-fold vs 27.2-fold; $P < 0.001$; Figure 4b). The EC₅₀ value of Aptscl56 was 101.4 nM, whereas the EC₅₀ value of

Aptsc156-9b11s was 31.7 nM (Figure 4c), which indicated that DeepModify-predicting aptamers with post-SELEX modifications attenuated the inhibitory effect of the transfected sclerostin on Wnt signaling *in vitro*.

MC3T3-E1 cells (preosteoblast-like cells) were used to evaluate the impact of Aptsc156-9b11s on downregulation of bone formation marker (*Runx2*, *Alp*, and *Ocn*) mRNA expression induced by sclerostin. Co-transfection of Wnt-1 and human sclerostin plasmids in MC3T3-E1 cells was performed, then Aptsc156 or Aptsc156-9b11s was added to the cells. Bone formation marker mRNA expression was then quantified using real-time qPCR. Consistent with the results of the luciferase assay, treatment with Aptsc156-9b11s led to a remarkable recovery in the relative level of mRNA expression induced by Wnt-1 (85.3% for *Runx2*, 90.9% for *Alp*, and 86.4% for *Ocn*). Notably, these levels were >1.7-fold higher compared to treatment with Aptsc156 (49.8% for *Runx2*, 49.9% for *Alp*, and 49.9% for *Ocn*; Figure 4d). Therefore, DeepModify-predicting aptamers with post-SELEX modifications promoted bone formation marker (*Runx2*, *Alp*, and *Ocn*) mRNA expression *in vitro*.

Collectively, all data indicated that the DeepModify model successfully predicted high-affinity aptamers with post-SELEX modifications, which attenuated the transfected sclerostin inhibitory function on Wnt signaling and promoted bone formation marker (*Runx2*, *Alp*, and *Ocn*) mRNA expression *in vitro*.

4. DISCUSSION

Aptamers have emerged as promising biosensors and therapeutic candidates due to target specificity [34, 35]. The binding affinity to the target has a critical role in aptamer efficacy for successful application [36]. Generally, post-SELEX modifications are necessary to improve the binding affinity of selected aptamers and compensate for inherent limitations. However, the large number of modification sites and diverse modification options make it impractical to exhaustively explore all possibilities manually. In the current study a deep learning algorithm (DeepModify) was established to efficiently predict aptamer binding affinity with post-SELEX modifications.

A total of 242 modified aptamers were synthesized by solid-phase phosphoramidite chemistry for training data with as much information as possible, while 180 modified aptamers were synthesized by liquid-phase CuAAC chemistry. A modified library with 422 modified sequences was constructed and the corresponding binding affinities to the target protein were measured for a large dataset. The DeepModify model trained on this dataset used a machine learning architecture, including different training algorithms and feature representation approaches. Using the stacking algorithm with > 422 training data represented by 40-bit one-hot vectors for modification sites and fingerprints for each candidate small molecule yielded a reliable stacking model.

In addition to the development and application of the DeepModify model, a thorough evaluation of the predictive performance was performed using various regression metrics, such as PCC, R^2 , and other error metrics (mean absolute error and root mean squared error). The stacking model demonstrated consistent superiority over the base models (RF, XGBoost, LightGBM), indicating an enhanced capability in predicting modified aptamer affinity. This comprehensive evaluation supports the reliability and robustness of the model. Furthermore, interpretability analysis using SHAP values provided additional insight into the significance of different modification types and sites, elucidating how specific features contribute to binding affinity. This interpretability is crucial for understanding the underlying biological mechanisms and ensuring that the model predictions are scientifically meaningful and actionable. By emphasizing these evaluation results, the machine learning approach not only accurately predicted high-affinity modifications but also offered valuable biological insight, making the machine learning approach a powerful tool for optimizing aptamer modifications.

