Acta Pharmaceutica Sinica B 2026;16(7):4552−4574

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

AI-powered therapeutic aptamer drug discovery: Targeting the CT-domain of CTGF for duchenne muscular dystrophy

Shanshan Yao[a,†], Xin Yang[b,c,†], Meishen Ren[b,d,e,†], Huarui Zhang[a,†], Sifan Yu[a,b,d,f,*,†], Tienan Chen[g], Shijian Ding[c], Yufei Pan[e], Luyao Wang[b], Yuan Ma[b], Wei Kang[h], Xiaoling Peng[i], Jiming Liu[j], Aiping Lyu[c], Zongkang Zhang[a,d,*], Yuanyuan Yu[c,d,*], Yixin He[e,*], Ge Zhang[b,d,*], Bao-Ting Zhang[a,d,*]

[a]*School of Chinese Medicine, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR 999077, China*
[b]*Law Sau Fai Institute for Advancing Translational Medicine in Bone and Joint Diseases (TMBJ), School of Chinese Medicine, Hong Kong Baptist University, Hong Kong SAR 999077, China*
[c]*Institute of Integrated Bioinformedicine and Translational Science (IBTS), School of Chinese Medicine, Hong Kong Baptist University, Hong Kong SAR 999077, China*
[d]*Guangdong-Hong Kong-Macao Greater Bay Area International Research Platform for Aptamer-based Translational Medicine and Drug Discovery (HKAP), Hong Kong SAR 999077, China*
[e]*Aptacure Therapeutics Limited, Hong Kong SAR 999077, China*
[f]*Shenzhen Institute for Research and Continuing Education (IRACE), Hong Kong Baptist University, Shenzhen 518000, China*
[g]*Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200030, China*
[h]*Department of Anatomical and Cellular Pathology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR 999077, China*
[i]*Beijing Normal University-Hong Kong Baptist University, Division of Science and Technology, Zhuhai 519000, China*
[j]*Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR 999077, China*

Received 11 August 2025; received in revised form 6 November 2025; accepted 13 November 2025

*Corresponding authors.
E-mail addresses: yusifan123@gmail.com (Sifan Yu), maxzhangzk@cuhk.edu.hk (Zongkang Zhang), yuyuanyuan@hkbu.edu.hk (Yuanyuan Yu), berryhe@aptacure.com (Yixin He), zhangge@hkbu.edu.hk (Ge Zhang), zhangbaoting@cuhk.edu.hk (Bao-Ting Zhang).
[†]These authors made equal contributions to this work.
Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2026.02.016

KEY WORDS

AI-Powered drug discovery;
Aptamer;
CTGF;
DMD;
Muscle fibrosis

Abstract Duchenne muscular dystrophy (DMD), a fatal X-linked disorder, features progressive muscle fibrosis as a key driver of mortality. While CTGF represents a therapeutic target for DMD, its VWC-domain-targeting antibody (FG-3019) failed in clinical trials. Through experimental validation, we identified CT-domain as a superior target domain, as it contributed more fibrosis activity of CTGF than VWC-domain without elevating compensatory TGF-β1 level. Aptamers are synthetic oligonucleotides identified through SELEX, which can specifically bind to flexible protein domains through their unique 3D conformations. Their small molecular size enables effective tissue penetration while maintaining high target specificity, making them ideal CT-domain inhibitors. Nevertheless, conventional SELEX involves time-consuming and inefficient multiple screening rounds. Here, we employed our generative AI model, AptGEN, to rapidly discover a potent CT-domain specific aptamer within 42 days. This chemically modified aptamer (Apc003OA) distributed and remained in muscle tissues for an extended period, whereas FG-3019 could not. Importantly, it demonstrated better fibrosis inhibitory activity *in vitro* and in mdx mice when compared to FG-3019. Furthermore, Apc003OA demonstrated a favorable safety profile in mdx mice. Within 10 months, we progressed from target domain discovery, aptamer drug discovery, and then obtained both Orphan Drug Designation and Pediatric Rare Disease Designation by US Food and Drug Administration.

1. Introduction

Duchenne muscular dystrophy (DMD) is a lethal X-linked genetic disorder caused by mutations in the dystrophin gene, leading to progressive muscle degeneration, fibrosis, and premature mortality[1,2]. Although dystrophin gene-targeting strategies aimed at restoring dystrophin expression hold promise, they face significant challenges and have yielded limited success[3].

Recently, approaches addressing underlying pathological changes of DMD have received attentions[4]. Fibrosis is one of the most prominent pathological changes of DMD[5]. In DMD, fibrosis could inhibit muscle regeneration, reduce muscle strength and flexibility, and ultimately contribute to mortality[6]. Among the key mediators of fibrosis, connective tissue growth factor (CTGF/CCN2) plays a central role[7]. CTGF protein levels are significantly elevated in the skeletal muscle of DMD patients and correlate positively with the extent of muscle fibrosis[8]. Reducing CTGF levels in a DMD animal model (mdx mice) could significantly reduce muscle fibrosis level and enhance muscle strength[9]. Therefore, CTGF is considered a potential therapeutic target for DMD treatment.

CTGF has four functional domains (IGFBP-domain, VWC-domain, TSP1-domain, and CT-domain)[7]. Both the VWC-domain and CT-domain contribute to the fibrosis activity of CTGF, indicating their potential as therapeutic target domains for DMD[10]. However, the VWC-domain specific antibody (FG-3019/Pamrevlumab, FibroGen) despite showing promise in preclinical models of DMD, failed to meet efficacy endpoints in Phase III clinical trials (https://investor.fibrogen.com/news-releases/news-release-details/fibrogen-announces-topline-results-lelantos-2-phase-3-clinical, https://www.mdaconference.org/abstract-library/pamrevlumab-failed-to-meet-primary-and-secondary-endpoints-for-nonambulatory-patients-with-duchenne-muscular-dystrophy-dmd-in-lelantos-1/)[11]. This clinical outcome has prompted a critical re-evaluation of the VWC-domain as the optimal therapeutic target domain within CTGF[12]. In this study, our data demonstrated that FG-3019 treatment elevated compensatory TGF-β1 level, which could attenuate the fibrosis inhibitory activity of FG-3019 both *in vitro* and *in vivo*. Moreover, our data suggested that CT-domain contributed more fibrosis activity of CTGF, without elevating compensatory TGF-β1 level, than VWC-domain. These findings suggested that the CT-domain is a more optimal therapeutic target of CTGF in DMD fibrosis treatment.

Nucleic acid aptamers (single-stranded DNA/RNA oligonucleotides typically 20−80 nt in length) are selected through the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) process[13,14]. As promising therapeutic agents, aptamers have several key advantages: 1) facile chemical modification through solid-phase synthesis allowing precise incorporation of functional groups; 2) high thermal stability with shelf life exceeding 12 months at −20 °C; 3) excellent batch-to-batch reproducibility due to chemical synthesis process; and 4) minimal immunogenicity as demonstrated in clinical trials[15-18]. Specifically, aptamers can rapidly recognize and bind to flexible protein regions through their unique 3D folding structures, achieving nanomolar binding affinity[19]. Their small size (8−15 kDa, about 1/10 of IgG antibodies) enables efficient tissue penetration, particularly in muscle tissues where with low vascular permeability[20]. Therefore, it is desirable to develop a CT-domain specific aptamer for DMD treatment.

The SELEX process leverages binding stringency and PCR amplification to drive affinity maturation, selecting aptamers with progressively stronger target binding[21]. Nevertheless, conventional SELEX involves time-consuming and inefficient multiple rounds of screening[13,22,23]. AI-based screening technologies have enabled more efficient, computer-driven methods for drug discovery. Here, we developed AptGEN, a generative model using a hybrid VAE (variational autoencoder)−GAN (generative adversarial network) architecture, capturing the evolutionary patterns and structural motifs to speed the CTGF aptamer drug discovery. By using AptGEN, we rapidly screened a CTGF CT-domain specific aptamer with high affinity, shortening the screening rounds from 20 rounds to 5 rounds, and reducing the screening time from 100 to 42 days. Also, this aptamer could exert fibrosis inhibitory activity without elevating compensatory TGF-β1 level *in vitro*.

Building upon this breakthrough, we further engineered this CTGF CT-domain specific aptamer into an optimized therapeutic candidate through rational modification strategies. The truncated aptamer (Apc003) containing the essential "CCCTC" motif maintained high binding affinity to the CT-domain of CTGF. To enhance its stability, we introduced 2′-*O*-methyl modifications to all cytosine and guanine nucleotides, which significantly improved resistance to serum nucleases and liver metabolism. To extend its elimination half-life *in vivo*, we used our well-developed a long-lasting fatty acid modification *via* octadecadienoic acid (OA) to develop the long-lasting the CTGF aptamer (Apc003OA). OA could facilitate extending the half-life of the conjugated Apc003 to 7 days in normal rats. In mdx mice, Apc003OA not only demonstrated prolonged retention in muscle tissues compared to antibody therapies (FG-3019), but also exerted better fibrosis inhibitory activity without elevating compensatory TGF-β1 levels than FG-3019. Treatment with Apc003OA significantly improved muscle specific force and forelimb grip strength while down-regulating fibrosis markers (fibronectin and collagen III) in diaphragm muscle. Masson's staining confirmed reduced fibrosis index in both gastrocnemius and diaphragm muscles. When treatment duration was extended to 5 months, unlike FG-3019, the Apc003OA treatment group still showed no elevation in TGF-β1 levels. Apc003OA treatment showed no obvious organ toxicity in mdx mice. These findings suggested that the CT-domain is a more optimal therapeutic target of CTGF in DMD fibrosis treatment.

Within just 10 months, we successfully moved from target domain discovery to aptamer drug discovery, and subsequently obtained both Orphan Drug Designation (DRU-2024-10138) and Pediatric Rare Disease Designation (RPD-2024-838) from the US Food and Drug Adminstration. This study presents an AI-driven tool for the development of aptamer drugs targeting specific protein domains, offering a promising therapeutic candidate drug for DMD patients and paving the way for innovative treatments in precision medicine.

2. Material and methods

2.1. Single-stranded DNA library construction

We constructed a library of single-strand DNAs with a whole length of 76 nt, the ssDNA library used in SELEX contained a 40-nucleotide (nt) randomized region and an 18 nt conserved region at each end for primer annealing: 5′-CGTACGGTCGACGC-TAGC-(N)40-CACGTGGAGCTCGGATCC-3′. FP 5′-CGTACGG TCGACGCTAGC-3′ and biotinylated RP 5′-biotin-GGATCC-GAGCTCCACGTG-3′ were used for the amplification of ssDNA during selection. All oligos were synthesized and purified by HPLC. Binding and washing buffers were prepared with phosphate buffered saline (PBS) pH 7.4 with 1 mmol/L $MgCl_2$, 5 mmol/L imidazole and 0.05% Tween 20.

2.2. Coupling of the CT-domain of CTGF (CTGF-CT) protein to magnetic beads

An aliquot (400 μL) of Magnabind™ Carboxyl Derivatized beads suspension (10 mg/mL) was washed several times with 1 mL coupling buffer (25 mmol/L MES, pH 5). Then, 25 mg EDC and 25 mg NHS were dissolved in 600 μL coupling buffer and immediately added to the 400 μL of washed beads. The mixture was rotated for 1 h at 25 °C. After the reaction, the beads were washed three times with 1 mL coupling buffer and aliquoted into two 1.5 mL centrifuge tubes. In one tube, CTGF-CT protein (10 μg) was added and the final reaction volume of 200 μL coupling buffer to prepare the CTGF-CT conjugated beads (CTGF-CT beads) for SELEX. After 0.5 h of end-over-end rotation at 25 °C, the unreacted protein in the supernatant was recovered and the beads were washed twice with 200 μL coupling buffer. Then, ethanolamine (1 mol/L) was incubated with the CTGF-CT beads for 10 min to block the activated and unreacted carboxyl groups on the beads. In addition, the recovered protein in the supernatant was quantified by an enzyme-linked immunosorbent assay (ELISA) kit to confirm that CTGF-CT was successfully immobilized on the beads. In the other tube, ethanolamine (1 mol/L) was also added to prepare negative beads for counter-SELEX. Finally, CTGF-CT and negative beads were washed extensively several times with binding buffer (5 mmol/L $MgCl_2$ and 0.01% Tween-20 in Dulbecco's PBS) and stored in binding buffer at 4 °C until use.

2.3. Protein SELEX procedure

5 rounds of the selection and amplification of ssDNA aptamers were performed to obtain specific aptamers for the CT-domain of CTGF[24]. An initial ssDNA library (200 nmol) containing a 40-nucleotide randomized region flanked by constant primer-binding sequences was used as the starting pool. The library was thermally conditioned to promote proper folding prior to each round of selection by heating at 95 °C for 10 min, cooling on ice at 4 °C for 10 min, and equilibrating at room temperature (RT, 25 °C) for 10 min. Recombinant human CTGF CT-domain protein was immobilized on magnetic beads at 4 °C for 1 h under gentle rotation. The conditioned ssDNA library was incubated with the CTGF-coated beads at RT (25 °C) for 30−60 min to allow binding. The total incubation volume was 1 mL, and BSA (10 μg for the first round, 1 μg for subsequent rounds) was included in the reaction mixture to minimize nonspecific adsorption.

Following incubation, the beads were separated magnetically, and the unbound sequences were removed by washing with selection buffer containing 50 mmol/L EDTA, 1 mol/L NaOH, 5 mol/L NaCl, and 500 mmol/L Tris−HCl (pH 7.4). The washing stringency was gradually increased across successive rounds by increasing the washing frequency and buffer strength, while the amount of immobilized target protein was progressively decreased to enhance selection pressure and enrichment for high-affinity aptamers.

To improve specificity, negative selection was performed using magnetic beads coated with recombinant human CT-deficient CTGF protein. The ssDNA pool was first incubated with these control beads, and unbound sequences were collected and subsequently used for positive selection against the target CTGF CT-domain in the next SELEX round.

The bound ssDNA−protein−bead complexes were directly used as templates for PCR amplification using a forward primer (FP) and a biotin-labeled reverse primer (Bio-RP).

PCR conditions were as follows: a) Initial denaturation at 95 °C for 60 s; b)12−20 cycles of 95 °C for 15−30 s, 56−60 °C for 15−30 s, and 72 °C for 10−30 s; c) Final extension at 72 °C for 1−2 min. PCR products were verified by agarose gel electrophoresis and purified using the QIAquick PCR Purification Kit.

Biotinylated PCR products were incubated with Streptavidin Magnetic Beads (100 μL) for 15 min at RT with gentle rotation to immobilize the double-stranded DNA (dsDNA).

After magnetic separation and washing with PBS to remove unbound material, the immobilized dsDNA was denatured with 50

μL of 0.1 mol/L NaOH for 5 min at RT to release the non-biotinylated ssDNA strand. The eluted ssDNA was neutralized with 50 mmol/L Tris−HCl buffer (pH 7.4) and further purified using the QIAquick PCR Purification Kit.

After 5 rounds of selection, the ssDNA pools from each round—including the initial random library—were subjected to next-generation sequencing (NGS) to identify sequences enriched for specific binding to the CTGF CT-domain[25]. The detailed screening conditions for each round, including target amount, elution volume, and washing frequency, are summarized in Supporting Information Table S1 (Selection Stringency in Each Round).

2.4. Statistical analysis of next-generation sequencing data

The forward and reverse sequencing data from each round were pairwise checked to ensure that the bidirectional data matched. When the forward sequence was consistent with the corresponding reverse sequence after complementary base pairing, and if every base of the sequence met the minimum quality requirement, the forward sequence was then extracted from the raw data as a valid sequence. The total error rate occurring in each round and the frequency at which the forward sequence failed to pair with the reverse sequence were counted. The reads per million (RPM) of each sequence in each round was calculated to determine the enrichment level of the aptamer as Eq. (1). This was done by dividing the number of occurrences of each sequence in each round by the total number of sequences in each round, and then multiplying by 1 million.

$$\mathrm{RPM} = \frac{\text{Number of reads mapped to an aptamer}}{\text{Total number of reads in the corresponding round}} \times 10^6 \tag{1}$$

The top 20,000 high-frequency sequences were extracted for further analysis, and their corresponding frequency changes in each round were depicted.

2.5. Computational workflow for aptamer discovery

The discovery of high-affinity aptamers targeting the CT-domain of CTGF was achieved through a multi-stage computational pipeline. This process began with the acquisition and rigorous quality assessment of next-generation sequencing (NGS) data from the initial SELEX pools. This high-quality data then served as the input for our bespoke generative AI model, AptGEN, which was constructed using a hybrid architecture. Following model training, a systematic post-generation screening workflow was employed to analyze the model's output, identify clusters of high-potential sequences, and rank candidates based on predicted affinity. Finally, a series of computational validation and mechanistic investigation steps were performed to refine the lead candidates and elucidate their binding interactions. The following sections detail each stage of this comprehensive AI-driven approach.

2.6. Data processing and quality assessment

To ensure the integrity of the input data for our model, raw NGS data from the SELEX pools underwent a rigorous quality control process.

