DOI: https://doi.org/10.54393/fbt.v6i2.245

FUTURISTIC BIOTECHNOLOGY
https://fbtjournal.com/index.php/fbt
ISSN (E): 2959-0981, (P): 2959-0973
Vol 06 Issue 02, (April-June, 2026)

Review Article

OPEN ACCESS

AI-Designed Synthetic CRISPR Enzymes: Evolving Molecular Scissors Beyond Natural Constraints

Shaista Khalid[1*]

[1]Department of Biotechnology, Forman Christian College, University, Lahore, Pakistan

ARTICLE INFO

Keywords:
CRISPR-Cas9, Generative AI, Protein Language Models, Protein Design, PAM Engineering, Gene Editing, Synthetic Biology, Biosecurity

How to Cite:
Khalid, S. (2026). AI-Designed Synthetic CRISPR Enzymes: Evolving Molecular Scissors Beyond Natural Constraints: AI-Designed Synthetic CRISPR Enzymes. Futuristic Biotechnology, 6(2), 10-17. https://doi.org/10.54393/fbt.v6i2.245

***Corresponding Author:**
Shaista Khalid
Department of Biotechnology, Forman Christian College, University, Lahore, Pakistan
shaaaiista@gmail.com

Received Date: 25th April, 2026
1st Revision Received: 10th May, 2026
2nd Revision Received: 4th June, 2026
Acceptance Date: 22nd June, 2026
Published Date: 30th June, 2026

ABSTRACT

Therapeutic gene editing has historically relied on naturally evolved bacterial nucleases (SpCas9, Cas12a), imposing constraints: rigid protospacer-adjacent motif (PAM) requirements, protein sizes challenging adeno-associated virus packaging, and pre-existing human immunity. This narrative review assesses the paradigm shift from natural-product discovery toward generative artificial intelligence (AI) for engineering non-natural CRISPR-Cas nucleases with user-defined properties. This study synthesizes the design–build–test–learn pipeline encompassing protein language models (ESM3), diffusion-based structure generators (RFdiffusion, ProteinMPNN), and property-specific architectures (Protein2PAM). High-throughput platforms like Sequence Display generate >10 million sequence–activity data points per experiment for closed-loop validation. AI-generated OpenCRISPR-1 (>400 mutations from SpCas9) matches natural on-target efficiency while achieving sub-1% median off-target indels and up to 95% reduction at specific sites. Protein2PAM-evolved Nme1Cas9 variants show broadened PAM tolerance and up to 50-fold increased activity without iterative evolution. Generative AI addresses PAM restriction, packaging, and immunogenicity simultaneously. However, in vivo validation, long-term genotoxicity data, and biosecurity governance gaps remain unresolved.

INTRODUCTION

The first 10 years of genome editing were discovery years. Nuclease orthologs of Cas9 were first described in Streptococcus pyogenes in 2012 [1], and almost all the therapeutic and research type nucleases discovered since have been orthologs of Cas9: Cas12a (Cpf1) was reported in Acidaminococcus and Lachnospiraceae species in 2015 [2], CasX and CasY were mined from uncultivated bacteria and archaea by genome-resolved metagenomics in 2017 [3], and CasΦ, Casπ and CasF were isolated from large bacteriophage genomes and environmental samples between 2020 and 2022 [4, 5]. These natural scaffolds were also expanded upon: base editing was integrated with a catalytically impaired Cas9 and coupled with a deaminase, while prime editing, engineered in 2019, was coupled with a Cas9 nickase and a reverse transcriptase for the direct, template-guided rewriting of DNA sequence without the need for a double-strand break or donor DNA [6]. Even then, these newer tools remain structurally tied to their evolutionary roots, making a move toward de novo invention timely and needed. The decade of discovery and engineering breakthroughs still encountered a universal biological limit; three were especially relevant to therapeutic applications. Repeated in vivo delivery and repeat dosing are directly impacted by pre-existing antibodies and T cells specific for Cas enzymes from bacteria that commonly infect humans [7, 8]. Secondly, each Cas enzyme demands certain short DNA motifs surrounding the target site to bind to the target DNA,

DOI: https://doi.org/10.54393/fbt.v6i2.245

rendering large sections of the genome inaccessible to any one Cas enzyme [9]. Third, protein size continues to be a limitation; compact effectors are highly desirable for delivery in size-constrained delivery vectors like adeno-associated virus (AAV), which motivated much of the previous metagenomic discovery effort [10]. These three factors PAM restriction, size, and immunogenicity constitute therapeutic barriers that AI-designed enzymes must overcome and serve throughout this review as the evaluative framework. A way to overcome these limitations is generative AI, and specifically protein language models and diffusion-based protein structure generators trained at evolutionary scale. These models are not searching the sequence space for existing sequences, but rather responding to a specification: a target PAM, a size limit, a stability profile, and returning candidate protein sequences that have never been found in any organism [11, 12]. This represents a journey from discovery to specification and constitutes the main theme of this review, where AI serves as a design tool rather than a computational filter to predict side effects on natural candidates. The results summarize the representative natural and AI-designed or evolved Cas enzymes reported here, their PAM specificities, size, and origin.