Finally, a classification model was trained to predict the AS of these aptamers. The first and second iterations focused on 2'-hydroxyl modification, resulting in the AI model demonstrating strong PCCs from 0.638–0.926. However, base modification and nucleo-sugar scaffold changes were introduced in the third iteration, leading to significant structural variations that subsequently decreased the PCC (0.655). In the final iteration, a total of 422 aptamers further increased the PCC (0.822). Notably, the average ASs significantly increased with each batch of modified aptamers (0.82 for the first batch, 1.77 for the second batch, 3.27 for the third batch, and 6.89 for the last batch). Collectively, a DeepModify model for predicting the AS of modified aptamers was successfully developed and validated. The model demonstrated reliable performance and showed promise for guiding the design of high-affinity aptamer libraries.

The effects of post-SELEX modifications on modified aptamer affinity was identified using the DeepModify model. For example, Aptsc156-9b, Aptsc156-4s, Aptsc156-6s, Aptsc156-8s, and Aptsc156-11s showed improved binding affinities with K_D values of 12.6 nM, 13.5 nM, 5.88 nM, 5.62 nM, and 5.28 nM, respectively. Conversely, Aptsc156-7i, Aptsc156-5k, Aptsc156-32o, and Aptsc156-35o had K_D values of 121.6 nM, 160.3 nM, 196.5 nM, and 205.7 nM, respectively. Notably, a superimposed effect on binding affinity was observed when specific types of dual post-SELEX modifications were applied. Aptsc156-9b11s ($K_D = 0.634$ nM) exhibited a 105-fold affinity enhancement compared to naturally unmodified aptamers (Aptsc156 [$K_D = 62.8$ nM]). The enhancement effect of chemical modifications on binding affinity can be attributed to the following factors: (1) The high-affinity modified aptamer (Aptsc156-9b11s) exhibited a broader spectrum of interaction types, encompassing hydrophobic interactions,

Research Article

hydrogen bonds, and salt bridges (Table S6). In contrast, the unmodified aptamer (Aptsc156) primarily featured hydrogen bonds and salt bridges (Table S3). The modifications in Aptsc156-9b11s potentially introduced novel functional groups or induced conformational changes in the aptamer that enabled a more diverse set of interactions with sclerostin. (2) The interaction count of Aptsc156-9b11s to sclerostin was 15 (Table S6), whereas the interaction count of Aptsc156 to sclerostin was 12 (Table S3). The modifications in Aptsc156-9b11s potentially introduced additional binding sites that actively participated in the interaction with sclerostin. (3) The RMSD value for Aptsc156-9b11s was 41.57 Å, whereas the RMSD for Aptsc156 was 56.18 Å (Figure 4a). The modifications in Aptsc156-9b11s contributed to enhancing the structural stability of the aptamer, potentially resulting in a more robust and reliable binding interface.

To evaluate the therapeutic efficacy of Aptsc156-9b11s, the effects of Aptsc156-9b11s on the suppressed Wnt-signal by the transfected sclerostin and bone formation marker expression were investigated. Luciferase signaling was higher in the Aptsc156-9b11s treatment group compared to the Aptsc156 treatment group, suggesting that Aptsc156-9b11s significantly reactivated Wnt signaling that was suppressed by sclerostin. Compared to Aptsc156, Aptsc156-9b11s also significantly promoted the levels of bone formation marker expression, indicating the potential to enhance bone formation processes. *Runx2* is a key transcription factor involved in osteoblast differentiation [37], while *Alp* and *Ocn* are markers of osteoblast activity and maturation, respectively [38]. Upregulation of these markers suggests that Aptsc156-9b11s enhances the differentiation and activity of preosteoblasts, leading to increased bone formation. Collectively, Aptsc156-9b11s not only attenuated the suppressive effect of sclerostin on Wnt signaling but also promoted bone formation marker expression, indicating the potential to enhance bone formation processes. This dual mechanism of action makes Aptsc156-9b11s a promising candidate for therapeutic interventions targeting bone-related disorders.