Sequencing error assessment: Raw sequencing reads were evaluated to quantify sequencing errors and unmatched rates between forward and reverse primers. Error and unmatched rates were consistently below 5%, confirming the high fidelity of the sequencing data for downstream modeling. Base Distribution and Sequence Count Analysis: The nucleotide distribution was analyzed for each SELEX cycle to confirm an even representation, while the sequence count distribution was monitored to track the enrichment of specific aptamer families throughout the selection process.

2.7. Model input data

The AptGEN model was trained using two primary sources of data: SELEX-Seq Data: This serves as the foundational input, comprising sequence pools from the initial rounds of an iterative *in vitro* selection process. This dataset captured the evolutionary trajectory of aptamers as they become progressively enriched for higher affinity towards the CT-domain of CTGF. Biochemical Features: To provide a richer, physicochemical context, the raw sequences were augmented with predicted biochemical properties. This included features such as predicted secondary structures, nucleotide composition, and the presence of potential binding motifs, which helped the model learn the structure−function relationships that govern high-affinity binding.

2.8. AptGEN model architecture and hyperparameters

AptGEN's architecture synergistically combined a Variational Autoencoder (VAE) and a Generative Adversarial Network (GAN) to generate novel and optimized aptamer sequences. All input ssDNA sequences were one-hot encoded into a tensor of shape (N, 4, L), where N is the batch size, 4 represents the nucleotide channels (A, C, G, T), and L is the sequence length.

Variational Autoencoder (VAE): The VAE component functions as the generative backbone. Encoder Network: The encoder consisted of a stack of three 1D convolutional layers designed to extract hierarchical features. The layers were configured as follows: (1) 128 filters, kernel size 5, stride 2, padding 2; (2) 256 filters, kernel size 5, stride 2, padding 2; (3) 512 filters, kernel size 5, stride 2, padding 2. Each convolutional layer was followed by a batch normalization layer and a LeakyReLU activation function (negative slope of 0.2) to prevent dying ReLU issues and promote stable training. The output of the final convolutional layer was flattened and passed to two separate fully connected layers to produce the parameters of the latent distribution: a mean vector (μ) and a log-variance vector ($\log\sigma 2$), each of dimension 64.

Decoder Network: The decoder mirrored the encoder's architecture using transposed convolutional layers to reconstruct the original sequence from a latent vector z, which was sampled using the reparameterization trick ($z = \mu + \sigma \odot \epsilon$). The network began with a fully connected layer to project the 64-dimensional latent vector back to the appropriate size for the convolutional stack. This was followed by three 1D transposed convolutional layers with configurations mirroring the encoder's convolutional layers in reverse, each coupled with batch normalization and LeakyReLU activation. The final layer used a Softmax activation function across the channel dimension to output a probability distribution over the four possible nucleotides for each position in the sequence.

Generative adversarial network (GAN): The GAN component operated as a refinement engine, improving the biological relevance and affinity of the generated sequences. Generator (G): The generator's architecture was designed to up-sample a noise vector into a full-length aptamer sequence. It began with a fully

connected layer that projected the input vector (a 64-dimensional latent vector z sampled from the VAE's latent space) into a high-dimensional tensor suitable for convolutional operations. This was followed by a stack of three 1D transposed convolutional layers with the following configurations: (1) 256 filters, kernel size 5, stride 2, padding 2; (2) 128 filters, kernel size 5, stride 2, padding 2; (3) 4 filters, kernel size 5, stride 2, padding 2. Similar to the VAE decoder, each layer except the last was followed by batch normalization and a LeakyReLU activation. The final layer employed a Softmax activation to produce the nucleotide probability distribution for the generated sequence.

Discriminator (D): The discriminator was a convolutional neural network designed to classify sequences as real or generated. Its architecture mirrored the VAE's encoder, processing an input sequence through a stack of three 1D convolutional layers with progressively increasing filter counts (128, 256, and 512 filters, respectively; all with kernel size 5, stride 2, padding 2). Each convolutional layer was followed by batch normalization and LeakyReLU activation. The output from the final convolutional layer was flattened and passed to a single fully connected layer which outputted a single scalar logit. A Sigmoid activation function was applied to this logit to produce the final probability, representing the likelihood that the input sequence was from the real experimental dataset.

Conditional GAN (cGAN) implementation: To guide the generation process, a conditional GAN approach was employed. An auxiliary condition, representing a predicted affinity class was concatenated to the latent vector z before being fed into the generator. For the discriminator, this conditional vector was concatenated to the flattened feature map from the final convolutional layer before being processed by the fully connected output layer. This allowed the model to learn a mapping between specific conditions and sequence characteristics, enabling the generation of aptamers with desired properties.

2.9. Training process

The AptGEN model was trained using data from the initial SELEX rounds, significantly accelerating the discovery timeline compared to traditional procedures. Loss Functions: The model's training was governed by distinct loss functions for the VAE and GAN components. The VAE training optimized the evidence lower bound (ELBO), which involved minimizing a combined loss function of reconstruction loss and the Kullback-Leibler (KL) divergence. The reconstruction term ensured the decoder can accurately rebuild the input data, while the KL divergence acts as a regularizer, forcing the latent space to follow a prior distribution (typically a standard normal distribution). The VAE loss was defined as Eq. (2):

$$\mathscr{L}_{\mathrm{VAE}}(\theta,\phi;x) = E_{q\phi(z|x)}\left[\log_{p\theta}(x|z)\right] - iii_{\mathrm{KL}}\left(q_{\phi}(z|x)\middle\|p(z)\right) \tag{2}$$

The GAN training utilizes an adversarial loss based on a minimax game. The discriminator (D) tries to maximize its ability to correctly classify real and fake samples, while the generator (G) tries to minimize the probability of its generated samples being classified as fake. The objective function is:

$$\min_{G}\max_{D} V(\mathrm{D},\mathrm{G}) = E_{x\sim P_{\mathrm{data}}(x)}[\log D(x)] + E_{z\sim P_z(z)}\left[\log\left(1-D(G(z))\right)\right. \tag{3}$$

Optimization and hyperparameter tuning: The Adam optimizer was used for training both networks with parameters set to $\beta 1 = 0.9$ and $\beta 2 = 0.999$. A batch size of 128 was used. A learning rate warm-up strategy was implemented where the rate was linearly increased from 1×10^{-5} to a peak of 4×10^{-4} over the first 2000 training steps. Following the warm-up, a cosine annealing schedule decayed the learning rate over the remaining 50,000 training epochs. To ensure stable training, gradient clipping (with a maximum gradient norm of 1.0) and batch normalization were employed in all convolutional layers. A dropout rate of 0.3 was applied to the fully connected layers to prevent overfitting.

2.10. Model output

The primary outputs of the trained AptGEN model were: Generated Aptamer Sequences: A set of *de novo* designed aptamer sequences predicted to have high binding affinity for the CT-domain of CTGF. These sequences were the product of the optimized interplay between the VAE's generative capacity and the GAN's refinement process. Affinity Prediction: Each generated sequence was accompanied by a predicted binding affinity score. This probabilistic score was derived from the model's learned correlation between sequence features and binding strength, enabling effective ranking and prioritization of candidates for experimental validation.

2.11. Post-generation screening workflow

A systematic *in silico* workflow was used to screen the generated sequences and identify the most promising candidates.

High-affinity probability distribution: The reconstructed hidden layer of the VAE was analyzed to calculate the likelihood of each generated sequence being a high-affinity binder. A probabilistic model, considering structural features and similarity to known binders, was used to predict the binding affinity of each sequence.

Cluster Estimation: To explore the diversity of high-affinity solutions, Gaussian Mixture Models (GMM) were applied to the latent space representation to identify distinct clusters of sequences. A GMM models the data as a weighted sum of several Gaussian distributions, with the probability density function given by Eq. (4):

$$p(x) = \sum_{k=1}^{K} \pi_k \mathscr{N}(x|\mu k, \Sigma k) \tag{4}$$

where K is the number of clusters, πk are the mixing coefficients, and N ($x|\mu k$, Σk) is the Gaussian density for each cluster k. Clusters with the highest probability of containing high-affinity sequences were selected for further validation. A hierarchical clustering approach was also used to organize sequences and enable efficient selection of promising candidates.

2.12. Validation and mechanistic investigation

A comprehensive pipeline of computational and experimental methods was used to validate predictions and investigate binding mechanisms.

Multiple sequence alignment (MSA): MSA was used to analyze high-affinity sequences and predict key nucleotide sites and conserved motifs critical for binding. These insights were used to guide the refinement of generated sequences to enhance binding affinity and specificity.

Truncation and Mutation Analysis: The functional importance of the predicted key sites was validated by evaluating the binding affinity of aptamers after site-directed mutagenesis or truncation. This iterative refinement process was supported by *in silico* mutagenesis and molecular dynamics simulations to predict the effects of mutations on affinity and stability.

RosettaFold2NA predictions: The RosettaFold2NA algorithm was employed to predict the three-dimensional structure of the aptamer−protein complex, providing detailed insights into the binding interactions between the optimized aptamer sequences and the CT-domain of CTGF. These structural predictions guided the design of blocking peptides to experimentally probe and confirm the binding mechanism.

2.13. Computational dynamics modeling of the interaction between CTGF and TGF-β1

To investigate the interaction between CTGF and TGF-β1, the protein sequences (CTGF: P29279, TGF-β1: P01137) were obtained from the UniProt database. The complex structure was predicted using AlphaFold3, and residue contacts ($\leq$4.5 Å) were analyzed with UCSF ChimeraX 1.8.

Molecular dynamics (MD) simulations were performed using GROMACS 2023.5 with the AMBER99SB-ILDN force field and the OPC water model. Following energy minimization, the system was equilibrated with position restraints (1000 kJ/mol·nm^2), followed by a 10 ns NVT simulation and two NPT simulations (10 ns with a Berendsen barostat and 40 ns with a Parrinello-Rahman barostat). A 50 ns production run was then conducted for analysis.

Trajectory analysis was performed using MDA nalysis to calculate the root mean square deviation (RMSD) and root mean square fluctuation (RMSF).

RMSD measures the average distance between the atoms of the simulated structure and a reference structure over time, indicating conformational stability. It is calculated as Eq. (5):

$$\mathrm{RMSD}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|r_i(t) - r_i^{\mathrm{ref}}\|^2} \tag{5}$$

RMSF measures the fluctuation of each individual residue from its time-averaged position, indicating regions of high flexibility. It is calculated as Eq. (6):

$$\mathrm{RMSF}_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} \|r_i(t) - \langle r_i \rangle\|^2} \tag{6}$$

2.14. Bio-layer interferometry (BLI) binding assay

The binding affinities between the tested aptamers and the CT-domain of CTGF were measured using the BLI-based Octet Platform. Biotin-tagged aptamers, purified to a concentration of 100 nmol/L, were immobilized on biosensors, resulting in a saturation response of 5−6 nm after 300 s. Subsequently, the loaded biosensors were washed for 3 min in PBS containing 0.05% Tween-20 (PBST) buffer to remove loosely nonspecifically bound aptamers and establish a stable baseline. For binding kinetic measurements, the association of aptamers and FL-CTGF (concentration gradient: 800, 400, 200, 100, 50, 25, and 12.5 nmol/L) was monitored for 60−180 s, and their dissociation was measured for 120 s in assay buffer. Reference wells using buffer instead of FL-CTGF were included to correct baseline shifts. A parallel set of sensors incubated in buffer only served as negative reference controls to correct for the non-specific binding of FL-CTGF to the biosensor surface. Raw kinetic data were analyzed using a double-reference subtraction approach, subtracting both background and non-specific binding. The binding affinity constant (K_d) values were calculated using a 1:1 binding model through global fitting of multiple kinetic traces. Data Analysis 9.0 software was employed for real-time monitoring data analysis. All measurements were conducted in three independent experiments.

2.15. Examination of the immigration rate of fibroblasts by wound healing assay

Rat 2 cells were cultured to form a monolayer on six-well plates. Then, Rat 2 cells were serum-deprived overnight and then mechanically "wounded" by scraping with a Fisher brand ready-tip (1−200 μL). Cell monolayers were washed twice with PBS and cultured with different aptamers in Dulbecco's Modified Eagle Medium (DMEM). After 0, 24 and 48 h of incubation, fibroblasts were washed with PBS and stained by Diff-Quick solution. Pictures were taken at × 2.5 magnification. Migration rate was calculated by counting the cells that cross into the "scratch" (wounded) area and presented as a percentage of total cells on identical "non-scratch" area (100%). The mobility ratio was evaluated by migrated cell area/scraped area by Image J.

2.16. Enzyme-linked oligo nucleotide assay (ELONA)

160 ng CT-domain of CTGF and negative targets were coated per well in 96-well microtiter plate by hydrophobic interactions at 4 °C overnight. Non-binding sites in the well were blocked with blocking buffer (PBS, %1 BSA and %1 Tween 20) at 25 °C for 1 h followed by washing with washing buffer (PBS, 1 mmol/L $MgCl_2$, 0.1% Tween 20 and 0.1% BSA) for 5 min for 4 times. Appropriate concentrations of biotinylated aptamer candidates (100 μL) were added into each well and incubated for 45 min at 25 °C with continuous gentle shaking. After binding, the plate was washed with washing buffer for 5 min for 4 times to remove non-specific and weak binding. 100 μL streptavidin-Horseradish peroxidase (HRP) (1:10,000 dilute into PBST+0.1% BSA) was added to each well and incubated for 30 min and washed with binding buffer for 4 times. 50 μL 3,3′,5,5′-tetramethylbenzidine (TMB) which is the substrate of HRP was added to each well and incubated for 20 min. The reaction was stopped by adding 50 μL 2 mol/L H_2SO_4. Absorbance at 450 nm was measured with microplate reader (Molecular Device i3x). Data was analyzed with Origin software (OriginLab, Northampton, MA, USA). A nonlinear curve fitting model Hyperbl was used to plot the binding curve.

2.17. Western blot analysis of fibrosis markers

The expression levels of proteins were analyzed by Western blot. Briefly, Rat 2 cells were seeded at a density of 2×10^5 cells/well and cultured with DMEM containing 10% FBS. After seeding, cells were starved using 2% fetal bovine serum (FBS) DMEM for 12 h. After starvation, the culture medium was changed to DMEM treated with human full-length CTGF protein and aptamers for 48 h. Subsequently, both the culture medium and cell lysates were collected for analysis. To harvest and lyse the cells, they were first detached using a detachment buffer (250 mmol/L sucrose, 10

mmol/L Tris, pH 7.6, 1 mmol/L EDTA, 1 mmol/L EGTA, 1 × protease inhibitor cocktail [Roche], and 1 mmol/L phenylmethylsulfonyl fluoride [PMSF]). The detached cells were then lysed in RIPA buffer (P0013B, Beyotime, Shanghai, China) supplemented with 2% PMSF (P7626, Sigma) and a protease inhibitor cocktail, followed by incubation on ice for 30 min. The cell lysates were centrifuged at 13,000×*g* for 12 min, and the culture medium was centrifuged at 5000×*g* for 15 min. The resulting supernatants from both the cell lysates and the medium were solubilized in 2% SDS sample buffer for subsequent analysis. The protein concentrations of samples were evaluated by BCA assay (P0010, Beyotime, Shanghai, China). Protein samples were separated through 10% SDS-PAGE and transferred to PVDF membranes. The membranes were blocked with 5% skim milk for 1 h, then incubated with primary antibodies against GAPDH (Santa cruz, sc-32233), fibronectin (Santa cruz, sc-8422) and collagen IIIa1 (Thermo Fisher, 6Y9C6) at 4 °C overnight, washed with TBST and incubated with corresponding anti-mouse or anti-rabbit secondary antibodies for 2 h at 25 °C (1:5000 dilution). The immunocomplexes were visualized with using a Immobilon® Western (P90719, Millipore). The protein expression levels were analyzed with Image J software (NIH, Bethesda, MD, USA) and normalized against GAPDH.

2.18. Construction of the CTGF recombinant plasmids

The full-length human CTGF coding sequence (NCBI GenBank Accession No. NM_001901, corresponding to nucleotide positions 201 to 1250) served as the template. Specific CTGF variants were designed: full-length CTGF (FL-CTGF), CTGF with VWC-domain deficiency (CTGF-ΔVWC), and CTGF with CT-domain deficiency (CTGF-ΔCT). The precise nucleotide sequences encoding these constructs, including the defined deletion boundaries for ΔVWC and ΔCT-domains, were synthesized *de novo* by a commercial gene synthesis service. These synthesized DNA fragments were subsequently cloned into the mammalian expression vector pcDNA3.1 using standard restriction enzyme digestion and ligation techniques. The resulting recombinant plasmids were transformed into competent *E. coli* DH5α cells for amplification. Positive clones were identified by colony PCR using vector-specific primers flanking the insertion site. Plasmids isolated from PCR-positive clones were subjected to Sanger sequencing across the entire insert and flanking regions to verify the integrity and accuracy of the CTGF sequences, including the intended deletions for ΔVWC and ΔCT constructs. Sequence-verified plasmids were purified using an endotoxin-free plasmid maxiprep kit, quantified spectrophotometrically, and stored at −20 °C for subsequent transfection experiments.