Out of the three constraints, the most important for target selection is that the flanking motif must be found close to a disease-relevant locus; a significant fraction of clinically relevant loci lack a motif recognized by any well-characterized natural enzyme [13]. This drawback has been tackled in the past by either continuing to exhaustively search for natural products or by performing labor-intensive directed evolution, such as phage-assisted continuous evolution, which can lead to off-target activity due to relaxed PAM stringency [14]. There is also a parallel issue of immunogenicity: both anti-Cas9 IgG antibodies and common antigen-specific T cell reactivity against both orthologs (SaCas9 and SpCas9) have been found in 78% and 58% of blood donors, respectively [7]. This has since been repeated in independent studies [8] and inspired the development of Cas9 variants with "immunosilenced" T-cell epitopes [15], a small-scale, structure-driven example of the sort of property engineering that can now be attempted in the context of foundation models. This seroprevalence data, combined with the comparative data in Table 1, provides the constraints above in a practical clinical context. While directed evolution and metagenomic mining have overcome some of these shortcomings, both are laborious, have been shown to yield off-target effects resulting from relaxed specificity [14], and are limited to existing evolutionary starting points and therefore to the capacity for scalable, property-specified enzyme design.

Although generative AI is increasingly being incorporated into the design–build–test–learn pipeline, there is no systematic review that critically evaluates how this pipeline works in practice, nor comprehensive documentation of functional capabilities of the resulting candidates like OpenCRISPR-1 and Protein2PAM-evolved versions, or the technical, regulatory, and governance challenges that remain before therapeutic translation. This narrative review aims to (i) outline the computational and experimental platforms that have been used for the design of Cas nucleases; (ii) contrast the evidence on the on-target efficiency, off-target fidelity, and PAM-redirection capacity of AI-generated nucleases to their natural counterparts; and (iii) highlight the unresolved challenges in in vivo delivery, long-term functional validation, regulatory characterization, and biosecurity screening of AI-developed nucleases. This review is significant because it provides the first comprehensive synthesis of how generative AI is transforming CRISPR nuclease engineering from discovery-based to specification-based design, while also identifying critical gaps that must be addressed for clinical translation.

Computational and Experimental Platforms for AI-Driven Cas Design

In both academic and industry-based labs, the four phases of design–build–test–learn in AI-guided Cas engineering have become unified. These subsections summarize the literature for each step: sequence-based generative modeling, structure-first design, structure-specific architectures for PAM engineering, and the high-throughput experimental infrastructure that bridges the gap between computational prediction and functional validation.

Sequence-Trained Protein Language Models

The generation of sequences is closely related to using large models that have been pre-trained on sequences. Examples of this include the ESM family of models, which was trained on UniRef50, featuring up to 15 billion parameters on the EvolutionaryScale platform (formerly Meta AI) that can model protein structure from sequence alone, and ESMFold, which has built a metagenomic atlas of over 600 million structures predicted from a large, diverse database of protein sequences [14]. Its successor, ESM3, is a multimodal generative model that reasons simultaneously in sequence, structure, and function; it has produced a protein variant (ESM3) with just 58% sequence identity to any natural homolog, which is estimated to have simulated 500 million years of evolution [15]. Profluent Bio (Emeryville, California) compiled a database of ~5.1 million Cas9-like protein sequences and then employed a large language model trained on this database to create millions of new, non-natural CRISPR-Cas candidate sequences [10].

Structure-First Diffusion and CRISPR-Specific Design Tools

DOI: https://doi.org/10.54393/fbt.v6i2.245

The first step is the implementation of structure-first diffusion and CRISPR-specific design tools. Implementing structure-first diffusion and CRISPR-specific design tools comes first. The structure-first approaches are complementary methods that build the protein backbones before the determination of the sequence. They rewrote RoseTTAFold as a denoising diffusion model for de novo backbone generation (RFdiffusion), combined with inverse-folding modeling, like ProteinMPNN [16, 17], to generate native sequences with significantly improved accuracy over physics-based models like Rosetta, and have demonstrated this on a variety of structural classes using crystallography and cryo-EM. There are also agentic systems that specifically assist the creation of CRISPR guides: CRISPR-GPT [18] brings computational tooling, guide-RNA design, and experimental protocol planning in one assistant.