Traditional approaches rely heavily on manual experimentation and intuition, making it challenging to explore the vast array of potential modification sites and options. In contrast, the DeepModify model provides a significant advance by leveraging deep learning to predict the binding affinity of aptamers with post-SELEX modifications. Compared to traditional methods, the DeepModify model enables efficient prediction of binding affinity, significantly reducing the need for extensive manual experimentation. Moreover, the DeepModify model provides data-driven insight into the effects of post-SELEX modifications on aptamer binding affinity. Overall, the DeepModify model offers significant improvement over traditional approaches by providing a more efficient, accurate, and reliable method for predicting the binding affinity of aptamers with

post-SELEX modifications. This advancement paves the way for the development of high-affinity aptamers with enhanced therapeutic potential and probe sensitivity.

DECLARATION OF COMPETING INTERESTS

G.A., X.Y., Y.P., and Y.H. were employed by Aptacure Therapeutics Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Young Scientists Fund of the Natural Science Foundation of Zhejiang Province [grant number LQN25H300006], Start-up Fund of The First Affiliated Hospital of Zhejiang University School of Medicine [grant number BQD2423], Theme-based Research Scheme [grant number T12-201-20R], Guangdong Basic and Applied Basic Research Foundation [grant number 2020A15110630], National Key R&D Program of China [grant number 2018YFA0800802], and the CUHK Direct Grant [grant number 4054714].

SUPPLEMENTARY DATA

Supplementary data for this article can be found online at: https://amm-journal.org/wp-content/uploads/2025/01/amm20240065_Suppl.pdf.

REFERENCES

- [1] Zhao QG, Zhou YJ, Cao DX, Tang AN, Kong DM: DNA-Functionalized Porphyrinic Metal-Organic Framework-Based Drug Delivery System for Targeted Bimodal Cancer Therapy. *Journal of Medicinal Chemistry* 2023, 66:15370–15379.
- [2] Bashir A, Yang Q, Wang J, Hoyer S, Chou W, McLean C, et al.: Machine Learning Guided Aptamer Refinement and Discovery. *Nature Communications* 2021, 12:2366.
- [3] Chen Z, Luo H, Gubu A, Yu S, Zhang H, Dai H, et al.: Chemically Modified Aptamers for Improving Binding Affinity to the Target Proteins via Enhanced Non-covalent Bonding. *Frontiers in Cell and Developmental Biology* 2023, 11:1091809.
- [4] Alkhamis O, Canoura J, Wu Y, Emmons NA, Wang Y, Honeywell KM, et al.: High-Affinity Aptamers for In Vitro and In Vivo Cocaine Sensing. *Journal of the American Chemical Society* 2024, 146:3230–3240.
- [5] Ma Y, Zhang H, Shen X, Yang X, Deng Y, Tian Y, et al.: Aptamer Functionalized Hypoxia-potentiating Agent and Hypoxia-inducible Factor Inhibitor Combined with Hypoxia-activated Prodrug for Enhanced Tumor Therapy. *Cancer Letters* 2024, 598:217102.
- [6] Chan KY, Kinghorn AB, Hollenstein M, Tanner JA: Chemical Modifications for a Next Generation of Nucleic Acid Aptamers. *ChemBioChem: A European Journal of Chemical Biology* 2022, 23:e202200006.
- [7] Elskens JP, Elskens JM, Madder A: Chemical Modification of Aptamers for Increased Binding Affinity in Diagnostic Applications: Current Status and Future Prospects. *International Journal of Molecular Sciences* 2020, 21:4522.