2.19. Examination of the levels of fibrosis markers and TGF-β1 in CTGF knock out fibroblast expressing CTGF with VWC-domain or CT-domain deficiency

Fibroblasts (Rat 2 cells) with stable knockout of endogenous CTGF (*CTGF^{ko}*) were generated using CRISPR/Cas9 technology and maintained in DMEM supplemented with 10% FBS and 1% penicillin−streptomycin (100 U/mL penicillin and 100 μg/mL streptomycin) at 37 °C in a humidified atmosphere containing 5% CO_2. For transient expression studies, *CTGF^{ko}* fibroblasts were seeded in 6-well plates at a density of 2×10^5 cells per well and cultured until they reached 70%−80% confluence. Transfection was performed using Lipofectamine 3000 reagent according to the manufacturer's protocol. Briefly, for each well, 2.5 μg of the respective plasmid DNA (FL-CTGF, CTGF-ΔVWC, CTGF-ΔCT, or empty pcDNA3.1 vector as control) was diluted in 125 μL of Opti-MEM I Reduced Serum Medium. Separately, 5 μL of Lipofectamine 3000 reagent was diluted in 125 μL of Opti-MEM I Medium. After 5 min of incubation at 25 °C, the diluted DNA was combined with the diluted Lipofectamine 3000 reagent, mixed gently, and incubated for a further 15 min at 25 °C to allow complex formation. The DNA−lipid complex mixture (250 μL total) was then added drop-wise to the cells in each well containing 1.5 mL of fresh complete medium (without antibiotics). Cells were incubated with the complexes for 48 h at 37 °C and 5% CO_2. Following the incubation period, the culture supernatant was carefully collected, centrifuged at 300×*g* for 5 min to remove any cellular debris, and aliquoted. Both the clarified supernatant and the cell lysates were stored at −80 °C until analysis. The levels of fibrosis markers (fibronectin, collagen III) in the supernatant and cell lysates, as well as TGF-β1 specifically in the supernatant, were subsequently analyzed by Western blotting and/or ELISA.

2.20. Examination of the levels of fibrosis markers in fibroblast treated with FG-3019 and TGF-β1 antibody

The FG-3019 antibody (Synonyms: Pamrevlumab, Anti-Human CTGF Recombinant Antibody; CAS No.: 946415-13-0) and TGF-β1 antibody (YA5515, Cat. No.: HY-P85823) were purchased from the MedChemExpress (MCE, CN). Fibroblasts were cultured in DMEM supplemented with 10% FBS and 1% penicillin−streptomycin at 37 °C with 5% CO_2. The fibroblasts were seeded in 6-well plates to reach 90% confluence. Then, the fibroblasts were cultured in DMEM and treated with CTGF (100 μg/mL), FG-3019 (100 μg/mL), TGF-β1 antibody (100 μg/mL), or IgG (100 μg/mL) for 48 h, respectively. Both supernatant and cell lysis were collected for evaluating fibronectin and collagen III levels by Western blot analysis.

2.21. Aptamer methoxy modification

Methoxy modifications were introduced at all cytosine (C) and guanine (G) positions within the Apc003 aptamer sequence. Modifications were systematically performed in a bidirectional manner—progressing from both 5′ and 3′ termini toward the central region—using a five-nucleotide unit as the modification window. Within each 5-nt segment, eligible C and G underwent 2′-O-methylation. The stability of modified aptamers was subsequently assessed in 10% and 100% FBS under physiological conditions (37 °C, 24 h), respectively.

2.22. Preparation of the amino-modified Apc003 (Apc003-NH_2)

The Apc003-NH_2 (sequence: NH_2-C6-TGmCmCmTACmTGmCmTCmTCmCmCmT

CmGmGmATCmCmGmAGmCmTCmCmACm GmTGm-(5′→3′)-idT) was synthesized on a 32 μmol scale on a ÄKTA oligopilot plus 100 standard DNA/RNA synthesizer using commercially available 3′-IDT CPG, 5′-*O*-DMT-2′-deoxynucleoside (A_{Bz}, C_{Ac}, Gi_{Bu} and T) phosphoramidite monomers, 5′-*O*-DMT-2′-*O*-methyl nucleoside ($A_{Bz,}$ C_{Ac}, Gi_{Bu} and U) phosphoramidite monomers, and amino-modifier C6 CE phosphoramidite monomer. All oligonucleotides were synthesized in DMT-OFF mode. After completion of the synthesis, the solid support was suspended in 28% ammonium hydroxide solution and heated at 60 °C for 4 h

to release the product from the support and to complete the removal of all protecting groups. The solid support was filtered, and the filtrate was desalted/buffer exchanged into ddH_2O and lyophilized to dryness.

2.23. Preparation of the Apc003OA/DA/PA/PE conjugate

In a flame-dried 100-mL flask, $NaHCO_3$ (150 mg) was dissolved in 4 mL ddH_2O by stirring at 25 °C. Then the solution (400 mg, 80 equiv) of octadecanedioic acid (OA), dodecanedioic acid (DA), palmitic acid (PA) and PEG40K (PE, Polyethylene Glycol) in 5 mL DMF was added to the mixture, respectively.100 mg Apc001-NH_2 in 6 mL ddH_2O was subsequently added to the mixture. The reaction was completed *via* the reaction of the 5′-amine and the *N*-hydroxysucccinimide-derivative of the coupling agents. The resulting mixture was stirred for overnight. After reaction was completed, the mixture was concentrated *in vacuo* to obtain the crude Apc003OA/DA/PA/PE conjugate.

2.24. Purification of Apc003-NH_2, and the Apc003OA/DA/PA/PE conjugate

The crude Apc003-NH_2 or Apc003OA/DA/PA/PE conjugate was dissolved in 20 mL dd H_2O, and then the mixture was filtered with 0.45-micron PTFE membrane to obtain clear filtrate. The crude product was purified by preparative liquid chromatography Autotide 100 Oligo Purification System to obtain the pure Apc003-NH_2 or the Apc003OA/DA/PA/PE conjugate (condition: column: Autotide-RP-20 C18, 250 mm × 22 mm, 1 mm; 26 °C; a flow rate of 6 mL/min; a linear gradient of 5%−35% *v/v* solvent B in 40 min, A was 0.05 mol/L triethylamine acetate solution, B was HPLC grade acetonitrile).

2.25. Mass spectrometric analysis

Both mass of aptamers, and aptamer conjugates were determined by an ultra-high performance liquid chromatography coupled with quadrupole time of flight mass spectrometer (UHPLC-ESI-Q-TOF-MS) in negative ion mode on a single quadrupole liquid chromatograph mass spectrometer (Xevo G2-XS QTof, Waters). The mass of purified Apc003-NH_2 was confirmed by ESI-Q-TOF-MS. The measuring weight (12066.0615) matched the calculating weight (12064.0015) (Supporting Information Fig. S1). The mass of purified Apc003OA conjugate was confirmed by ESI-Q-TOF-MS. The measuring weight (12782.5382) matched the calculating weight (12782.8912) (Fig. S1). The measuring weight (12636.2666) matched the calculating weight (12636.27394) for Apc003DA. The measuring weight (12770.0791) matched the calculating weight (12770.0851) for Apc003PA.

2.26. HPLC analysis of Apc003PE

The characterization of Apc003PE was assessed by RP-HPLC (Waters Acquity HPLC system) using a C18 column (Waters/BEH-C18, 2.1 mm × 100 mm, 1.7 um). Apc003PE was run at a flow rate of 0.2 mL/min with a linear gradient of solvent B over 15 min (A: 94.9% *v/v* H_2O, 5% *v/v* MeCN and 0.1% *v/v* TFA; B: 94.9% *v/v* MeCN, 5% *v/v* H_2O and 0.1% *v/v* TFA).

2.27. The design of blocking peptide

Based on the calculated binding sites of the CT-domain of CTGF to Apc003, it was predicted that K276 on CTGF were involved. Then, we designed a peptide sequence (BP) including the above binding sites and verified its binding affinity to Apc003 by Bio-layer Interferometry (BLI) analysis. All the peptide sequences were purchased from GL Biochem (Shanghai) Ltd.

2.28. Pharmacokinetics experiment

1) The pharmacokinetics study of the Apc003OA/DA/PA/PE was performed in six-week-old male SD rats. The rats were fed ad libitum with a standard laboratory diet and housed under controlled conditions (12 h light cycle, 20 °C). The rats were treated with Apc003OA/DA/PA/PE at 25 mg/kg by a single subcutaneous injection. Blood samples were collected at different time points (0, 1, 2, 4, 8, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120, 144, 168, 192, 216, 240, 264, 312, 360, 408, 456 and 504 h) from rats ($n = 3$ for each group) and plasma were isolated; 2) After subcutaneous injection of Apc003OA/DA/PA/PE, 500 μL of blood were taken *via* orbital vein from each rat at different time points, and were collected into tubes containing sodium-heparin as an anticoagulant (0.9 mL vacutainers, BD Biosciences), then were put into wet ice immediately; 3) Plasma were isolated by centrifugation at 3000×*g* for 10 min at 4 °C within 1 h after collection and stored at −80 °C until analysis. After allowing each sample to thaw, the plasma was mixed and an aliquot (100 μL) transferred to a polypropylene tube. Methanol (1.4 mL) were added to each sample and, after mixing, the plasma protein was removed by centrifugation at 14,000×*g* for 10 min. An aliquot of the supernatant was transferred to a 250 μL limited volume insert *via*. Then the supernatants were dried and reconstituted for Aptamer quantification. Standards were prepared in blank mice plasma containing sodium-heparin; 4) The remaining compounds in the plasma were quantified by Molecular beacon (MB). The pharmacokinetics profile of Apc003OA/DA/PA/PE were analyzed by software DAS 2.0.

2.29. Aptamer conjugate quantification

The Apc003OA/DA/PA/PE could be assayed by molecular beacon (MB) technology[26]. The MB sequence was Cy3-CCGCGCGmAGmAGmGmGmAGmCmCmTAGmGmCmTGCGCGG-BHQ2. The sequence of Apc003 was TGmCmCmTACmTGmCmTCmTCmCmCmTCmGmGmATCmCmGmAGmCmTCmCmACmGmTGm-idT (the underlined sequence was the complementary sequence between MB and Apc003). The 96-well plate was used for loading samples. For each well, 5 μL plasma samples, 5 μmol/L MB sequence and 85 μL PBS were added. Then, mixed the reagents by gently tapping the sides of the plate. Incubated the plate, protected from light for 1 h at 37 °C. The Apc003 could be detected when hybridization occurred between the complementary sequence of MB and Apc003. This caused the separation of the stem and hence of the fluorophore and the quencher. Once the fluorophore was no longer next to the quencher, illumination of the hybrid with light resulted in the fluorescent emission. The presence of the emission reported that the event of hybridization has occurred and hence Apc003 sequence was present in the test sample. Read the plate using a fluorescence plate reader at Excitation: 550 nm and Emission: 575 nm. The concentrations of Apc003OA/DA/PA/PE were calculated according to the standard

curves, respectively. Standard curves were obtained by analyzing different concentrations of Apc003OA/DA/PA/PE that were dissolved in plasma and processed as well as analyzed using the same procedure.

2.30. Aptamer stability assay

To evaluate the stability of the acquired products, aptamers were incubated with 10 μL of FBS at 37 °C. In contrast, an equivalent aptamer was also added to 10 μL of PBS as a control. After 0, 2, 4, 8, 12, 24, 36, 48, and 72 h of incubation, the samples were collected and incubated at 95 °C for 10 min to inhibit nuclease activity. Finally, the samples were run on a 3% agarose gel (Sigma−Aldrich, Saint Louis, USA) in 1 × TAE buffer and visualized with GelRed (Sigma−Aldrich). The lanes were quantified and analyzed using Image Lab software.

2.31. Metabolic stability test

The 150 μL incubation system contained aptamer (10 μg/mL), liver microsome (1 mg/mL), and NADPH regeneration system (final concentrations at 1.3 mmol/L for $NADP^+$, 3.3 mmol/L for glucose-6-phosphate, 3.3 mmol/L for magnesium chloride and 0.4 U/mL for glucose-6-phosphate dehydrogenase) in 0.1 mol/L potassium phosphate buffer (pH 7.4). Samples after incubation for 0, 2, 18 h was snap frozen in liquid nitrogen to terminate the reaction, and stored in < -65 °C, followed by addition of DNA loading buffer and formamide (deionized). Next, the resulting samples were electrophoresed on 20% denaturing urea polyacrylamide gel at 220 V for 40 min. After that, the gels were photographed by usage of ChemiDoc XRS System (BioRad) and analyzed in Image Lab software.

2.32. The pharmacokinetics analysis

The Apc003OA/DA/PA/PE concentration versus time profile was plotted and analyzed for each rat by software DAS 2.0 (BioGuider Co., Shanghai, China). All reported concentrations of the Apc003OA/DA/PA/PE were based on the area under the curve (AUC), respectively. Using the data from standards, calibration curve was generated for the Apc003OA/DA/PA/PE. Half-life ($t_{1/2}$) of the Apc003OA/DA/PA/PE was calculated according to the time it takes for Apc003OA/DA/PA/PE to eliminate half of the maximum plasma concentration, respectively. Maximum plasma concentration (C_{max}) and the time to maximum plasma concentration (T_{max}) were obtained according to pharmacokinetics curves[27]. The AUC was computed starting at the time the drug was administered and ending when the concentration in plasma was negligible. Subsequently, the dosing intervals in multiple-dose of the Apc003OA/DA/PA/PE were calculated according to Eq. (7):

$$D = 1/1-e^{-Ke\cdot t} \tag{7}$$

K_e: elimination constant, $K_e = \ln 2/t_{1/2}$; t: dosing interval; D: dosing ratio = loading dose/maintenance dose).

2.33. Tissue distribution by IVIS imaging system

Cy3-Apc003OA and Cy3-FG-3019 at a single dose of 5 mg/kg were given to two groups of the mdx mice, respectively, by subcutaneous route (s.c.). The mice were killed 2, 4, 12 and 24 h after injection. Muscle (diaphragm and gastrocnemius) and major organs (hearts, lungs, livers, spleens, and kidneys) were collected for biophotonic imaging. Biophotonic imaging was performed using an IVIS Lumina XR imaging system. All images were acquired with an exposure time of 5 s. The distribution of fluorescence within tissues was represented as a pseudo-color image. The photographic and fluorescent images were individually acquired and then overlaid[25].

2.34. Blood sampling and plasma isolation

The blood samples were taken *via* orbital vein from each rat at different time points and collected into tubes containing sodium heparin as an anticoagulant (0.9 mL vacutainers, BD Biosciences) on wet ice. The plasma was isolated by centrifugation at $6000\times g$ for 10 min at 4 °C after collection and stored at −80 °C until analysis. After allowing each sample to thaw, the plasma was be mixed, and an aliquot was transferred to a polypropylene tube. Methanol was added to each sample, after mixing, the plasma protein was removed by centrifugation at $14{,}000\times g$ for 10 min.

2.35. Tissue sampling and collection

To ensure sample integrity and prevent contamination or degradation, all tissue collection procedures were performed meticulously within an ultraclean biosafety cabinet. Immediately following euthanasia and prior to tissue dissection, major organs (heart, liver, spleen, lungs, kidneys) and relevant muscles (gastrocnemius, diaphragm, soleus) were rapidly excised. Tissues were rinsed briefly in ice-cold PBS to remove excess blood. For histopathological analysis (H&E, Masson's staining, immunohistochemistry), tissue samples designated for fixation were immediately immersed in an adequate volume (at least 10x tissue volume) of freshly prepared 4% paraformaldehyde (PFA) in PBS or 10% neutral buffered formalin. For biochemical analysis (Western blot, ELISA, RNA), tissue samples were flash-frozen in liquid nitrogen within 2−5 min of excision to halt enzymatic degradation. Between the collection of each distinct tissue type, all surgical instruments (forceps, scissors) were thoroughly sterilized by immersion in 70% ethanol followed by flaming or using fresh sterile instruments to prevent cross-contamination. All fixed tissues were processed for embedding within 48 h. Flash-frozen tissues were stored at −80 °C in clearly labeled, sterile cryovials until further processing.