Property-Specific Architectures for PAM Engineering

In this section, property-specific architectures for PAM Engineering are discussed. Property-specific architectures address design issues directly that are narrower. Protein2PAM, developed by Profluent Bio and trained using over 45,000 characterized CRISPR-Cas PAM sequences extracted from around 26.2 terabases of microbial genomic sequences, directly infers the PAM specificity from the primary amino acid sequence for Type I, II, and V systems. Since it is sequence-based and does not need a solved crystal structure, Protein2PAM enables in silico deep mutational scanning to systematically predict functional effects of any possible substitution at the PAM-recognition residues [11].

High-Throughput Experimental Screening and Closed-Loop Validation

The second step is closed-loop validation, a high-throughput experimental screening. Given the size of the sequence space (a 50-residue protein has about 10^65 possible variants), validating such sequences requires high-throughput screening infrastructure. In April 2026, a barcoding system for recording the functional behavior of individual protein variants at scale was published in Nature Biotechnology. Using a small CRISPR-Cas protein as a proof of concept, this platform produced more than 10 million sequence–activity data points in a single experiment [19]. The resultant data was then used to build a model that predicts mutations most likely to result in increased activity of the target protein, thereby closing the loop between generative design and experimental ground truth and serving as a template for closed-loop, self-improving engineering

The Clinical Benchmark: Casgevy and Primary-Cell Validation

The last and least automated part of the pipeline is the validation of AI-designed candidates in disease-relevant primary cells, where they have to compete with established drugs. This is the first FDA-approved CRISPR-Cas9 therapeutic, Casgevy (exagamglogene autotemcel), which was approved in December 2023 and treats sickle cell disease [20] through ex vivo modification of the patient's own hematopoietic stem cells to increase fetal hemoglobin levels. Developed by Vertex Pharmaceuticals (Boston, Massachusetts) and CRISPR Therapeutics (Zug, Switzerland), Casgevy sets the clinical and regulatory benchmark for all future ex vivo hematopoietic stem cell-editing nucleases, from naturally occurring to AI-generated, by demonstrating safety, efficacy and long-term genotoxicity in Good Manufacturing Practice-compliant clinical development environments before moving from discovery to approved therapy (Figure 1).

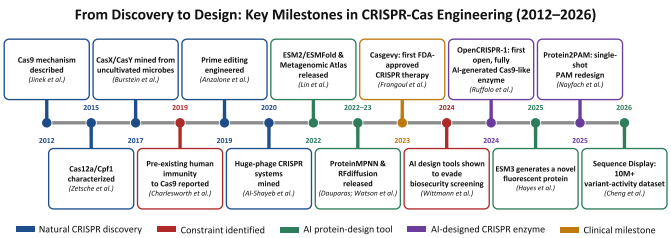

Figure 1: Timeline of Key Milestones Spanning Natural CRISPR-Cas Discovery, Identification of Natural-Enzyme Constraints, and the Emergence of AI Protein-Design Tools and AI-Generated CRISPR Enzymes, 2012–2026

The generalized design–build–test–learn pipeline used across AI-driven Cas-engineering efforts: generation of candidate sequences by a trained foundation model, computational filtering and property prediction, high-throughput experimental screening, and validation in therapeutically relevant cell models, with activity data feeding back into subsequent design rounds (Figure 2).

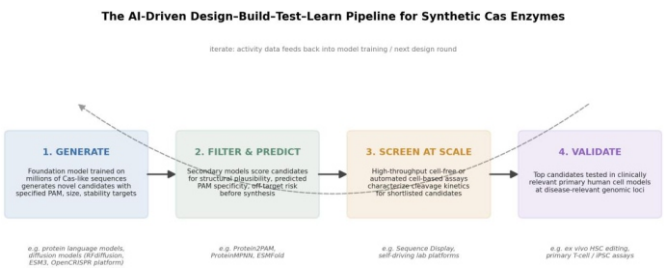

Figure 2: The AI-Driven Design–Build–Test–Learn Pipeline for Cas-Engineering Efforts

Benchmarking AI-Generated Nucleases Against Natural Enzymes

This section synthesizes the comparative evidence for the two most extensively characterized AI-designed Cas platforms, OpenCRISPR-1 and Protein2PAM-evolved Nme1Cas9, alongside the broader infrastructure that supports field-wide convergence toward generative nuclease design. This section highlights the comparative evidence supporting two of the most well characterized AI-

DOI: https://doi.org/10.54393/fbt.v6i2.245

designed Cas platforms, namely, OpenCRISPR-1 and Protein2PAM-evolved Nme1Cas9, as well as the other platform components that enable convergence in the field toward generative nuclease design. The following is available to help you learn more about 3.1 OpenCRISPR-1: A Fully Synthetic Nuclease Matching Natural Specificity. OpenCRISPR-1 is the first example of how generative AI can be used to create a completely novel, non-natural Cas9-like nuclease that is on-target efficient and significantly off-target efficient, with median indel rates of <1% and up to a 95% reduction at some off-target sites. This finding suggests that protein language models can sample and traverse the vast sequence-fitness landscape to find protein variants that do not exist in nature. However, unlike the other Cas9s, OpenCRISPR-1 still has limited genomic space that can be targeted, which can be increased through sequence generation; the expansion of the PAMs would need additional methods like Protein2PAM (Table 1).