- [8] Hasegawa H, Savory N, Abe K, Ikebukuro K: Methods for Improving Aptamer Binding Affinity. *Molecules* 2016, 21:421.
- [9] Davies DR, Gelinas AD, Zhang C, Rohloff JC, Carter JD, O'Connell D, et al.: Unique Motifs and Hydrophobic Interactions Shape the Binding of Modified DNA Ligands to Protein Targets. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109:19971–19976.
- [10] Gawande BN, Rohloff JC, Carter JD, von Carlowitz I, Zhang C, Schneider DJ, et al.: Selection of DNA Aptamers with Two Modified Bases. *Proceedings of the National Academy of Sciences of the United States of America* 2017, 114:2898–2903.
- [11] Wu D, Katilius E, Olivas E, Dumont Milutinovic M, Walt DR: Incorporation of Slow Off-Rate Modified Aptamers Reagents in Single Molecule Array Assays for Cytokine Detection with Ultrahigh Sensitivity. *Analytical Chemistry* 2016, 88:8385–8389.
- [12] Ji D, Feng H, Liew SW, Kwok CK: Modified Nucleic Acid Aptamers: Development, Characterization, and Biological Applications. *Trends in Biotechnology* 2023, 41:1360–1384.
- [13] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK: Artificial Intelligence in Drug Discovery and Development. *Drug Discovery Today* 2021, 26:80–93.
- [14] Ballarotto M, Willems S, Stiller T, Nawa F, Marschner JA, Grisoni F, et al.: De Novo Design of Nurr1 Agonists via Fragment-Augmented Generative Deep Learning in Low-Data Regime. *Journal of Medicinal Chemistry* 2023, 66:8170–8177.
- [15] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al.: Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. *Nature Biotechnology* 2019, 37:1038–1040.
- [16] Knight CG, Platt M, Rowe W, Wedge DC, Khan F, Day PJ, et al.: Array-based Evolution of DNA Aptamers Allows Modelling of An Explicit Sequence-fitness Landscape. *Nucleic Acids Research* 2009, 37:e6.
- [17] Iwano N, Adachi T, Aoki K, Nakamura Y, Hamada M: Generative Aptamer Discovery Using RaptGen. *Nature Computational Science* 2022, 2:378–386.
- [18] Yu Y, Wang L, Ni S, Li D, Liu J, Chu HY, et al.: Targeting Loop3 of Sclerostin Preserves Its Cardiovascular Protective Action and Promotes Bone Formation. *Nature Communications* 2022, 13:4241.
- [19] Wang L, Yu Y, Ni S, Li D, Liu J, Xie D, et al.: Therapeutic Aptamer Targeting Sclerostin Loop3 for Promoting Bone Formation Without Increasing Cardiovascular Risk in Osteogenesis Imperfecta Mice. *Theranostics* 2022, 12:5645–5674.
- [20] Zhang H, Yu S, Ni S, Gubu A, Ma Y, Zhang Y, et al.: A Bimolecular Modification Strategy for Developing Long-lasting Bone Anabolic Aptamer. *Molecular Therapy. Nucleic Acids* 2023, 34:102073.
- [21] Amu G, Zhang G, Jing N, Ma Y: Developing Stapled Aptamers with a Constrained Conformation for Osteogenesis Imperfect Therapeutics. *Journal of Medicinal Chemistry* 2024, 67:18883–18894.
- [22] Shen WX, Liu Y, Chen Y, Zeng X, Tan Y, Jiang YY, et al.: AggMapNet: Enhanced and Explainable Low-sample Omics Deep Learning with Feature-aggregated Multi-channel Networks. *Nucleic Acids Research* 2022, 50:e45.
- [23] Zhou Z-H: *Ensemble Methods: Foundations and Algorithms*. New York: CRC Press; 2012.
- [24] Gubu A, Ma Y, Yu S, Zhang H, Chen Z, Ni S, et al.: Unique Quinoline Orientations Shape the Modified Aptamer to Sclerostin for Enhanced Binding Affinity and Bone Anabolic Potential. *Molecular Therapy. Nucleic Acids* 2024, 35:102146.