2.36. Examination of fibrosis markers within tissue samples by immunohistochemical staining

Lung tissues fixed in 4% PFA for 48 h were processed for cryosection. Tissues were cryoprotected by sequential immersion in increasing concentrations of sucrose in PBS (10%, 20%, 30%, each for 24 h at 4 °C) and then embedded in Optimal Cutting Temperature (OCT) compound. Serial sections of 5 μm thickness were cut using a CryoStar NX50 cryostat (Thermo Fisher Scientific, USA) at −20 °C and mounted onto Superfrost Plus glass slides. Sections were air-dried briefly and then fixed in 4% PFA for 10 min at 25 °C. After washing with PBS, sections were permeabilized with 0.1% Triton X-100 in PBS for 10 min (if required by the antibody protocol) and blocked with 10% normal goat serum in PBS containing 1% bovine serum albumin (BSA) for 1 h at 25 °C to reduce nonspecific binding. Sections were incubated overnight at 4 °C with primary antibodies diluted in blocking solution. Primary antibodies used included those against TGF-β1, α-smooth muscle actin (α-SMA), fibronectin, collagen III, and CTGF (specific antibody

sources and catalog numbers should be provided here or referenced from section 2.15). Following extensive washing with PBST, sections were incubated with appropriate Alexa Fluor 568-conjugated secondary antibodies (*e.g.*, goat anti-rabbit or goat anti-mouse IgG, diluted 1:500 in blocking solution) for 1 h at 25 °C in the dark. Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI, 1 μg/mL) for 5 min. After final washes in PBS, sections were mounted using a fluorescence-compatible anti-fade mounting medium. Stained sections were visualized and imaged using an Olympus BX43 fluorescence microscope equipped with appropriate filter sets. Images were captured using a digital camera under consistent exposure settings.

2.37. Enzyme-linked immunosorbent assay for investigating levels of fibrosis markers in the serum

Blood samples collected from mice were allowed to clot at 25 °C for 30 min and then centrifuged at 1000×*g* for 20 min at 4 °C to separate serum. The supernatant (serum) was carefully aspirated, aliquoted into fresh tubes, and stored at −80 °C until analysis. Prior to assay, serum samples were thawed on ice and centrifuged again at 10,000×*g* for 10 min at 4 °C to remove any precipitated material or residual particulates. The levels of specific fibrosis markers (*e.g.*, TGF-β1) and other relevant diagnostic markers (*e.g.*, surfactant protein A/D, Krebs von den Lungen-6, circulating caspase-cleaved cytokeratin-18) in the clarified serum were quantified using commercially available ELISA kits according to the manufacturer's instructions (*e.*g., kits from Thermo Fisher Scientific, USA, or R&D Systems). Briefly, standards and diluted serum samples were added to antibody-coated wells and incubated for the specified time. After washing steps to remove unbound material, a biotinylated detection antibody was added, followed by incubation and washing. Streptavidin-Horseradish Peroxidase (HRP) conjugate was then added. After further incubation and washing, TMB substrate solution was added for color development. The enzymatic reaction was stopped by adding a stop solution, and the absorbance was immediately measured at the appropriate wavelength (*e.g.*, 450 nm with a reference wavelength of 570 nm) using a microplate reader. Concentrations of the analytes in the samples were interpolated from the standard curve generated using the provided standards.

2.38. In vivo assessment of anti-fibrotic activity of Apc003OA in mdx mice

Twelve-week-old male mdx mice (C57BL/10ScSn-Dmdmdx/J), serving as a model for DMD, were utilized to evaluate the therapeutic efficacy of Apc003OA against fibrosis. Mice were randomly assigned to one of four experimental groups: (1) Vehicle Control Group: Received subcutaneous (s.c.) injections of the formulation buffer alone (*e.g.,* PBS or vehicle used for Apc003OA), once weekly for 12 weeks. (2) Apc003OA Treatment Group: Received s.c. injections of Apc003OA at a dose of 100 mg/kg body weight, dissolved in the appropriate vehicle, once weekly for 12 weeks. (3) FG-3019 Treatment Group: Received s.c. injections of the CTGF VWC-domain specific antibody FG-3019 (purchased from Med Chem Express) administered according to an equivalent schedule and dose regimen as a comparator. (4) Apc003OA + BP Pre-treatment Group: Received an s.c. injection of the blocking peptide (BP) designed against the CT-domain of CTGF binding site for Apc003 (*e.g.*, at a dose and timepoint defined in pilot studies, such as 1 h prior to each Apc003OA dose), followed by s.c. injection of Apc003OA (100 mg/kg), once weekly for 12 weeks. All injections were administered in a consistent volume relative to body weight (*e.g.*, 10 μL/g body weight). Body weights were monitored weekly. At the end of the 12-week treatment period, mice were subjected to functional assessments and then humanely euthanized. Tissue samples (diaphragm, gastrocnemius, soleus, heart, liver, spleen, lung, kidney, serum) were collected for subsequent histopathological analysis and biochemical analysis.

2.39. Muscle grip strength test

After 12 weeks of treatment, mice were placed over the grip rod until they naturally grasped it while handling them by their tails. The tail was gently pulled horizontally until the mouse released the rod grip to investigate the force by a BIO-GS4 Grip Strength Test instrument (BIOSEB Ltd., USA). As a mouse grasped the bar, the peak pull force in grams was recorded on a digital force transducer. In the conventional test, a mouse was allowed to grasp the bar mounted on the force gauge. The gauge was reset to 0 g after stabilization, and the mouse's tail was slowly pulled back by an inspector. Tension was recorded by the gauge at the time the mouse released its forepaws from the bar. For the modification of the conventional test, the gauge was rotated vertically and fixed to the metal stand to keep the system immobilized. The measurement procedure was identical to that in the conventional test except for the direction in which the mouse's tail was pulled by an inspector. In each test, trials in which only one forepaw, or the hindlimbs were used and in which the mouse turned during the pull or leaves the bar without resistance were excluded. Given that the speed of the tail pull can influence the measurement, we conducted the procedure at a constant speed sufficiently slow to permit mice to build up a resistance against it. We performed 6 consecutive measurements per day at 1 min intervals. The grip strength was normalized to body weight and expressed in N/g[28].

2.40. Muscle specific force test

The mice were euthanatized to collect the soleus muscle followed by mounting the muscles in the mechanics bath of a muscle testing system (1200A: Isolated Muscle System, Aurora Scientific Ltd., Aurora, ON, Canada). After placing the muscle in the bath, a single 0.5 ms stimulation pulse was used to produce a twitch for monitoring the force output. The current was gradually increased until the force reached a maximal but stable level. Using isometric twitch stimulations, adjust the length of the muscle gradually until a maximal force is obtained. The muscles were stimulated at 150 Hz to collect force data. The muscle specific force was calculated by using Eqs. (8) and (9):

$$\text{Specific force (mN/mm}^2) = \text{maximum isometric tetanic force (mN) / PCSA (mm}^2) \quad (8)$$

$$\text{The PCSA (mm}^2) = \text{mass (mg) / [}(L_o \text{ mm}) \times (L/L_o) \times (1.06 \text{ mg/mm}^3)\text{]} \quad (9)$$

The L/L_o is the fiber to muscle length ratio.

2.41. Histopathological test

Samples of the heart, liver, spleen, lungs, kidneys, and stomach (along with muscles like gastrocnemius and diaphragm)

designated for histopathology were immediately immersed in an adequate volume of 10% neutral buffered formalin for fixation for a minimum of 48 h at 25 °C. Fixed tissues were then processed through a graded series of ethanol solutions (70%, 80%, 95%, and 100%) for dehydration, cleared in xylene, and infiltrated and embedded in paraffin wax using standard histological protocols. Paraffin blocks were sectioned at a thickness of 5 μm using a rotary microtome. Sections were mounted onto glass slides, deparaffinized in xylene, and rehydrated through a descending series of ethanol (100%, 95%, and 70%) to water. For general histology assessment, sections were stained with hematoxylin (nuclear stain) for 5−8 min, rinsed in water, differentiated in 1% acid alcohol, blued in Scott's tap water substitute, rinsed again, and then counterstained with eosin (cytoplasmic stain) for 1−2 min. Sections were then dehydrated through ascending alcohols, cleared in xylene, and coverslipped using a permanent mounting medium. To evaluate the degree of fibrosis, Masson's trichrome staining was performed specifically on muscle sections (*e.g.*, gastrocnemius, diaphragm). Briefly, deparaffinized and rehydrated sections were stained in Weigert's iron hematoxylin working solution for 10 min, rinsed in water, and then stained in Biebrich scarlet-acid fuchsin solution for 10−15 min. After rinsing, sections were differentiated in phosphomolybdic-phosphotungstic acid solution for 10−15 min, transferred directly to aniline blue solution for 5−10 min, rinsed briefly in distilled water, differentiated in 1% acetic acid for 2−5 min, dehydrated, cleared, and coverslipped. Collagen fibers stain blue, muscle fibers stain red, and nuclei stain black. All stained sections (H&E and Masson's) were examined under light microscopy (*e.g.*, Olympus BX43) by a pathologist/researcher blinded to the treatment groups. Images were captured using a digital camera attached to the microscope.

2.42. Statistical analysis

All data are presented as mean ± standard deviation (SD). Normality of data distribution was assessed using the Shapiro-Wilk test. For comparisons between multiple groups, one-way analysis of variance (ANOVA) was performed. If the ANOVA indicated significant differences ($P < 0.05$), Tukey's honestly significant difference *post-hoc* test was applied to determine specific pairwise differences between group means. Statistical significance was defined as $P < 0.05$. All statistical analyses weres performed using GraphPad Prism software (version 8; GraphPad Software, Inc., San Diego, CA, USA). The required sample size for *in vivo* experiments was predetermined using a power calculation based on effect sizes observed in our previously published protocols investigating similar endpoints. Animals were randomly assigned to treatment groups, and researchers performing measurements and analyses were blinded to the group allocation throughout the experimental procedures and data analysis. Any animals exhibiting poor health or body condition unrelated to the experimental treatment prior to the endpoint were excluded from the final analysis.

3. Results

3.1. CT-domain is an optimal domain without elevating compensatory TGF-β1 level rather than VWC-domain

To compare the differences in the contribution to the fibrosis activity of CTGF between VWC-domain and CT-domain, fibroblast with CTGF knock out ($CTGF^{ko}$) was constructed. CTGF with VWC-domain deficiency (CTGF-ΔVWC), CTGF with CT-domain deficiency (CTGF-ΔCT) and full-length CTGF (FL-CTGF) was expressed in $CTGF^{ko}$ fibroblast, respectively. Compared to the FL-CTGF group, the levels of the supernatant fibrosis markers (fibronectin and collagen III) were down-regulated by 50%−80% in CTGF-ΔCT group, whereas the levels of the supernatant fibrosis markers were down-regulated by 20%−30% in CTGF-ΔVWC group (Fig. 1a). It indicated that CT-domain contributed more to the fibrosis activity of CTGF than VWC-domain.

TGF-β1 is a well-known profibrotic factor contributing to the development and progression of fibrosis[29,30]. It has been known that CTGF VWC-domain could physically interact with TGF-β1[31,32]. Cooperative interactions between CTGF and TGF-β1 signaling are required to elicit overt tissue fibrosis[32]. In our study, the supernatant compensatory TGF-β1 level was not elevated in CTGF-ΔCT group, whereas the supernatant compensatory TGF-β1 level was significantly elevated in CTGF-ΔVWC group, when compared to that in FL-CTGF group (Fig. 1b). Consistently, FG-3019 treatment elevated the compensatory TGF-β1 level in supernatant of fibroblasts in a concentration-dependent manner *in vitro* (Fig. 1c). To investigate whether the elevated TGF-β1 could attenuate the fibrosis inhibitory activity of FG-3019, fibroblast was treated with CTGF, CTGF + FG-3019, CTGF + TGF-β1 antibody, CTGF + FG-3019 + TGF-β1 antibody, and CTGF + IgG, respectively. The levels of fibrosis markers (fibronectin and collagen III) in CTGF + FG-3019 + TGF-β1 antibody group were lower than those in CTGF + FG-3019 group *in vitro* (Fig. 1d). Meanwhile, the levels of fibrosis markers (fibronectin and collagen III) in mdx + FG-3019 + TGF-β1 antibody group were lower than those in mdx + FG-3019 group *in vivo* (Fig. 1e), suggesting that the elevated compensatory TGF-β1 level could attenuate the fibrosis inhibitory activity of FG-3019 both *in vitro* and *in vivo*. From the above, it indicated that CT-domain is an optimal domain without elevating compensatory TGF-β1 level rather than VWC-domain.

3.2. An aptamer-based generative model (AptGEN) was developed to obtain high-affinity aptamers against the CT-domain of CTGF

Based on our *in vitro* and *in vivo* data, we demonstrated that CTGF CT-domain is an optimal therapeutic target domain within the CTGF. Therefore, it is desirable to develop an innovative CTGF inhibitor that could specifically bind to CTGF CT-domain and exert muscular fibrosis inhibitory activity.

Aptamers are synthetic oligonucleotides identified through the SELEX process, which can specifically bind to flexible protein domains through their unique 3D conformations. SELEX technology is a powerful and flexible method for screening of high-affinity aptamers[33]. However, the SELEX process is still time consuming, and the successful rates are low[13]. To reduce screening time and overcome the inefficiencies of traditional SELEX, we developed AptGEN, a hybrid generative AI model for rapid *de novo* aptamer design against the CTGF CT-domain.

Firstly, 5 positive rounds of positive SELEX have been performed, and data of next-generation sequencing (NGS) was obtained. Next, to ensure the fidelity of the data underpinning our model, the quality of the NGS output was thoroughly assessed. We analyzed the raw data from the initial library and the five subsequent enrichment pools to extract valid sequences, calculating both sequencing error rates within conserved primer regions and the rates of primer mismatch. Across all datasets, both error and

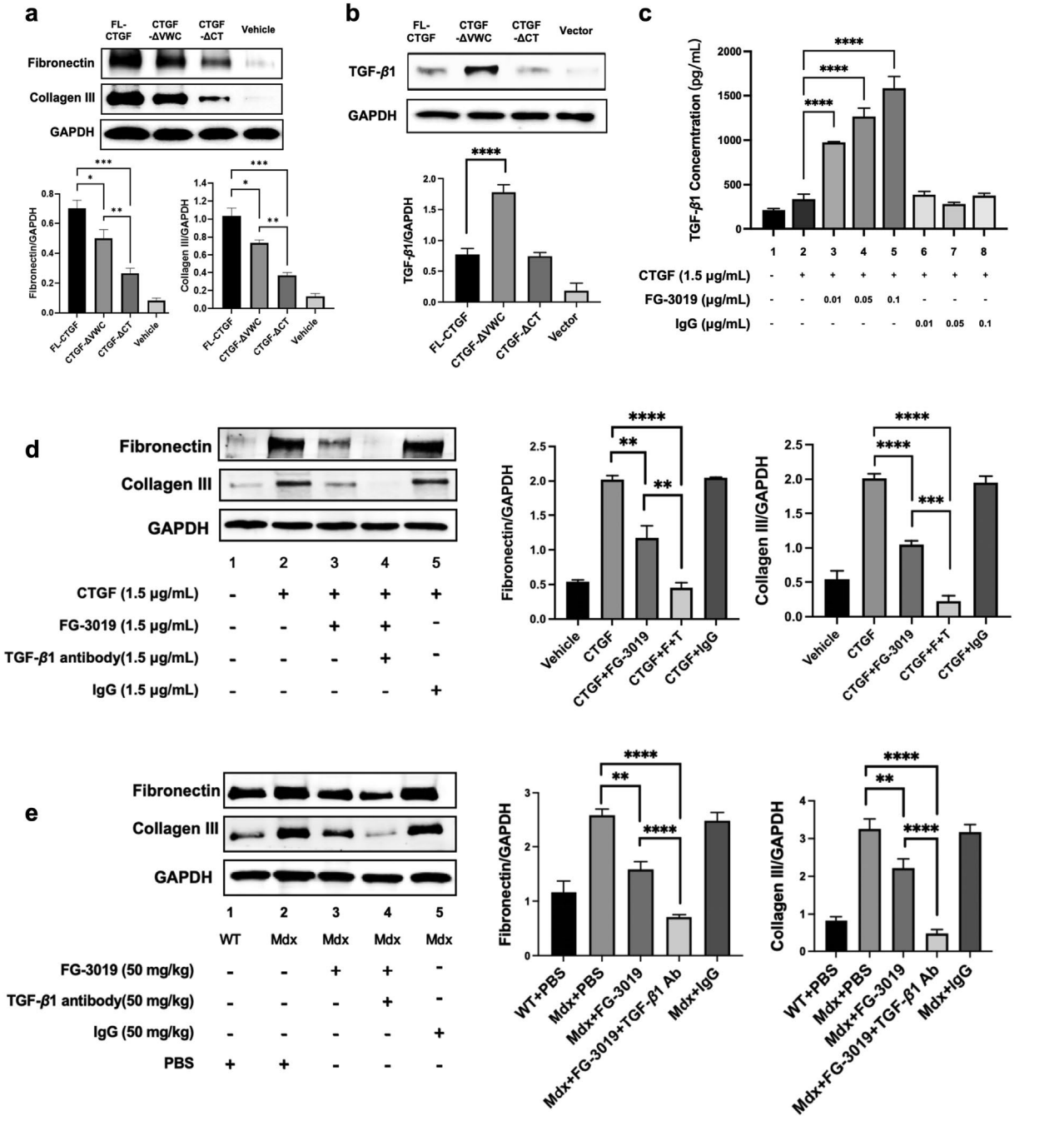


Figure 1 CT-domain is an optimal domain without elevating compensatory TGF-β1 level rather than VWC-domain. (a) The levels of supernatant fibrosis markers (fibronectin and collagen III) in *CTGF*^{ko} (CTGF knock out) fibroblast expressing full-length CTGF (FL-CTGF), CTGF with VWC-domain deficiency (CTGF-ΔVWC), and CTGF with CT-domain deficiency (CTGF-ΔCT), respectively (upper). Statistical analysis of ratios of fibronectin and collagen III to GAPDH (lower). (b) The levels of supernatant TGF-β1 in *CTGF*^{ko} fibroblast expressing FL-CTGF, CTGF-ΔVWC and CTGF-ΔCT, respectively (upper). Statistical analysis of ratios of TGF-β1 to GAPDH (lower). (c) The ELISA analysis of TGF-β1 levels in fibroblast treated with CTGF (1.5 μg/mL), CTGF + FG-3019 (0.01 μg/mL), CTGF + IgG (0.05 μg/mL), CTGF + FG-3019 (0.1 μg/mL), CTGF + IgG (0.01 μg/mL), CTGF + FG-3019 (0.05 μg/mL), CTGF + IgG (0.1 μg/mL) for 48 h, respectively. (d) The Western blot analysis of fibronectin and collagen III levels in fibroblast treated with CTGF, CTGF + FG-3019, CTGF + TGF-β1 antibody (CTGF + TGF-β1 ab), CTGF + FG-3019 + TGF-β1 antibody (CTGF + F + T), and CTGF + IgG for 48 h, respectively (left). Statistical analysis of ratios of fibronectin and collagen III levels to GAPDH, respectively (right). (e) The Western blot analysis of fibronectin and collagen III levels in diagram muscle of WT mice treated with PBS, Mdx mice treated with PBS, FG-3019, TGF-β1 antibody (TGF-β1 ab), FG-3019 + TGF-β1 antibody (CTGF + F + T), and IgG for 6 weeks, respectively (left). Statistical analysis of ratios of fibronectin and collagen III levels to GAPDH, respectively (right). Data are expressed as mean ± SD followed by one-way ANOVA with Tukey's *post hoc* test, $n = 3$ per group. *$P < 0.05$. **$P < 0.01$. ***$P < 0.001$, ****$P < 0.0001$.