Table 1: OpenCRISPR-1: A Fully Synthetic Nuclease Matching Natural Specificity

Feature / Parameter	SpCas9 (Natural)	OpenCRISPR-1 (AI-Generated)	Comparison / Reference
Origin	Streptococcus pyogenes (natural)	De novo, generated by an LLM trained on 5.1M Cas9-like sequences	Synthetic, non-natural [20, 21]
Protein size (aa)	~1,368	~1,300 (Cas9-like architecture)	Comparable size; no packaging advantage reported yet [21]
Protein size (aaaMutations from SpCas9	N/A (reference)	>400 mutations from SpCas9; ~200 from all known natural orthologs	Highly diverged from natural Cas sequence space [21])
PAM preference	5'-NGG-3'	5'-NGG-3'	Compatible with existing guide RNA protocols [21]
On-target editing efficiency	Baseline (100% reference)	Comparable to SpCas9 in HEK293T human cells	Matches natural enzyme performance [22]
Median off-target indel rate	Variable (typically >1-5% at known hotspots)	<1% (reported)	Substantially improved specificity [22]
Off-target reduction at specific sites	Baseline	Up to 95% reduction at certain tested off-target sites	Directional, site-dependent improvement [22]
Availability/accessibility	Widely available (Addgene, commercial)	Openly released for research and commercial use	Described as a "drop-in replacement" for Cas9-like applications [23]

Note: Benchmarking AI-generated nuclease performance: OpenCRISPR-1 versus natural SpCas9 (Aim 1)

Protein2PAM shows that the specificity of PAMs is arguably one of the most critical therapeutic constraints – can be computationally redirected in one design step, giving up to 50-fold higher cleavage activity with a dramatically expanded PAM tolerance without iterative evolution in the lab. This ability transforms protein engineering from a multi-month directed evolution screen to a single computational query, which could help increase adoption of PAM optimization across academia and industrial labs. The Protein2PAM-evolved Nme1Cas9 backbone maintains a compact size of ~1,080 amino acids, effectively overcoming AAV packaging constraints, albeit with the reported activity improvements yet to be validated in human primary cells and in vivo models (Table 2).

Table 2: Protein2PAM-Evolved Nme1Cas9: Computational Redesign of PAM Specificity

Parameter / Feature	Wild-Type Nme1Cas9 (Natural)	Protein2PAM-Evolved Variant	Fold-Change / Design Achievement
Backbone origin	Neisseria meningitidis (natural, compact Cas9 ortholog)	Engineered from natural Nme1Cas9 backbone	Retains compact ~1,080 aa framework [24,25]
PAM specificity	Narrow, defined natural motif	Dramatically broadened PAM tolerance	Directed redesign of a defining natural constraint [25]
Cleavage activity (in vitro)	Baseline	Up to 50-fold increase relative to wild type	Substantial activity enhancement under tested conditions [25]
Design method/effort	N/A (evolved by nature)	Transformer-predicted mutations; in silico deep mutational scanning	Single computational design query vs. multi-month iterative directed evolution [24]
Structural model required?	Yes (for traditional rational design)	No—model operates on sequence alone	Predicts PAM specificity without solving protein-DNA co-crystal structures [24]
Availability	Limited to research labs	Publicly available Protein2PAM server for academic/industry researchers	Enables community-driven PAM optimization on any Cas enzyme [26]

Targeted property redesign: Protein2PAM-evolved Nme1Cas9 versus wild type (Aim 2)

By creating over 10M sequence activity datapoints/experiment, Sequence Display is closing the gap between generative models and experimental ground truth, solving a fundamental bottleneck in AI-driven protein design by generating more

DOI: https://doi.org/10.54393/fbt.v6i2.245

training data. Together with general-purpose protein design tools (ESM, RFdiffusion, ProteinMPNN) and CRISPR-specific agentic systems (CRISPR-GPT), this infrastructure forms a continuum in which natural discovery, rational engineering, and AI-assisted development work in harmony as complementary stages of a pipeline, not as competing strategies. The field-wide convergence reported in independent 2025–2026 reviews is evidence that these Cas enzymes created by AI are now at the heart of an increasingly closed-loop, automated engineering ecosystem (Table 3).