- [25] Ma Y, Zhao W, Li Y, Pan Y, Wang S, Zhu Y, et al.: Structural Optimization and Additional Targets Identification of Antisense Oligonucleotide G3139 Encapsulated in a Neutral Cytidiny-lipid Combined with a Cationic Lipid In Vitro and In Vivo. *Biomaterials* 2019, 197:182–193.
- [26] Shahrezaee M, Salehi M, Keshtkari S, Oryan A, Kamali A, Shekarchi B: In Vitro and In Vivo Investigation of PLA/PCL Scaffold Coated with Metformin-loaded Gelatin Nanocarriers in Regeneration of Critical-sized Bone Defects. *Nanomedicine: Nanotechnology, Biology, and Medicine* 2018, 14:2061–2073.
- [27] Sugatani T, Agapova O, Malluche HH, Hruska KA: SIRT6 Deficiency Culminates in Low-turnover Osteopenia. *Bone* 2015, 81:168–177.
- [28] Koblan LW, Erdos MR, Wilson C, Cabral WA, Levy JM, Xiong ZM, et al.: In Vivo Base Editing Rescues Hutchinson-Gilford Progeria Syndrome in Mice. *Nature* 2021, 589:608–614.
- [29] Kim W, Song ES, Lee SH, Yang SH, Cho J, Kim S-J: A New DNA Aptamer Which Binds to SARS-CoV-2 Spike Protein and Reduces Pro-inflammatory Response. *Scientific Reports* 2024, 14:7516.
- [30] Poojara LRK, Rawal RM: Computational Approaches Screening DNA Aptamers Against Conserved Outer Membrane Protein W of *Vibrio Cholerae* O1- An Investigation Expanding the Potential for Point-of-Care Detection with Aptasensors. *Journal of Biomolecular Structure & Dynamics* 2023, 41:14438–14449.
- [31] Yan Y, Tao H, He J, Huang SY: The HDock Server for Integrated Protein-Protein Docking. *Nature Protocols* 2020, 15:1829–1852.
- [32] Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, et al.: PLIP 2021: Expanding the Scope of the Protein-ligand Interaction Profiler to DNA and RNA. *Nucleic Acids Research* 2021, 49: W530–W534.
- [33] Aldrees A, Khan M, Taha ATB, Ali M: Evaluation of Water Quality Indexes with Novel Machine Learning and SHapley Additive ExPlanation (SHAP) Approaches. *Journal of Water Process Engineering* 2024, 58:104789.
- [34] Ma Y, Zhu Y, Wang C, Pan D, Liu S, Yang M, et al.: Annealing Novel Nucleobase-lipids with Oligonucleotides or Plasmid DNA Based on H-bonding or π - π Interaction: Assemblies and Transfections. *Biomaterials* 2018, 178:147–157.
- [35] Niu B, Wu Y, Zhou M, Lin R, Ge P, Chen X, et al.: Precise Delivery of Celastrol by PEGylated Aptamer Dendrimer Nanoconjugates for Enormous Therapeutic Effect via Superior Intratumor Penetration Over Antibody Counterparts. *Cancer Letters* 2023, 579:216461.
- [36] Ma Y, Xie D, Chen Z, Shen X, Wu X, Ding F, et al.: Advancing Targeted Combination Chemotherapy in Triple Negative Breast Cancer: Nucleolin Aptamer-mediated Controlled Drug Release. *Journal of Translational Medicine* 2024, 22:604.
- [37] Tang W, Xu J, Xu C: Noncoding RNAs in the Crosstalk between Multiple Myeloma Cells and Bone Marrow Microenvironment. *Cancer Letters* 2023, 556:216081.
- [38] Cong W, Shen H, Jiang Y, Li L, Kong X, Chen S, et al.: Design, Synthesis, and Anti-Osteoporotic Characterization of Arginine N-Glycosylated Teriparatide Analogs via the Silver-catalyzed Solid-Phase Glycosylation Strategy. *Journal of Medicinal Chemistry* 2024, 67:1360–1369.