unmatched rates were found to be consistently below 5%, confirming the high quality of the data and its suitability for subsequent model training and analysis (Supporting Information Fig. S2a). Meanwhile, we evaluated key metrics of the SELEX process itself. Uniform nucleotide distribution across SELEX rounds indicated negligible synthesis bias, while progressive enrichment of specific sequence families demonstrated effective selection of high-affinity binders (Fig. S2b and S2c).

Using this high-quality dataset, we constructed AptGEN by integrating a Variational Autoencoder (VAE) with a Generative Adversarial Network (GAN). This model was specifically designed to learn a rich, structured representation of the aptamer sequence landscape. The VAE component served as the generative backbone. It included an encoder that compressed input ssDNA (single-stranded DNA) sequences into a low-dimensional space and a decoder that reconstructed them. This forced the model to capture the most salient sequence and structural features linked to key biological function. The GAN acted as a refinement module. Its generator created new ssDNA sequences based on the VAE's latent space. And a discriminator was trained to distinguish these generated sequences from real sequences in the SELEX data. We further improved the model using deconvolution techniques and added conditional constraints. These adjustments guided the model to generate sequences with features that suggest high binding affinity (Fig. 2a).

AptGEN was trained exclusively on data from the first five SELEX rounds. This endpoint was strategically selected based on selection dynamics: early rounds (1−4) exhibit excessive noise from limited binder enrichment, impairing feature learning, while later rounds (>6) diminish the acceleration advantage inherent to AI-driven discovery. The fifth round represented an optimal inflection point—sufficient selection pressure had revealed binding patterns, yet diversity remained high enough for robust model learning and extrapolation. At this stage, sufficient selection pressure had been applied to allow meaningful binding patterns to emerge from the background noise, yet enough sequence diversity remained to provide the model with a rich feature set for learning and accurate extrapolation. Leveraging a previously accumulated model training process, we utilized the processed data from only the first five SELEX rounds as input to train the AptGEN model. This enabled AptGEN to learn the complex sequence-affinity landscape and generate a diverse library of potential high-affinity aptamers within just 7 days of computational work—a profound acceleration that reduced the conventional 20-round discovery timeline (approximately 3 months) to merely 5 rounds and 42 days total. This dramatic efficiency improvement demonstrates AptGEN's power as a rapid therapeutic aptamer discovery platform, capable of identifying viable candidates from minimal initial experimental data.

3.3. One CTGF CT-domain specific aptamer (CT-3) with the highest fibrosis inhibitory activity was obtained

With a vast library of thousands of potential high-affinity sequences generated by the AptGEN platform, the critical next challenge was to identify the single most promising therapeutic lead. To this end, we designed and executed a rigorous, multi-stage discovery funnel that systematically integrated high-throughput computational screening with decisive experimental validation, ensuring a logical and evidence-based progression from a large candidate pool to a single, optimized lead aptamer (Fig. 2b).

The first stage of this funnel consisted of a large-scale *in silico* pre-screening. We first analyzed the high-affinity probability distribution within the reconstructed hidden layer of the VAE, which serves as a learned "fitness landscape" for the aptamers. This enabled the initial identification of broad regions within the sequence space that were enriched with potentially strong binders (Fig. 2c). To navigate these promising regions with greater precision, we then employed Gaussian Mixture Models (GMM), an unsupervised machine learning algorithm, to partition the high-dimensional latent space into distinct clusters of sequences sharing similar learned features. This approach moves beyond evaluating individual sequences in isolation and instead allows for the identification of "elite families" of candidates. The cluster exhibiting the highest mean predicted affinity was subsequently prioritized for wet-lab validation (Fig. 2d).

The second stage marked the transition from computational prediction to empirical testing. Top-ranking candidate sequences from the elite GMM cluster, along with the most highly enriched sequences from the experimental SELEX pools, were synthesized. Their binding affinities to the CT-domain of CTGF were then quantitatively measured using Biolayer Interferometry (BLI). This crucial step provided definitive validation that AptGEN's *in silico* predictions correlated strongly with real-world binding kinetics (Fig. 2e). Based on these quantitative affinity data, the top five aptamer candidates exhibiting the highest affinity (designated CT-1 through CT-5) were advanced for functional characterization (Supporting Information Table S2).

In the final step, the biological function of these 5 high-affinity candidates was investigated (Fig. 2f, Table S2), recognizing that strong binding affinity does not necessarily equate to functional potency. The fibrosis inhibitory activities of these 5 high-affinity aptamer candidates (CT-1, CT-2, CT-3, CT-4 and CT-5) were investigated *in vitro* by examining the protein levels of fibrosis markers (fibronectin and collagen III) and the migration rate, respectively. The fibroblasts were treated with CTGF protein and CT-1, CT-2, CT-3, CT-4 and CT-5, respectively. The protein levels of fibronectin and collagen III were examined by WB assay (Fig. 2f). And the migration rate of each group was examined by Wound healing assay (Fig. 2g). Both Western blot and wound healing assays demonstrated that all five high-affinity candidates exhibited anti-fibrotic activity, with CT-3 showing the most potent fibrosis inhibitory effect (Fig. 2f and g). Moreover, the fibrotic inhibitory activity of CT-3 was concentration-dependent. Dot blot assay was employed to define the specificity of CT-3. The data showed that CT-3 could specifically bind to the CT-domain, with no detectable cross-reactivity to other domains of CTGF (Fig. 2h).

Through this systematic process, which progressed from high-throughput computational screening to the experimental validation, CT-3 was identified as the lead aptamer candidate targeting the CTGF CT-domain due to its high affinity, specificity, and potent anti-fibrotic activity.

3.4. One truncated aptamer (Apc003) containing the key nucleotides sites with high affinity against the CT-domain of CTGF was identified through MSA (multiple sequence alignment)

Following the identification of the potent lead aptamer, CT-3, the subsequent objective was to engineer an optimized therapeutic sequence by identifying its minimal functional core. Reducing the size of an aptamer is a critical step in drug development, as shorter oligonucleotides generally offer improved synthesis efficiency, lower production costs, and potentially enhanced tissue

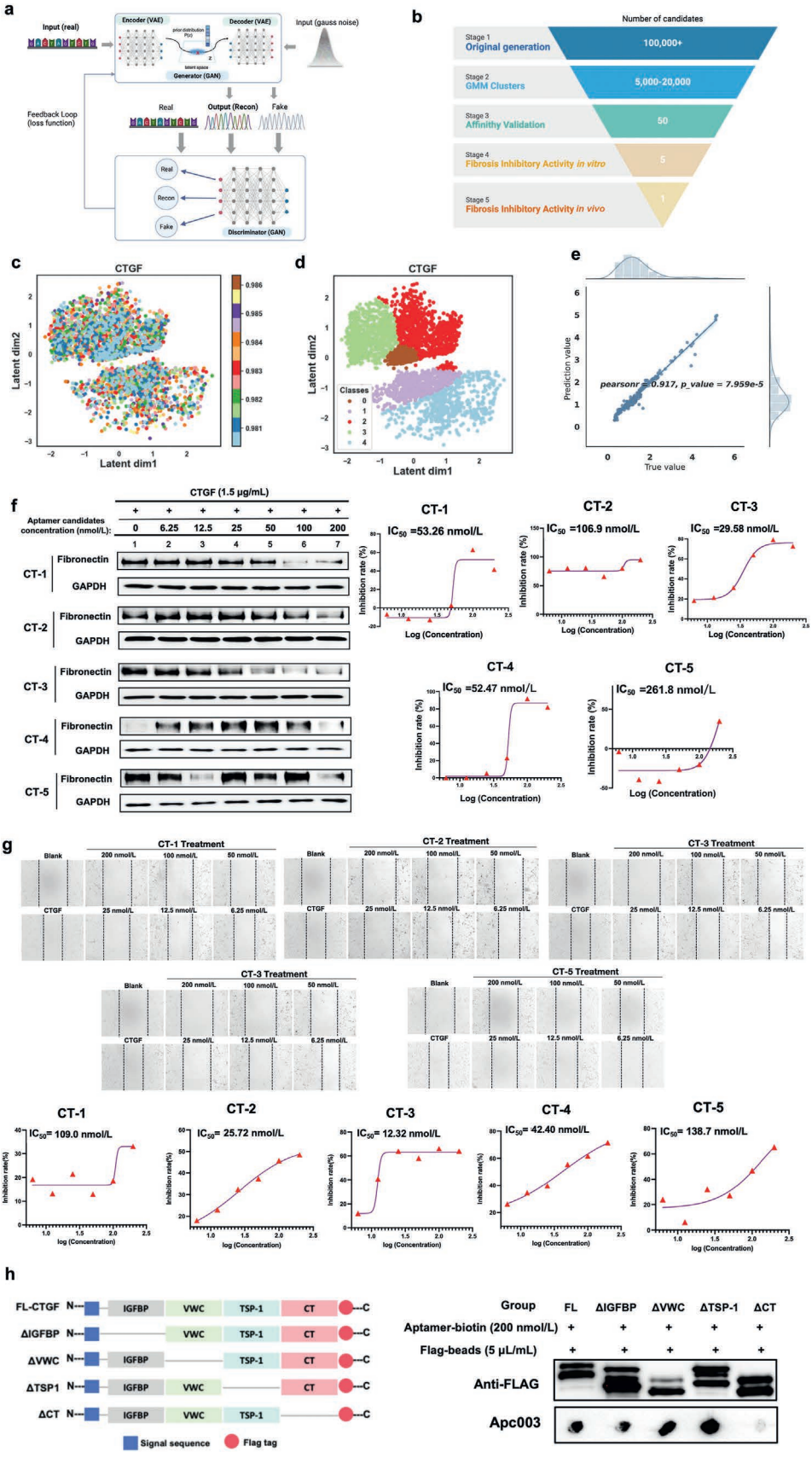


Figure 2 AI-driven generation of high-affinity aptamers targeting the CT-domain of CTGF demonstrated potent fibrosis inhibitory activity *in vitro*. (a) The framework of the neural network trained to generate the high-affinity aptamers targeting the CT-domain of CTGF from the raw NGS data. (b) *In silico* post-generation screening workflow. Selection process: Step 1, (c) The distribution of the reconstructed hidden layer for

penetration and pharmacokinetic properties. To do this efficiently and systematically, we used the AptGEN dataset to guide the truncation process.

Our strategy was predicated on the hypothesis that the functional essence of CT-3 was not unique to its sequence alone but was likely shared among the family of high-affinity aptamers identified by our model. We therefore performed a Multiple Sequence Alignment (MSA) on all sequences belonging to the elite GMM cluster from which CT-3 was originally selected. This comprehensive analysis of functionally related sequences revealed a highly conserved nucleotide motif, "CCCTC" which stood out as the putative binding core responsible for the aptamer's interaction with the CT-domain of CTGF (Fig. 3a).

To investigate the functional importance of the predicted "CCCTC" motif, we performed a series of *in silico* and *in vitro* validation experiments. Using RNAfold, we found that mutating this motif disrupted the aptamer's stable secondary structure, indicated by a less favorable minimum free energy (MFE) compared to the original, suggesting loss of its proper 3D conformation for binding (Fig. 3b). Then, RosettaFold2NA was employed to model the aptamer−protein complex at the atomic level. RosettaFold2NA modeling showed that the "CCCTC" motif directly mediated critical interactions with the CTGF CT-domain, and mutations could abolish these contacts (Fig. 3c). Experimentally, the binding affinities of CT-3 and CT-3 sequence with the "CCCTC" motif truncated were examined by ELONA assay,

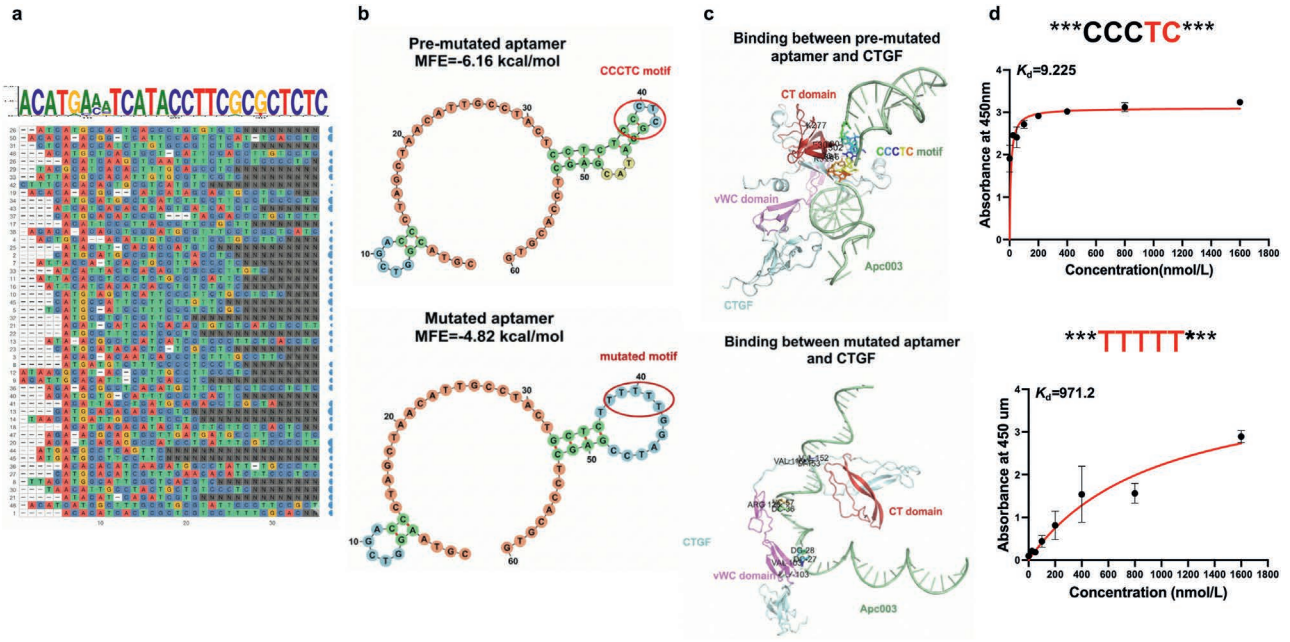


Figure 3 The binding affinity of the Apc003 targeting the CT-domain of CTGF was abolished after truncating the predicted key nucleotides involved in the interaction to CTGF *in vitro*. (a) The distribution of nucleotide sits by MSA (multiple sequence alignment), using the generated sequences. (b) The difference in secondary structural conformation of aptamer candidate (upper), compared to the mutated aptamer (lower) predicted by RNAFold. The stability (minimum free energy, MFE) of the pre-mutated aptamer was −6.16 kcal/mol and that of the post-mutated aptamer was −4.82 kcal/mol. The smaller MFE indicates higher conformational stability. (c) The difference in binding sites of aptamer candidate (upper) to CTGF, compared to the mutated aptamer (lower) predicted by RosettaFold2NA (K277, L301, P302, E304, K314, K315, N316). (d) The binding affinity of aptamer candidate were examined by ELONA Assay (data are expressed as mean ± SD followed by one-way ANOVA with Tukey's *post hoc* test, $n = 3$ per group), after mutating the predicted key nucleotides (lower) compared to the original aptamer candidate (upper).