Table 3: Enabling Infrastructure and Field-Level Convergence

Platform / Framework	Core Capability	Experimental / Computational Scale	Relevance to AI-Designed Cas Enzymes
Sequence Display (Rice University)	Activity-based barcoding system for high-throughput functional screening	>10 million sequence–activity datapoints per single experiment	Provides dense experimental ground truth to train/retrain generative models, closing the design–build–test–learn loop [27]
Protein language & diffusion models (ESM2/ESM3, RF diffusion, ProteinMPNN)	Generative design of protein backbones and sequences; ESM3 simulates ~500M years of evolution	Trained on UniRef50 (15B parameters); predicted >600M structures; validated by crystallography/cryo-EM	General-purpose tools increasingly applied to Cas scaffolds; enabled generation of OpenCRISPR-1 and related nucleases [28, 29]
CRISPR-specific agentic systems (CRISPR-GPT)	Integrates computational tooling	Automated design-planning interface	Streamlines transition from computational candidate selection to experimental validation [30]
Field-level integration (ACS Synthetic Biology 2026; EMM 2025)	Convergent pipeline: natural discovery → rational design → directed evolution → AI-assisted development	Field-wide consensus documented in two independent 2025–2026 reviews	Confirms that generative Cas design is a core node in a larger, integrated AI-augmented CRISPR toolset [31, 32]

This review synthesizes findings from peer-reviewed publications, preprints, company technical disclosures, and trade press current as of August 2026. Reported performance figures are drawn directly from the cited primary sources and are presented as those sources describe them, not independently re-derived. Given the pace of this field, readers pursuing specific technical claims should consult primary sources directly. (A) OpenCRISPR-1 versus SpCas9 in human cells: on-target editing efficiency was comparable between the two enzymes, while off-target editing at tested sites was substantially reduced for OpenCRISPR-1, consistent with a reported 95% reduction at certain off-target sites and a sub-1% median indel rate. (B) Cleavage-activity fold-change for a Protein2PAM-evolved Nme1Cas9 variant relative to wild type, following computational PAM-specificity redesign. Values are illustrative summaries of directional findings reported in the cited primary sources [28, 30], not independently re-derived data (Figure 3).

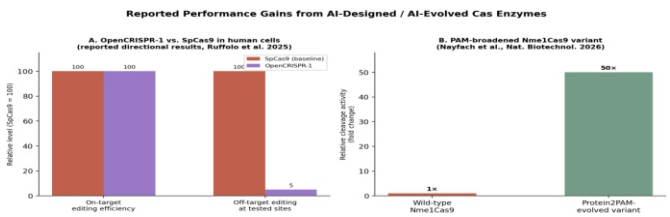

Figure 3: Reported Performance Metrics from the Two Lead AI-Design Case Studies Reviewed

Commercial and Translational Rationale

The commercial logic behind moving from discovery-based to specification-based nuclease engineering is simple: specifying a nuclease for a new target, be it a PAM, a size, or an immunogenicity, becomes a design-and-validate process, instead of a search. Profluent's headquarters are in Emeryville, California, and investors include Spark Capital, Insight Partners, Air Street Capital, and AIX Ventures [24]. Instead, a 2026 ACS review for Synthetic Biology considers natural discovery, rational engineering, and AI-generated generation as complementary steps in a continuum, not competing [27].

Biosecurity Governance Gaps

This commercial and scientific momentum is accompanied by governance challenges that extend beyond CRISPR specifically. An October 2025 study led by researchers together with nucleic acid synthesis and biosecurity organizations showed that open-source AI protein-design tools can generate variants of proteins of concern with low sequence identity to known hazards — meaning such variants can, in some cases, pass the sequence-identity-based screening software used by most DNA synthesis companies to flag potentially hazardous orders [21]. Following a period of confidential, coordinated disclosure comparable to practices used in cybersecurity, the same researchers worked with synthesis companies and policymakers to develop improved detection methods ahead of publication [22]. These sequence-divergent "synthetic homologs" are difficult for identity-based screening to catch, since biological function can be retained despite minimal or no sequence homology to any naturally occurring protein [23]. A 2024 U.S. federal program intended to make compliance with nucleic acid synthesis screening a condition of federal funding eligibility was suspended in 2025 and, as of this writing, has not been replaced [23]. Any pipeline capable of producing a functional, non-natural nuclease can, by the same logic, be used to design enzyme variants that existing screening infrastructure cannot detect.

Profluent's decision to open-source OpenCRISPR-1

DOI: https://doi.org/10.54393/fbt.v6i2.245

appears intended to seed broad, transparent use rather than concentrate the technology within a single organization [33], a defensible stance for a therapeutic tool, but one that places additional weight on synthesis-level screening, institutional biosafety review, and ongoing coordination among AI developers, synthesis providers, and regulators as generative design tools proliferate.

Remaining Barriers to Clinical Translation

Beyond governance, clinical translation of AI-designed nucleases is constrained by four interrelated challenges: delivery remains an inherent property of the nuclease itself; in vivo safety data are still lacking; regulatory pathways specific to AI-generated nucleases have yet to be defined; and model quality remains bounded by training data quality, which high-throughput screening alone cannot fully resolve [19, 34]. Addressing these barriers will likely require coordinated investment in several areas: standardized benchmarking platforms, comparable to CASP, that enable blinded evaluation in primary human cells; closed-loop, self-driving laboratories capable of generating dense functional datasets; interpretable AI architectures that can provide mechanistic explanations suitable for regulatory review; and function-based biosecurity screening capable of identifying hazardous enzymes independent of sequence divergence. The field would also benefit from treating nuclease engineering and delivery as a single integrated design problem for each therapeutic indication, rather than pursuing them as separate innovation streams.

The accelerating convergence of generative AI and CRISPR engineering has recently yielded several landmark advances that extend well beyond the initial demonstrations with OpenCRISPR-1 and Protein2PAM. In a pivotal 2026 study, Doudna and colleagues used a hybrid AI approach combining inverse protein-folding models with evolution-informed sequence constraints to design SynTnpBs—novel RNA-guided nucleases based on the minimal TnpB family—with some variants achieving editing activity equal to or greater than wild-type TnpB while sharing as little as 77% sequence identity with any natural homolog [25]. Notably, these AI-generated nucleases were validated not only in human cells but also in plant and bacterial systems, demonstrating the broad applicability of specification-based design across kingdoms of life [25]. Concurrently, the PAMmla machine learning framework enabled high-throughput engineering of nearly 1,000 SpCas9 variants, training a neural network to predict PAM specificity across 64 million theoretical enzymes and identifying bespoke editors that outperformed natural and engineered counterparts as both nucleases and base editors in human cells while reducing off-target effects validated in a mouse model of retinitis pigmentosa [26]. Complementing these advances, the CICERO model, built on the ESM2 protein language model and trained on the CRISPR-PAMdb database of 3.8 million bacterial and archaeal genomes, now enables PAM prediction for over 50,000 additional Cas9 proteins, dramatically accelerating the discovery of natural and synthetic variants with novel PAM specificities [31]. Finally, agentic AI systems such as CRISPR-GPT have demonstrated fully automated gene-editing experiment design and execution, successfully knocking out four genes with CRISPR-Cas12a and epigenetically activating two genes using CRISPR-dCas9 in human cell lines, validating the feasibility of AI co-pilots for end-to-end genome engineering workflows [32]. Collectively, these developments confirm that the field is transitioning from isolated AI-assisted design efforts toward an integrated ecosystem in which generative models, high-throughput screening, and autonomous experimentation operate in closed-loop synergy, though the translation of these powerful capabilities into approved clinical therapeutics will require continued investment in in vivo validation, regulatory pathway definition, and biosecurity infrastructure [35].

Limitations and Future Recommendations

The results of this study are limited by a number of factors. First, it was carried out in one emergency area in the Sahiwal Division, and the results might not be generalizable to other parts of Pakistan. Second, non-probability consecutive sampling might have resulted in selection bias. Future studies should include several hospitals and Rescue 1122 centres from various regions of Pakistan with larger and probability-based samples to enhance generalizability. Further research should be undertaken that better assesses the impact of early pre-hospital interventions on morbidity, mortality and disability

CONCLUSION

Generative AI has demonstrably begun to dismantle the three historical constraints limiting therapeutic genome editing to naturally evolved nucleases: PAM restriction, delivery-limiting size, and pre-existing immunogenicity, often addressing multiple constraints within a single design cycle. OpenCRISPR-1 shows that specificity can be substantially improved through sequence generation alone, while Protein2PAM demonstrates that PAM tolerance can be computationally redirected without iterative laboratory evolution. Realizing clinical potential depends on in vivo validation, regulatory frameworks for non-natural nucleases, interpretable AI architectures, and biosecurity screening capable of detecting sequence-divergent synthetic homologs. Progress on these fronts, rather than further gains in on-target efficiency alone, will determine how quickly AI-designed Cas enzymes advance from compelling preclinical proofs of concept to approved clinical therapeutics.

DOI: https://doi.org/10.54393/fbt.v6i2.245

Author's Contribution

Conceptualization: SK
Methodology: SK
Formal analysis: SK
Writing and Drafting: SK
Review and Editing: SK

The author approved the final manuscript and take responsibility for the integrity of the work

Conflicts of Interest

The author declare no conflict of interest.

Source of Funding

The author received no financial support for the research, authorship and/or publication of this article.

REFERENCES

[1] Ganguly C, Rostami S, Long K, Aribam SD, Rajan R. Unity among the Diverse RNA-guided CRISPR-Cas Interference Mechanisms. Journal of Biological Chemistry. 2024 Jun; 300(6). doi: 10.1016/j.jbc.2024.107295.