generated sequences. The plot colors indicated the different probability values obtained from models. Step 2, (d) The clusters of the reconstructed hidden layer estimated by GMM. The plot colors indicated the different clusters. Step 3, (e) The correlation between predicted scores of generated sequence and binding affinities, as validated by Biolayer Interferometry (BLI), Note: Data are expressed as pearson r = 0.917 followed by Pearson Correlation Coefficient. *P* value = 7.959e-5. Step 4, (f) The fibrosis inhibitory activity of the top five highest-affinity aptamer candidates were examined by WB analysis. The fibroblasts (Rat 2 cells) were incubated with serially diluted aptamers CT-1, CT-2, CT-3, CT-4 and CT-5 at final concentrations of 200, 100, 50, 25, 12.5, and 6.25 nmol/L in the medium of CTGF (500 μg/mL) for 48 h. Cell lysis and medium were tested by Western blot assay to investigate the protein levels of fibronectin and GAPDH (left). The intensities of protein bands were quantified by the software Image J. Obtained intensity data was used to calculate IC_{50} value of each aptamer by GraphPad Prism 8. The curve of each aptamer candidate was plotted, respectively (right). (g) The immigration rate of 5 aptamer candidates (CT-1, CT-2, CT-3, CT-4 and CT-5) were determined by wound healing assay (upper). The IC_{50} values of each aptamer was analyzed by GraphPad Prism 8. Nonlinear fit curves were drawn by plotting the inhibitory rates of CTGF aptamer candidates *versus* the logarithm of the serially diluted aptamer candidates' concentrations, respectively (lower). (h) The plasmids of full-length CTGF (FL-CTGF) and its deletion mutants were constructed (FL, ΔIGFBP, ΔVWC, ΔTSP-1 and ΔCT), respectively (left). The binding domain of CT-3 to CTGF was examined by dot blot assay (right). Data are expressed as mean ± SD followed by one-way ANOVA with Tukey's *post hoc* test, $n = 3$ per group. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

respectively. Compared to CT-3, the binding abilities of the mutants with single-base mutations within "CCCTC" motif to CTGF were significantly lower, suggesting that each nucleotide within "CCCTC" motif could be essential for the binding of CT-3 to CTGF. Importantly, compared to CT-3, CT-3 loss of "CCCTC" motif showed dramatically decreased binding affinity (K_d value increased from 9.225 to 971.2 nmol/L), indicating that the core "CCCTC" motif was indispensable for the binding of CT-3 to CTGF (Fig. 3d, Supporting Information Fig. S4, Table S3).

Experimental validation confirmed the "CCCTC" motif as the essential functional core, completing the AI-guided discovery cycle. This integrated computational-experimental approach yielded Apc003—a truncated aptamer retaining this minimal motif—as the optimized preclinical candidate.

3.5. The Apc003 could exert fibrosis inhibitory activity by binding to the CT-domain of CTGF without elevating compensatory TGF-β1 level in vitro

To mechanistically investigate how the Apc003 exerted its fibrosis inhibitory activity, a blocking peptide (BP) was designed based on the predicted binding sites of the CTGF CT-domain to the Apc003[27,34]. Computational docking identified that 4 high-affinity binding residues (K267, F268, K277, and C284) were essential for Apc003-CTGF binding (Fig. 4a). Based on this, the BP (sequence: ISKPIKFELSGCTSMKTYPRAKFCGV) was synthesized to competitively occupy these sites. Additionally, a negative control peptide (NS, sequence: ISKPAAAALSGCTSAAAYPRAKAAAV) was synthesized. Furthermore, BLI assay was employed to define the binding affinity of Apc003 to BP and NC, respectively. The BLI assay results showed that APC003 exhibited good affinity (86 nmol/L) for BP but extremely low affinity for NC (undetectable) (Fig. 4a). Moreover, the binding affinity of the Apc003 to CTGF was significantly reduced when it was pre-saturated with BP (from 86 to 376.8 nmol/L) (Fig. 4b). Then, the levels of fibrosis markers (fibronectin and collagen III) and TGF-β1 in the supernatant of fibroblasts incubated with CTGF were assessed following Apc003 treatment, with or without the addition of BP. The results showed that, compared to the CTGF-only group, the Apc003-treated group exhibited significantly lower levels of both fibronectin and collagen III (Fig. 4c). Moreover, no elevation in compensatory TGF-β1 levels was observed after Apc003 treatment (Fig. 4d). These results demonstrated that Apc003 could inhibit fibrosis activity of CTGF without elevating compensatory TGF-β1 level. Moreover, the BP treatment abolished the fibrosis inhibitory activity of Apc003 on fibronectin and collagen III level. This finding suggested that the fibrosis inhibitory activity of the Apc003 was abolished after blocking the interaction between the Apc003 and the predicted binding sites in the CT-domain of CTGF *in vitro* (Fig. 4c and d). Collectively, these data indicated that the Apc003 exerted fibrosis inhibitory activity by binding to the CTGF CT-domain without elevating compensatory TGF-β1 levels.

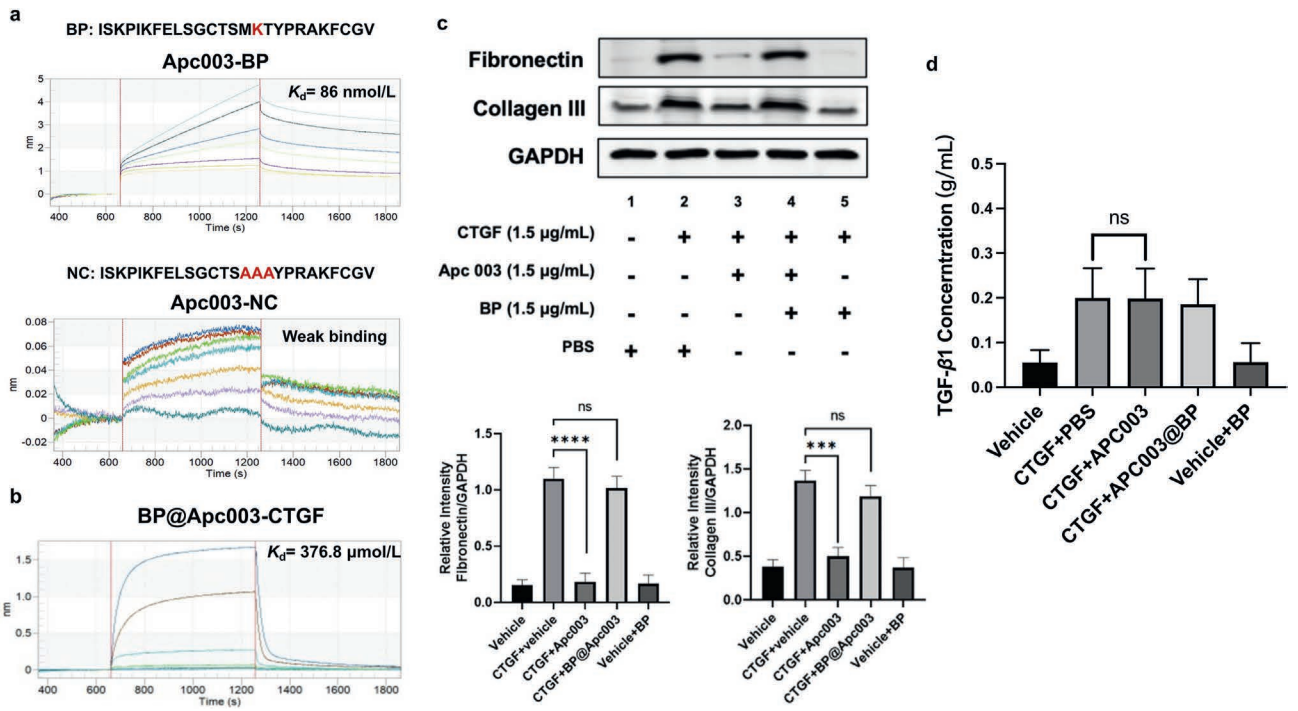
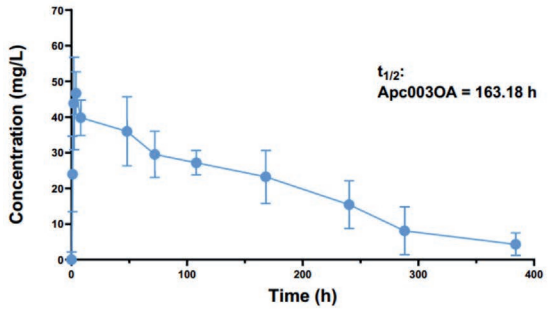


Figure 4 The Apc003 could exert fibrosis inhibitory activity by binding to the CTGF CT-domain without elevating compensatory TGF-β1 level *in vitro*. (a) The mulAi cycle kinetics analysis of binding affinity of Apc003 to blocking peptide (BP) (upper)/NC (lower). (b) The binding affinity of the Apc003 to CTGF was significantly reduced when it was pre-saturated with BP. (c) Western blot analysis of the fibrosis markers (fibronectin and collagen III) levels in CTGF treated fibroblasts after 48 h-treatment of Apc003, BP@Apc003, and BP, respectively (left). Statistical analysis of ratios of fibronectin and collagen III to GAPDH (right). (d) The ELISA assay analysis TGF-β1 levels in CTGF treated fibroblasts after 48 h-treatment of Apc003, BP@Apc003, and BP, respectively. Note: ns: not significant; BP: Blocking peptide; The amino acids in red represent the binding sites; NC: We have designed the negative peptide sequence (NC) by mutating the predicted binding amino acids sites to alanine (A) on BP. Data are expressed as mean ± SD followed by one-way ANOVA with Tukey's *post hoc* test, $n = 3$ per group. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$.

a



b

Parameter	Units	Mean value
T_{max}	h	4
C_{max}	μg/mL	46.67±5.75
AUC	(μg·h)/mL	7702.50±200.12
Elim. $t_{1/2}$	h	163.18±20.14

Figure 5 OA could facilitate prolonging the circulation half-life of the conjugated Apc003 in normal rats. (a) Pharmacokinetics of a single subcutaneous (s.c.) injection of Apc003OA (25 mg/kg) in normal rats. The half-life of the Apc003OA is 163.18 h. (b) Pharmacokinetic parameters of Apc003OA administered s.c. Note: $n = 3$ for each group.

3.6. CT-domain might function independently with TGF-β1 whereas VWC-domain might interact with TGF-β1 revealed by AlphaFold3 and MD simulations

To explore the compensatory elevation in the TGF-β1 level, AlphaFold3 was employed to predict the structural domains involved in the CTGF interaction with TGF-β1. By visualizing the (TGF-β1)-CTGF complex structure (Supporting Information Fig. S4a), we found that the VWC-domain was closest to TGF-β1 in 3D space, indicating a significant interaction. Further analysis of the number of interaction contacts showed that the VWC-domain had 1332 contacts with TGF-β1, compared to only 324 contacts for the CT-domain (Fig. S4b). This suggests that the VWC-domain predominantly interacts with TGF-β1, while the CT-domain is spatially separated and less accessible for direct interaction. Molecular dynamics simulations were conducted to further assess the stability and interactions of the complex. The root meant square deviation (RMSD) values of the TGF-β1-CTGF complex over a 50 ns trajectory confirmed that the complex remained stable throughout the simulation, with the VWC-domain demonstrating higher interaction stability compared to the CT-domain (Fig. S4c). The center of mass analysis (Fig. S4d) between TGF-β1 and the respective CTGF domains indicated that the VWC-domain maintained a closer and more consistent interaction range with TGF-β1 compared to the CT-domain during the simulation. Furthermore, RMSF analysis revealed that the VWC-domain exhibited lower flexibility compared to other CTGF domains, suggesting a more rigid and stable interaction with TGF-β1 (Fig. S4e). These structural and dynamic analyses imply that the spatial arrangement of CTGF allows it to function independently of TGF-β1, thereby avoiding compensatory elevation in TGF-β1 levels.

3.7. Methoxy modification enhanced the in vitro stability of the modified Apc003 in both serum and metabolic conditions

To enhance the stability and nuclease resistance of Apc003, 2′-*O*-methylation was introduced. To evaluate the optimal modification sites, the sequence was divided into 5-nucleotide units starting from both ends, and cytosine (C) and guanine (G) nucleotides within these units were 2′-*O*-methylated. Four modified variants were generated: Apc003-left, Apc003-right, Apc003-both, and Apc003-all (fully 2′-*O*-methylated on all C and G nucleotides). Their binding affinities were subsequently assessed using BLI assays. And the serum stability of these variants, alongside unmodified Apc003, was then evaluated in 10% FBS and 100% FBS, respectively.

The BLI data demonstrated that terminal modifications significantly enhanced target binding affinity relative to unmodified Apc003 (K_d = 86 nmol/L) (Supporting Information Fig. S5a). Modifications at the 5′- and 3′-termini reduced Kd values to 15.1 and 11.9 nmol/L, respectively, while dual-terminal modification (Apc003-both) achieved 12.4 nmol/L. Strikingly, comprehensive C/G modification (Apc003-all) further increased affinity to 9.89 nmol/L (Fig. S5a), demonstrating a positive correlation between modification density and binding optimization (Fig. S5a).

Serum stability profiling showed differential resistance patterns: In 10% FBS, all variants maintained structural integrity >72 h. Under stringent 100% FBS conditions, partially modified constructs (Apc003-left/right/both) underwent complete degradation within 12 h, but Apc003-all exhibited exceptional stability, retaining >95% structural integrity at 72 h with electrophoretic band intensity comparable to baseline (T0) (Fig. S5b). This superior serum resistance motivated its selection as the lead candidate (designated OMe-Apc003) for subsequent studies.

To evaluate the nuclease resistance conferred by methoxy modification, we incubated both unmodified (Apc003) and modified Apc003 (OMe-modified Apc003) with liver enzymes. The modified aptamer demonstrated significantly enhanced stability at all timepoints tested. After just 2 h of incubation, 89% of the OMe-modified Apc003 remained intact compared to only 72% of the unmodified version. This protective effect was even more pronounced after 18 h, with 73% of the OMe-modified aptamer remaining *versus* 51% of the unmodified control (Fig. S5c). Electrophoresis analysis supported these findings, showing minimal fragmentation bands for the methoxy-modified group at all timepoints (0, 2, and 18 h), while the unmodified aptamer exhibited progressive degradation. These results clearly demonstrate that methoxy modifications effectively protect Apc003 from enzymatic cleavage.

Collectively, methoxy modification could facilitate enhancing the *in vitro* stability of the modified Apc003 in both serum and metabolic conditions.

3.8. Long-lasting fatty acid modification via octadecadienoic acid (OA) could extend the half-life of the conjugated Apc003 in normal rats

Nucleic acid aptamers (~12 kDa) typically undergo rapid renal clearance due to their sub-threshold molecular weight (<30−50

kDa), severely limiting therapeutic utility[35]. To extend the *in vivo* half-life of Apc003 (with full 2′-*O*-methylation of all C and G nucleotides), we used our well-developed long-lasting fatty acid modification *via* OA to design the long-lasting the CTGF aptamer (Apc003OA) (Supporting Information Figs. S1 and S6). OA could prolong the half-life of the conjugated nucleic acid aptamers *via* binding human serum albumin (HSA) to form a molecular complex with an average mass above the cut-off threshold of the renal filtration[36]. The pharmacokinetics studies indicated that OA could facilitate extending the half-life of the conjugated Apc003 to 7 days in normal rats (Fig. 6). We also evaluated the pharmacokinetics profiles of Apc003 modified with other fatty acids of varying chain lengths (DA: C12; PA: C17) or polyethylene glycol (PE). The results showed that Apc003OA could exhibit a longer half-life ($t_{1/2}$ = 163.18 h) than Apc003DA ($t_{1/2}$ = 106.61 h), Apc003PA ($t_{1/2}$ = 73.93 h) and Apc003PE ($t_{1/2}$ = 62.57 h) (Supporting Information Fig. S7), indicating that DA/PA/PE modifications were less effective than OA in prolonging the half-life (Fig. 5).

3.9. *Apc003OA exhibited prolonged distribution and retention in muscle tissues compared to FG-3019*

To evaluate the *in vivo* distribution of Apc003OA and FG-3019 in mdx mice, 12-week-old mdx mice were subcutaneously administered with 5 mg/kg of Cy3 labeled Apc003OA and FG-3019, respectively. After 2, 4, 12 and 24 h, the mdx mice were euthanized, respectively. Gastrocnemius muscle, diaphragm and main organs (heart, liver, spleen, lungs, kidney), were collected and imaged by IVIS imaging system. Biophotonic imaging analysis demonstrated that fluorescence intensity was primarily enriched in the gastrocnemius muscle, diaphragm, liver, spleen, lungs, and kidneys, with minimal accumulation in the heart. At all observed time points (2, 4, 12, and 24 h), the Apc003OA group exhibited significantly higher fluorescence intensity in both the diaphragm and gastrocnemius muscle compared to the FG-3019 group. Apc003OA maintained sustained fluorescence intensity in the diaphragm for up to 12 h post-injection. Comparably, the FG-3019 group showed no detectable fluorescence in either tissue beyond 4 h (Fig. 6). The above data indicated that Apc003OA could distribute and remain in muscle tissues for an extended period, whereas FG-3019 could not do so.

3.10. *Apc003OA exerted better fibrosis inhibitory activity without elevating TGF-β1 levels than FG-3019 in mdx mice*

To evaluate the fibrosis inhibitory activity of the Apc003OA on muscle fibrosis of DMD, the effects of Apc003OA on muscle fibrosis were investigated in the well-established mouse model of DMD (mdx mice, C57BL/10ScSn-*Dmd*^{Mdx}/J mice)[37]. 12-week-old mdx mice (n = 10/group) were subcutaneously administered with: (1) vehicle control (PBS), (2) 100 mg/kg FG-3019 (VWC-domain specific antibody), (3) 100 mg/kg Apc003OA, (4) 100 mg/kg Apc003OA+100 mg/kg BP and (5) BP, respectively, once a week for a 12-week (4 months) period (Supporting Information

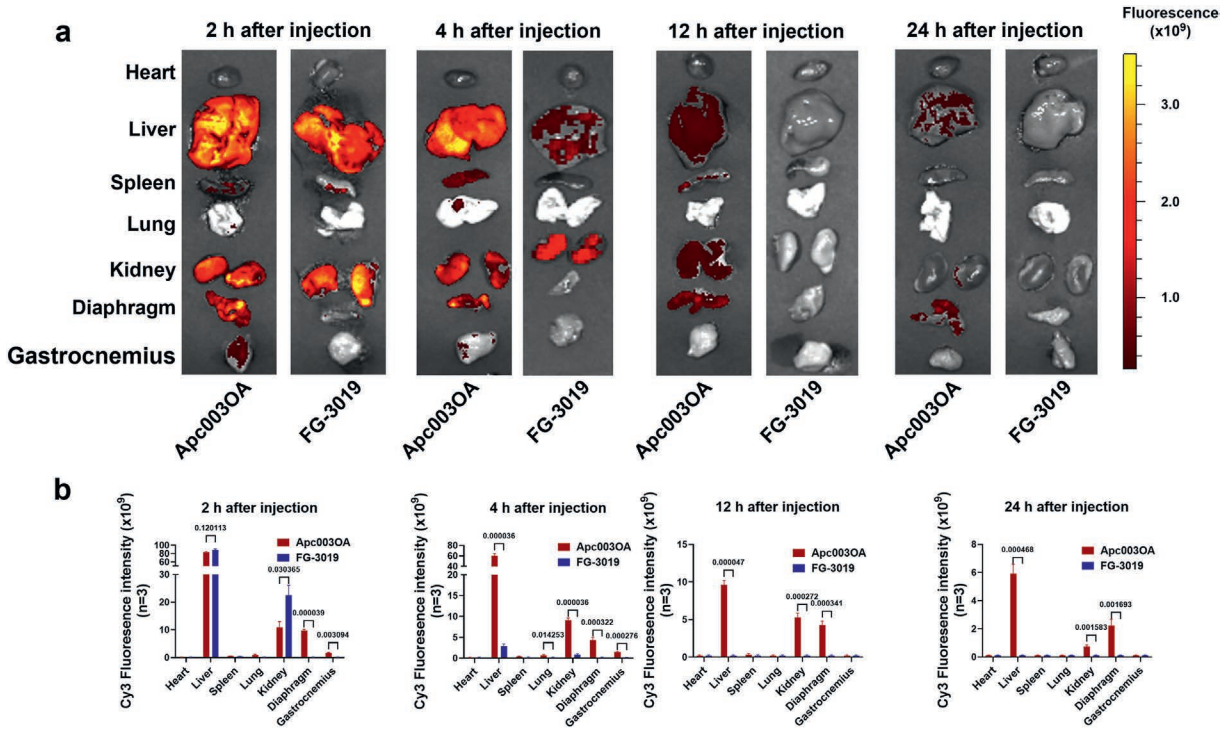


Figure 6 Apc003OA could distribute and remain in muscle tissues for an extended period, whereas FG-3019 could not do so. (a) The *in vivo* distribution of Cy3-Apc003OA and Cy3-FG-3019 in muscles (diaphragm and gastrocnemius) and major organs (heart, liver, spleen, lung, and kidney) 2, 4, 12 and 24 h after subcutaneous injection, respectively. (b) Fluorescence intensities of Cy3-Apc003OA and Cy3-FG-3019 in muscles and major organs 2, 4, 12 and 24 h after subcutaneous injection, respectively. Data are expressed as mean ± standard deviation followed by multiple unpaired *t* tests, n = 3 per group. *P < 0.05. **P < 0.01. ***P < 0.001. ****P < 0.0001. The P values for the respective comparisons are indicated on the graph. Note: The data were presented as the means ± standard deviation.

Table S4). Following the 12-week treatment, compared to vehicle or FG-3019 treatment, the soleus muscle of mdx mice produced higher muscle specific force (improved intrinsic contractile capacity) (Fig. 7a). Besides, the mdx mice in Apc003OA group showed higher forelimb grip strengths (a key measure of overall muscular performance in dystrophic mice) (Fig. 7b). Consistent with muscle functional improvement, the levels of fibrosis markers (fibronectin and collagen III) in diaphragm muscle were significantly down-regulated (Fig. 7c). Next, Masson's staining was employed to determine the fibrosis index of both gastrocnemius muscle and diaphragm muscle. The fibrosis index was significantly lower in Apc003OA group than that of vehicle group or FG-3019 group (Fig. 7d). Critically, the effects of Apc003OA on muscle function and muscle fibrosis level were completely abolished by BP co-administration (Fig. 7a−d), strongly suggesting that its fibrosis inhibitory activity was dependent on binding to the CT-domain of CTGF and not off-target effects (Fig. 7a−d). Moreover, serum TGF-β1 levels in Apc003OA group remained statistically indistinguishable from vehicle controls, no compensatory elevation in serum TGF-β1 levels was found after the Apc003OA treatment, whereas compensatory elevation in serum TGF-β1 levels was found after FG-3019 treatment (Fig. 7e).

Moreover, to investigate the long-term treatment effect of Apc003OA in mdx mice, 12-week-old mdx mice (n = 10/group) were subcutaneously administered with: (1) vehicle control (PBS), (2) 100 mg/kg FG-3019 and (3) 100 mg/kg Apc003OA once a week for 5 months. Apc003OA group showed higher muscle specific force and forelimb grip strengths compared to PBS group or FG-3019 group (Supporting Information Fig. S8a and S8b). Consistently, compared to the PBS or FG-3019 groups, Apc003OA exhibited significant downregulation of fibrosis markers (fibronectin and collagen III) in diaphragm muscle (Fig. S8c and S8d). Importantly, no compensatory elevation in serum TGF-β1 levels was found after the Apc003OA treatment, whereas compensatory elevation in serum TGF-β1 levels was found after FG-3019 treatment (Fig. S8e).

Collectively, these data demonstrated that Apc003OA showed superior fibrosis inhibitory activity compared to FG-3019 in the mdx model, as evidenced by enhanced muscle force generation, improved functional strength, and reduced fibrosis markers accumulation in key skeletal muscles. Moreover, Apc003OA achieved enhanced therapeutic effects without elevating compensatory TGF-β1 levels. This suggested Apc003OA could potentially offer an effective therapeutic strategy for muscle fibrosis in DMD.

3.11. The 12-week Apc003OA treatment showed no obvious organ toxicity in mdx mice

To evaluate the long-term safety profile of Apc003OA, a systematic histopathological assessment of major organs was performed. After 12 weeks of Apc003OA treatment (100 mg/kg), H&E staining was performed on the major organs (heart, liver, spleen, lung, kidney) to detect any structural abnormalities, cellular damage, or inflammatory responses, respectively (Supporting Information Fig. S9). No evidence of obvious lesions, necrosis, inflammatory infiltrates, or other histopathological injuries were detected in the major organs, indicating the safety of Apc003OA at the employed dosage in mdx mice.

4. Discussion

Our *in vitro* genetic studies demonstrated that CTGF-ΔCT led to lower fibrosis markers level than CTGF-ΔVWC. Consistently, our *in vivo* experiments showed that Apc003OA (targeting the CT-domain) outperformed the FG-3019 (targeting the VWC-domain) in reducing fibrosis and improving muscle force in mdx mice. Moreover, both CTGF-ΔVWC and FG-3019 treatment elevated TGF-β1 levels *in vitro*, which was consistent with our *in vivo* data that FG-3019 treatment elevated the TGF-β1 levels in mdx mice. Furthermore, we demonstrated that the elevated compensatory TGF-β1 could attenuate the fibrosis inhibitory activity of FG-3019 both *in vitro* and *in vivo*, which might explain the reason for the failure of FG-3019 in the DMD Phase III clinical trials (https://investor.fibrogen.com/news-releases/news-release-details/fibrogen-announces-topline-results-lelantos-2-phase-3-clinical).

Nucleic acid aptamers are single-stranded DNA/RNA oligonucleotides typically 20−80 nt in length[38]. As promising therapeutic agents, aptamers offer several key advantages: 1) facile chemical modification through solid-phase synthesis allowing precise incorporation of functional groups; 2) high thermal stability with shelf life exceeding 12 months at −20 °C; 3) excellent batch-to-batch reproducibility due to chemical synthesis process; and 4) minimal immunogenicity as demonstrated in clinical trials[39]. Specifically, aptamers can rapidly recognize and bind to flexible protein regions like CT-domain of CTGF through their unique 3D folding structures, achieving nanomolar binding affinity.1224 Their small size enables efficient tissue penetration, particularly in muscle tissues where vascular permeability is limited[40]. Moreover, the low immunogenicity of aptamers minimizes the risk of immune responses. Therefore, it is desirable to develop a CT-domain specific aptamer for DMD treatment. Traditionally, Aptamers could be identified through SELEX. However, this process can be time-consuming and technically demanding, requiring careful optimization of selection conditions and downstream validation. To accelerate the identification of a highly specific aptamer targeting the CT domain of CTGF, we developed an aptamer generative platform, AptGEN, to obtain a CT-domain specific aptamer. AptGEN is an integrated platform that combines SELEX-seq data and deep learning. It employs a novel framework integrating Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN), establishing a new paradigm for aptamer discovery. This approach overcomes the inefficiencies of traditional SELEX and the limitations of existing computational methods. Among the top ten ranked aptamer sequences generated by AptGEN, experimental validation confirmed high affinity, with a correlation of 0.917 between predicted sequence scores and affinity rankings. Traditional SELEX requires 20 rounds to enrich for high-affinity aptamers[24,25], while existing models (such as DeepAptamer, RaptGen etc.) still require 10−20 rounds as input to yield robust binding sequences[41,42]. In contrast, AptGEN achieves comparable data with only five rounds of SELEX data, significantly accelerating drug discovery by efficiently identifying viable therapeutic candidates. To quantitatively

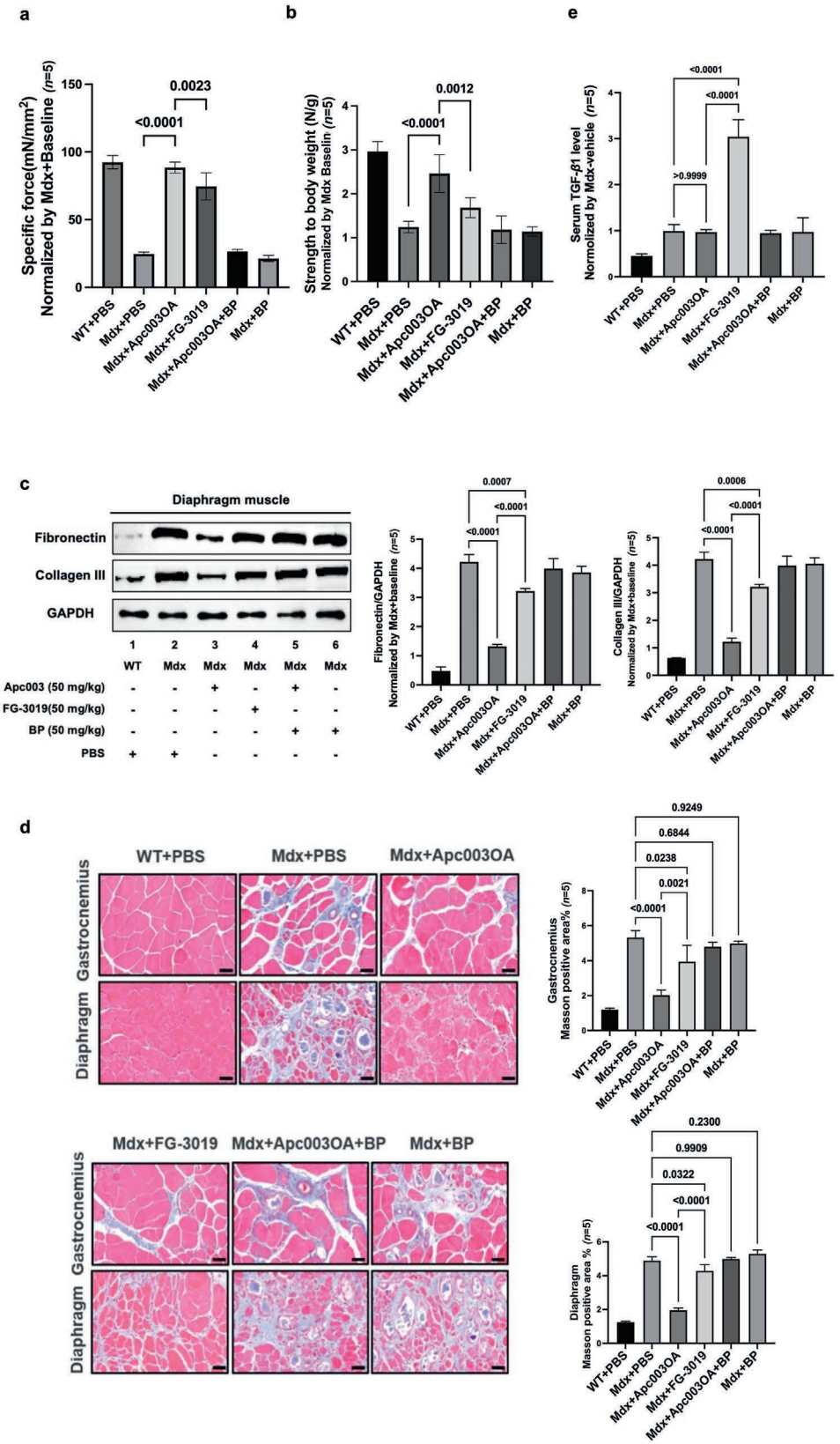


Figure 7 Apc003OA exerted better fibrosis inhibitory activity without elevating compensatory TGF-β1 levels than FG-3019 in mdx mice. (a) The specific force against stimulation frequency of soleus muscles in WT mice treated with PBS, and mdx mice treated with PBS, Apc003OA, FG-3019, Apc003OA + BP, and BP for 12 weeks, respectively. (b) The forelimb grip strengths of mdx mice treated with PBS, and mdx mice treated with PBS, Apc003OA, FG-3019, Apc003OA + BP and BP for 12 weeks, respectively. (c) The fibronectin and collagen III levels in diaphragm muscle of WT mice treated with PBS, mdx mice treated with PBS, Apc003OA, FG-3019, Apc003OA + BP and BP for 12 weeks,

demonstrate these improvements, we conducted comprehensive performance comparisons (Supporting Information Table S5 and Fig. S10). Probability density analysis revealed that traditional SELEX produced a broad affinity distribution (15.5−170 nmol/L, mean: 74.8 ± 54.7 nmol/L), while AptGEN generated a narrow, concentrated peak (15.9−23.6 nmol/L, mean: 19.0 ± 3.0 nmol/L), representing 4-fold improvement in average affinity (Fig. S10). Beyond affinity improvements, AptGEN reduced the screening cycle from ∼100 days (20 rounds) to 42 days (5 rounds), achieving 2.4-fold acceleration with 75% fewer SELEX rounds (Supporting Information Table S5). These quantitative comparisons, conducted under identical experimental conditions targeting CTGF CT-domain, validate AptGEN as a significant advancement in aptamer discovery.