[2] Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P *et al.* Cpf1 is A Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 2015 Oct; 163(3): 759-71. doi: 10.1016/j.cell.2015.09.038.

[3] Alexander LM, Aliaga Goltsman DS, Liu J, Lin JL, Temoche-Diaz MM, Laperriere SM *et al.* Novel and Engineered Type II CRISPR Systems from Uncultivated Microbes with Broad Genome Editing Capability. The CRISPR Journal. 2023 Jun; 6(3): 261-77. doi: 10.1089/crispr.2022.0090.

[4] Al-Shayeb B, Skopintsev P, Soczek KM, Stahl EC, Li Z, Groover E *et al.* Diverse Virus-Encoded CRISPR-Cas Systems Include Streamlined Genome Editors. Cell. 2022 Nov; 185(24): 4574-86. doi: 10.1016/j.cell.2022.10.020.

[5] Sun A, Li CP, Chen Z, Zhang S, Li DY, Yang Y *et al.* The compact Casπ (Cas12l) Bracelet Provides a Unique Structural Platform for DNA Manipulation. Cell Research. 2023 Mar; 33(3): 229-44. doi: 10.1038/s41422-022-00771-2.

[6] Matsoukas IG. Prime Editing: Genome Editing for Rare Genetic Diseases Without Double-Strand Breaks or Donor DNA. Frontiers in Genetics. 2020 Jun; 11: 527211. doi: 10.3389/fgene.2020.00528.

[7] Serpa G, Gong Q, De M, Rana PS, Montgomery CP, Wozniak DJ *et al.* Detection of Pre-Existing Immunity to Bacterial Cas9 Proteins in People with Cystic Fibrosis. Immunohorizons. 2025 Oct; 9(10): vlaf041. 10.1093/immhor/vlaf041.

[8] Mehta A and Merkel OM. Immunogenicity of Cas9 Protein. Journal of Pharmaceutical Sciences. 2020 Jan; 109(1): 62-7. doi: 10.1016/j.xphs.2019.10.003.

[9] Collias D and Beisel CL. CRISPR Technologies and the Search for the PAM-Free Nuclease. Nature Communications. 2021 Jan; 12(1): 555. doi: 10.1038/s41467-020-20633-y.

[10] Ruffolo JA, Nayfach S, Gallagher J, Bhatnagar A, Beazer J, Hussain R *et al.* Design of Highly Functional Genome Editors by Modelling CRISPR-Cas Sequences. Nature. 2025 Sep; 645(8080): 518-25. doi: 10.1038/s41586-025-09298-z.

[11] Nayfach S, Bhatnagar A, Novichkov A, Kim N, Hoffnagle AM, Hussain R *et al.* Customizing CRISPR-Cas PAM Specificity with Protein Language Models. Nature Biotechnology. 2026 Feb: 1-0. doi: 10.1038/s41587-025-02995-0.

[12] Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W *et al.* Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. Science. 2023 Mar; 379(6637): 1123-30. doi: 10.1126/science.ade2574.

[13] Hayes T, Rao R, Akin H, Sofroniew NJ, Oktay D, Lin Z *et al.* Simulating 500 Million Years of Evolution with A Language Model. Science. 2025 Feb; 387(6736): 850-8. doi: 10.1126/science.ads0018.

[14] Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE *et al.* De Novo Design of Protein Structure and Function with RFdiffusion. Nature. 2023 Aug; 620(7976): 1089-100. doi: 10.1038/s41586-023-06415-8.

[15] Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF *et al.* Robust Deep Learning–Based Protein Sequence Design using ProteinMPNN. Science. 2022 Oct; 378(6615): 49-56. doi: 10.1126/science.add2187.

[16] Qu Y, Huang K, Yin M, Zhan K, Liu D, Yin D *et al.* CRISPR-GPT for Agentic Automation of Gene-Editing Experiments. Nature Biomedical Engineering. 2026 Feb; 10(2): 245-58. doi: 10.1038/s41551-025-01463-z.

[17] Cheng L, Zheng X, Jiang SJ, Hu Y, Liu Y, Yang K *et al.* Sequence Display Enables Large-Scale Sequence–Activity Datasets for Rapid Protein Evolution. Nature Biotechnology. 2026 Apr: 1-5. doi: 10.1038/s41587-026-03087-3.

[18] Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. New England Journal of Medicine. 2021 Jan; 384(3): 252-60. doi: 10.1056/NEJMoa2031054.

[19] Wittmann BJ, Alexanian T, Bartling C, Beal J, Clore A, Diggans J *et al.* Strengthening Nucleic Acid

DOI: https://doi.org/10.54393/fbt.v6i2.245

Biosecurity Screening Against Generative Protein Design Tools. Science. 2025 Oct; 390(6768): 82-7. doi: 10.1126/science.adu8578.