In the AptGEN framework, VAE serves as the backbone for sequence generation, encoding early-round SELEX nucleotide data into a latent space that captures key biochemical features, such as conserved motifs (*e.g.,* "CCCTC") and inferred structural characteristics (*e.g.,* stem-loop preferences). By reconstructing sequences from this latent space, VAE ensures that generated candidates retain the functional elements required for binding while also exploring novel variations. The integration of GAN with VAE establishes a synergistic framework where the adversarial mechanism serves as an adaptive refinement module, progressively optimizing the generated sequences to achieve enhanced consistency with binding strength. The VAE provides sequence diversity and structural feasibility, and the GAN discriminator leverages SELEX-based high-affinity data to iteratively remove low-binding sequences. This iterative adversarial interaction addresses a critical limitation of diffusion-based aptamer generation models such as AptaDiff, which can generate diverse sequences but lack of a mechanism to enforce affinity-driven optimization[43]. Moreover, AptGEN differs from other aptamer-generating algorithms like RaptGen, which relies on rule-based mutations and its diversity is constrained by predefined sequence templates. Similarly, AptaTrans, which employs molecular docking and transformers to capture long-range dependencies, struggles to generate structurally viable conformations for highly flexible targets such as the CT-domain of CTGF[44]. By integrating both generative and discriminative mechanisms, AptGEN overcomes the limitation of lacking binding strength-driven optimization in diffusion models and surpassing the generation bottlenecks of rule-based mutations and single-structure predictions. This synergistic combination of VAE and GAN surpasses the capabilities of conventional approaches that increases the model's performance and reduces the screening cycles.

AptGEN not only enhances the efficiency of aptamer drug discovery but also provides a tool for studying nucleic acid−-protein interactions in dynamic conformations. The established platform enables rapid generation of highly specific and potent aptamer sequences against the targets with flexible conformations in dense tissue, like CT-domain in muscle tissue, which are hard for small molecule to bind (due to lack of stable pockets) and inaccessible for antibodies (due to the poor tissue permeability). Building on the success demonstrated in this study, the broader applicability of the AptGEN platform is a key area of future exploration. Leveraging our prior experience with predictive models like DeepAptamer, we are already extending AptGEN's application to other therapeutically relevant targets, including DKK1 (Dickkopf-related protein 1) and BCMA (B-cell maturation antigen) [41]. While initial results in generating novel sequences for these targets are promising, comprehensive pharmacodynamic evaluations will be essential to fully validate their *in vivo* efficacy. However, the true significance of AptGEN extends beyond any single target. This work establishes a new blueprint for therapeutic discovery, where generative AI can interpret and extrapolate from minimal, early-stage experimental data to rationally design potent drug candidates. This paradigm shift has the potential to dramatically reduce the time, cost, and failure rates associated with traditional discovery methods. In conclusion, AptGEN not only delivers a promising therapeutic candidate for DMD but also offers a validated, generalizable platform poised to accelerate the development of the next generation of aptamer-based diagnostics and therapies.

While nucleic acid aptamers showed great therapeutic potential, their clinical translation has been hampered by two fundamental limitations: rapid nuclease degradation in biological fluids and extremely short *in vivo* half-life due to renal clearance. To overcome these challenges, we implemented a comprehensive chemical optimization strategy. Firstly, methoxy modifications were introduced to dramatically improve its stability, with modified aptamers demonstrating >10-fold greater resistance to serum nucleases and hepatic metabolism compared to unmodified counterparts. For *in vivo* persistence, we conducted systematic evaluations of various half-life extension strategies including PEGylation (5−40 kDa) and fatty acid modifications. Pharmacokinetic profiling revealed OA modification as the optimal approach, with Apc003OA showing longer circulation half-life than PEGylated versions or shorter fatty acid-modified conjugates. This superiority stems from OA's unique ability to bind human serum albumin, forming stable complexes that effectively evade renal clearance. The combination of methoxy modification and OA conjugation yielded Apc003OA, a therapeutic candidate that simultaneously addresses both intrinsic limitations of therapeutic aptamers. This successful optimization demonstrates how strategic chemical modifications can transform these molecules from research tools into viable clinical candidates. The training dataset of AptGEN does not include activity data of chemically modified aptamers, as such data are extremely limited and insufficient for effective model training. However, our group has begun addressing this challenge through our previously developed DeepModify framework, which pioneered an iterative experimental-computational approach to predict how specific chemical modifications at different sequence positions affect aptamer binding affinity[45]. In our study on sclerostin-targeting aptamers, DeepModify achieved a correlation coefficient of 0.82

respectively (left). Statistical analysis of fibronectin and collagen III intensity ratios to GAPDH (right). (d) The diaphragm muscle using Masson's staining from WT mice treated with PBS, and mdx mice treated with PBS, Apc003OA, FG-3019, Apc003OA + BP, and BP for 12 weeks, respectively (left). The statistical analysis of fibrosis index (right). (e) The serum levels of TGF-β1 in WT mice treated with PBS, and mdx mice treated with PBS, Apc003OA, FG-3019, Apc003OA + BP and BP for 12 weeks, respectively. Data are expressed as mean ± standard deviation followed by one-way ANOVA with Tukey's *post hoc* test, $n = 5$ per group. *$P < 0.05$. **$P < 0.01$. ***$P < 0.001$. ****$P < 0.0001$. The P values for the respective comparisons are indicated on the graph. Note: BP: Blocking peptide. WT: wild type (C57BL/10ScSn mice). Administration route: subcutaneous injection. Dosing frequency: once a week.

using 422 modified aptamer-affinity data points, demonstrating the feasibility of computationally predicting modification effects. Building on this foundation, we envision integrating DeepModify's modification prediction capabilities with AptGEN's generative framework in future work. This pipeline would enable: (1) AptGEN generating diverse high-affinity unmodified sequences, (2) DeepModify computationally screening various modification patterns for optimal stability and pharmacokinetics, and (3) focused experimental validation of the most promising modified candidates. As larger datasets become available through systematic characterization efforts, we aim to develop end-to-end models capable of directly generating chemically optimized, drug-like aptamer sequences.

5. Conclusions

Collectively, in this study, we developed an aptamer generative platform, AptGEN, to obtain a CT-domain specific aptamer (Apc003OA) with high affinity in 42 days. This study proposed a novel precision inhibition strategy targeting the CT-domain of CTGF rather than VWC-domain and developed an efficient therapeutic aptamer candidate targeting the CT-domain of CTGF than FG-3019 targeting VWC-domain to counteract muscle fibrosis of DMD, offering a transformative approach to address an unmet clinical need.

Acknowledgments

This study was supported by Hong Kong General Research Fund (CUHK 14108322, CUHK 14109721, CUHK 14103121, CUHK 14103420, HKBU 12100125, HKBU 12100921, HKBU 12102524, HKBU 12102223, China), the Young Scientists Fund of the National Natural Science Foundation of China (82304378), Theme-based Research Scheme (T12-201-20R, China), Guangdong-Hong Kong Technology Cooperation Funding Scheme (GHP/149/21GD, 2023A0505010015, China), the National Natural Science Foundation of China (82372359), Hong Kong Pneumoconiosis Compensation Fund Board Research Fund (2023RS06, China), Hong Kong Innovation and Technology Fund-Partnership Research Programme (PRP/044/24FX, China), HKBU start-up Grant (165546, China).

Author contributions

Baoting Zhang, Ge Zhang, Yixin He, Yuanyuan Yu, Zongkang Zhang & SifanYu devised the project. Shanshan Yao identified the CTGF aptamer. Huarui Zhang and Sifan Yu synthesized the small molecules and aptamer conjugates. Shanshan Yao, Meishen Ren, Huarui Zhang, and Sifan Yu, designed and performed the *in vitro* and *in vivo* study. Xin Yang, and Shijian Ding performed the computational work. Yixin He, provided computing support. Tienan Chen & Yufei Pan assisted with *in vitro* and *in vivo* samples processing. Shanshan Yao, Xin Yang, Meishen Ren, Huarui Zhang, and Sifan Yu wrote the paper with the input from all authors. All authors discussed the results in detail and commented on the manuscript.

Conflicts of interest

The authors declare no competing interests.

Data availability statement

The data used to support the findings of this study are included within the article.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2026.02.016.

References

1. Yao S, Chen Z, Yu Y, Zhang N, Jiang H, Zhang G, et al. Current pharmacological strategies for Duchenne muscular dystrophy. *Front Cell Dev Biol* 2021;**9**:689533.
2. Fairclough RJ, Wood MJ, Davies KE. Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. *Nat Rev Genet* 2013;**14**:373−8.
3. Jones D. Duchenne muscular dystrophy awaits gene therapy. *Nat Biotechnol* 2019;**37**:335−7.
4. Markati T, Oskoui M, Farrar MA, Duong T, Goemans N, Servais L. Emerging therapies for Duchenne muscular dystrophy. *Lancet Neurol* 2022;**21**:814−29.
5. Pessina P, Cabrera D, Morales MG, Riquelme CA, Gutiérrez J, Serrano AL, et al. Novel and optimized strategies for inducing fibrosis *in vivo*: focus on Duchenne muscular dystrophy. *Skeletal Muscle* 2014;**4**:7.
6. Kharraz Y, Guerra J, Pessina P, Serrano AL, Muñoz-Cánoves P. Understanding the process of fibrosis in Duchenne muscular dystrophy. *BioMed Res Int* 2014;**2014**:965631.
7. Lipson KE, Wong C, Teng Y, Spong S. Ctgf is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis. *Fibrogenesis Tissue Repair* 2012;**5**:S24.
8. Song Y, Yao S, Liu Y, Long L, Yang H, Li Q, et al. Expression levels of tgf-β1 and ctgf are associated with the severity of Duchenne muscular dystrophy. *Exp Ther Med* 2017;**13**:1209−14.
9. Morales MG, Gutierrez J, Cabello-Verrugio C, Cabrera D, Lipson KE, Goldschmeding R, et al. Reducing ctgf/ccn2 slows down mdx muscle dystrophy and improves cell therapy. *Hum Mol Genet* 2013;**22**:4938−51.
10. Chen Z, Zhang N, Chu HY, Yu Y, Zhang ZK, Zhang G, et al. Connective tissue growth factor: from molecular understandings to drug discovery. *Front Cell Dev Biol* 2020;**8**:593269.
11. Connolly AM, Zaidman CM, Brandsema JF, Phan HC, Tian C, Zhang X, et al. Pamrevlumab, a fully human monoclonal antibody targeting connective tissue growth factor, for non-ambulatory patients with Duchenne muscular dystrophy. *J Neuromuscul Dis* 2023;**10**:685−99.
12. Ren M, Yao S, Chen T, Luo H, Tao X, Jiang H, et al. Connective tissue growth factor: regulation, diseases, and drug discovery. *Int J Mol Sci* 2024;**25**(9):4692.
13. Zhuo Z, Yu Y, Wang M, Li J, Zhang Z, Liu J, et al. Recent advances in selex technology and aptamer applications in biomedicine. *Int J Mol Sci* 2017;**18**(10):2142.
14. Keefe AD, Pai S, Ellington A. Aptamers as therapeutics. *Nat Rev Drug Discov* 2010;**9**:537−50.
15. Safarkhani M, Ahmadi S, Ipakchi H, Saeb MR, Makvandi P, Ebrahimi Warkiani M, et al. Advancements in aptamer-driven DNA nanostructures for precision drug delivery. *Adv Sci (Weinh)* 2024;**11**:e2401617.
16. Elskens JP, Elskens JM, Madder A. Chemical modification of aptamers for increased binding affinity in diagnostic applications: current status and future prospects. *Int J Mol Sci* 2020;**21**:4522.
17. Domsicova M, Korcekova J, Poturnayova A, Breier A. New insights into aptamers: an alternative to antibodies in the detection of molecular biomarkers. *Int J Mol Sci* 2024;**25**:6833.

18. Zhou J, Rossi J. Aptamers as targeted therapeutics: ccrrent potential and challenges. *Nat Rev Drug Discov* 2017;**16**:181−202.
19. Shraim AS, Abdel Majeed BA, Al-Binni MA, Hunaiti A. Therapeutic potential of aptamer−protein interactions. *ACS Pharmacol Transl Sci* 2022;**5**:1211−27.
20. Wong K-Y, Wong M-S, Lee JH, Liu J. From cell-selex to tissue-selex for targeted drug delivery and aptamer nanomedicine. *Adv Drug Deliv Rev* 2025;**224**:115646.
21. Hao L, Gu H. Introduction of aptamer, SELEX, and different SELEX variants. In: Dong Y, editor. *Aptamers for medical aapplications*. Singapore: Springer; 2021.
22. Fang Z, Feng X, Tang F, Jiang H, Han S, Tao R, et al. Aptamer screening: current methods and future trend towards Non-selex approach. *Biosensors* 2024;**14**:350.
23. Chinchilla-Cárdenas DJ, Cruz-Méndez JS, Petano-Duque JM, García RO, Castro LR, Lobo-Castañón MJ, et al. Current developments of selex technologies and prospects in the aptamer selection with clinical applications. *J Genet Eng Biotechnol* 2024;**22**: 100400.
24. 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**:5645−74.
25. 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. *Nat Commun* 2022;**13**:4241.
26. Hwang GT, Seo YJ, Kim BH. A highly discriminating quencher-free molecular beacon for probing DNA. *J Am Chem Soc* 2004;**126**: 6528−9.
27. 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. *Mol Ther Nucleic Acids* 2023;**34**:102073.
28. Zhang B-T, Yeung SS, Liu Y, Wang H-H, Wan Y-M, Ling S-K, et al. The effects of low frequency electrical stimulation on satellite cell activity in rat skeletal muscle during hindlimb suspension. *BMC Cell Biol* 2010;**11**:87.
29. Meng XM, Nikolic-Paterson DJ, Lan HY. Tgf-β: the master regulator of fibrosis. *Nat Rev Nephrol* 2016;**12**:325−38.
30. Jackson JW, Frederick Jr CS, Pal A, Coricor G, Boston C, Brueckner CT, et al. An antibody that inhibits TGF-β1 release from latent extracellular matrix complexes attenuates the progression of renal fibrosis. *Sci Signal* 2024;**17**. eadn6052.
31. Abreu JG, Ketpura NI, Reversade B, De Robertis EM. Connective-tissue growth factor (CTGF) modulates cell signalling by bmp and tgf-beta. *Nat Cell Biol* 2002;**4**:599−604.
32. Wang Q, Usinger W, Nichols B, Gray J, Xu L, Seeley TW, et al. Cooperative interaction of ctgf and TGF-β in animal models of fibrotic disease. *Fibrogenesis Tissue Repair* 2011;**4**:4.
33. Feng L, Sun Y, Jia W, Yu Y, Liu C, Yang J, et al. Advancements in selex technology for aptamers and emerging applications in therapeutics and drug delivery. *Biomolecules* 2025;**15**: 818.
34. 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. *Nat Commun* 2022;**13**:4241.
35. Zhang Y, Zhang H, Chan DWH, Ma Y, Lu A, Yu S, et al. Strategies for developing long-lasting therapeutic nucleic acid aptamer targeting circulating protein: the present and the future. *Front Cell Dev Biol* 2022;**10**:1048148.
36. 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. *Mol Ther Nucleic Acids* 2023;**34**:102073.
37. Yucel N, Chang AC, Day JW, Rosenthal N, Blau HM. Humanizing the mdx mouse model of dmd: the long and the short of it. *NPJ Regen Med* 2018;**3**:4.
38. Bashir A, Yang Q, Wang J, Hoyer S, Chou W, McLean C, et al. Machine learning guided aptamer refinement and discovery. *Nat Commun* 2021;**12**:2366.
39. Marton S, Reyes-Darias JA, Sánchez-Luque FJ, Romero-López C, Berzal-Herranz A. *In vitro* and *ex vivo* selection procedures for identifying potentially therapeutic DNA and RNA molecules. *Molecules* 2010;**15**:4610−38.
40. Song KM, Lee S, Ban C. Aptamers and their biological applications. *Sensors (Basel)* 2012;**12**:612−31.
41. Yang X, Chan CH, Yao S, Chu HY, Lyu M, Chen Z, et al. Deepaptamer: advancing high-affinity aptamer discovery with a hybrid deep learning model. *Mol Ther Nucleic Acids* 2025;**36**: 102436.
42. Iwano N, Adachi T, Aoki K, Nakamura Y, Hamada M. Generative aptamer discovery using raptgen. *Nat Comput Sci* 2022;**2**: 378−86.
43. Wang Z, Liu Z, Zhang W, Li Y, Feng Y, Lv S, et al. Aptadiff: *de novo* design and optimization of aptamers based on diffusion models. *Briefings Bioinf* 2024;**25**. bbae517.
44. Shin I, Kang K, Kim J, Sel S, Choi J, Lee JW, et al. Aptatrans: a deep neural network for predicting aptamer−protein interaction using pre-trained encoders. *BMC Bioinf* 2023;**24**:447.
45. Amu G, Yang X, Luo H, Yu S, Zhang H, Tian Y, et al. Machine learning-powered, high-affinity modification strategies for aptamers. *Acta Mater Med* 2025;**4**:122−36.