[20] Wheeler NE. Responsible AI in Biotechnology: Balancing Discovery, Innovation and Biosecurity Risks. Frontiers in Bioengineering and Biotechnology. 2025 Feb; 13: 1537471. doi: 10.3389/fbioe.2025.1537471.

[21] Ruffolo JA, Chu LS, Mahajan SP, Gray JJ. Fast, Accurate Antibody Structure Prediction from Deep Learning on Massive Set of Natural Antibodies. Nature Communications. 2023 Apr; 14(1): 2389. doi: 10.1038/s41467-023-38063-x.

[22] Nidhi S, Anand U, Oleksak P, Tripathi P, Lal JA, Thomas G *et al.* Novel CRISPR–Cas Systems: An Updated Review of the Current Achievements, Applications, and Future Research Perspectives. International Journal of Molecular Sciences. 2021 Mar; 22(7): 3327. doi: 10.3390/ijms22073327.

[23] Koonin EV, Gootenberg JS, Abudayyeh OO. Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. Biochemistry. 2023 May; 62(24): 3465. doi: 10.1021/acs.biochem.3c00159.

[24] Chatterjee P, Lee J, Nip L, Koseki SR, Tysinger E, Sontheimer EJ *et al.* A Cas9 with PAM Recognition for Adenine Dinucleotides. Nature Communications. 2020 May; 11(1): 2474. doi: 10.1038/s41467-020-16117-8.

[25] Alkhnbashi OS, Meier T, Mitrofanov A, Backofen R, Voss B. CRISPR-Cas Bioinformatics. Methods. 2020 Feb; 172: 3-11. doi: 10.1016/j.ymeth.2019.07.013.

[26] Edison R, Toner S, Esvelt KM. Evaluating the Robustness of Current Nucleic Acid Synthesis Screening. Preprint. 2024.

[27] Ferdosi SR, Ewaisha R, Moghadam F, Krishna S, Park JG, Ebrahimkhani MR *et al.* Multifunctional CRISPR-Cas9 with Engineered Immunosilenced Human T Cell Epitopes. Nature Communications. 2019 Apr; 10(1): 1842. doi: 10.1038/s41467-019-09693-x.

[28] Skopintsev P, Esain-Garcia I, DeTurk EC, Yoon PH, Zhou Z, Weiss T *et al.* Structure and Evolution-Guided Design of Minimal RNA-Guided Nucleases. Science. 2026 Jul; 393(6808): 313-8. doi: 10.1126/science.aed6123.

[29] Silverstein RA, Kim N, Kroell AS, Walton RT, Delano J, Butcher RM *et al.* Custom CRISPR–Cas9 PAM Variants Via Scalable Engineering and Machine Learning. Nature. 2025 Jul; 643(8071): 539-50. doi: 10.1038/s41586-025-09021-y.

[30] Walton RT, Christie KA, Whittaker MN, Kleinstiver BP. Unconstrained Genome Targeting with Near-PAMless Engineered CRISPR-Cas9 Variants. Science. 2020 Apr; 368(6488): 290-6. doi: 10.1126/science.aba8853.

[31] Simhadri VL, McGill J, McMahon S, Wang J, Jiang H, Sauna ZE. Prevalence of Pre-Existing Antibodies to CRISPR-Associated Nuclease Cas9 in the USA Population. Molecular Therapy Methods and Clinical Development. 2018 Sep; 10: 105-12. doi: 10.1016/j.omtm.2018.06.006.

[32] Wang D, Tan Z, Gao J, Zhang S, Shen J, Lu Y. AI4Protein: Transforming the Future of Protein Design. Science China Life Sciences. 2025 Oct; 68(10): 2880-90. doi: 10.1007/s11427-024-2906-3.

[33] Fang T, Bogensperger L, Feer L, Allam A, Bezshapkin V, Balázs Z *et al.* Uncovering Cas9 PAM Diversity Through Metagenomic Mining and Machine Learning. Nature Communications. 2026 Feb; 17(1): 2510. doi: 10.1038/s41467-026-69098-5.

[34] Karimian A, Azizian K, Parsian H, Rafieian S, Shafiei-Irannejad V, Kheyrollah M *et al.* CRISPR/Cas9 Technology as a Potent Molecular Tool for Gene Therapy. Journal of Cellular Physiology. 2019 Aug; 234(8): 12267-77. doi: 10.1002/jcp.27972.

[35] She L, Liang Z, Zou Q, Huo YX. Engineering DNA-Targeting CRISPR and CRISPR-Like Effectors: Advances from Rational Design and High-Throughput Screening to AI-Driven Development. ACS Synthetic Biology. 2026. doi: 10.1021/acssynbio.6c00386.