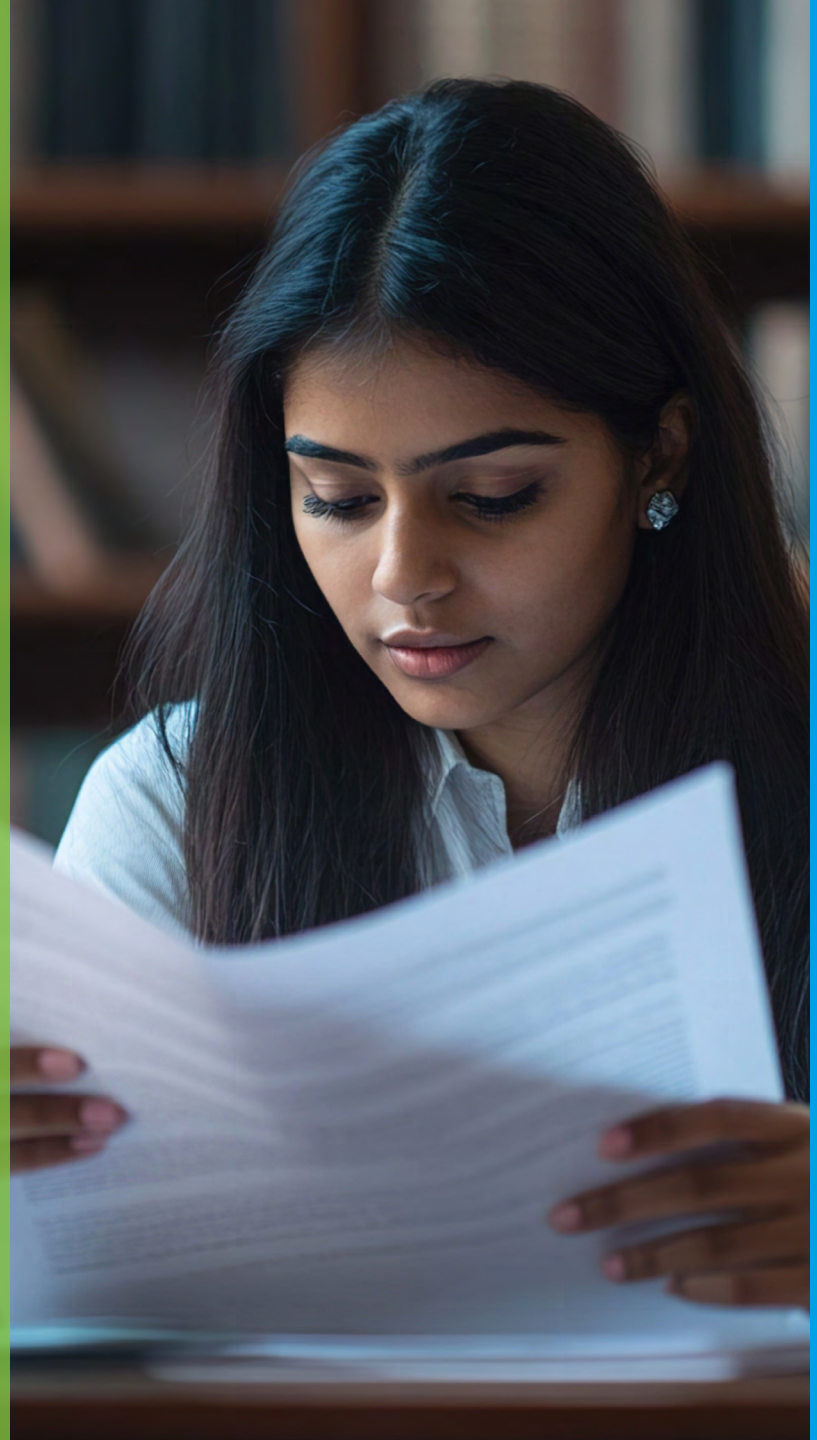
Oligo Pools in Pioneering RNA and Protein Engineering Research

Application and publication compendium

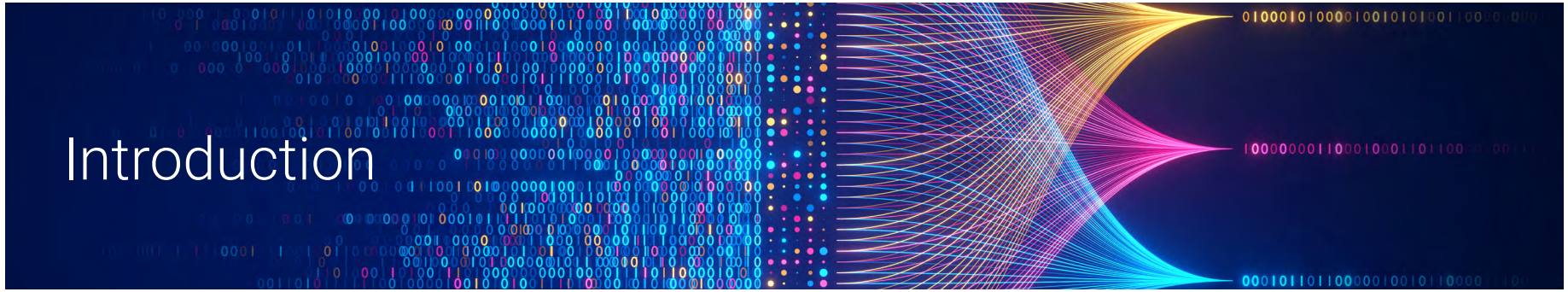


Agilent

Trusted Answers

Contents

Introduction



The 2024 Nobel Prize in Chemistry was recently awarded for pioneering work in computational protein design and protein structure prediction. This research has enabled scientists to predict three-dimensional protein structure and folding with remarkable precision and has opened the possibility for design of novel proteins with specific, unique functions.

Using advanced software tools, researchers can design thousands of small proteins (sometimes called ‘minibinders’), typically around 50-65 amino acids in length. An important part of converting the predicted sequences into actual proteins starts with encoding the amino acid sequences of the proteins as oligonucleotide pools.

Oligonucleotide pools are also a key tool used by researchers modeling complex RNA structure and dynamics. Here, atomic-scale nucleotide interactions are studied in an effort to better understand the unique 3D conformations of different RNA sequences.

In this compendium, we have selected articles that showcase how researchers are taking protein and RNA engineering from the digital world into the laboratory. Lastly, articles using or describing how oligo pools are used in functional genomics and gene assembly are provided.

A common thread for all references is that they’ve used Agilent’s Oligonucleotide Library Synthesis platform to synthesize the oligonucleotide sequences encoding the designed proteins or RNA structures. Agilent’s high-quality oligo pools provide cost effective access to enormous sequence diversity with exceptional uniformity and fidelity.

Spotlight on Oligo Pools and Protein Engineering

Nobel Prize Winner Dr. David Baker's focus on engineering minibinder proteins and their applications

Computational approaches to protein design

A computational approach to protein design provides a rational, more direct and faster approach to generating protein-binders than large scale, random library selection methods. When trying to explain the difference between old and the new computational approaches, one article[1] uses an analogy of a climber having to ascend a wall with just a few of the most reliable footholds or handholds which happen to be very far away from each other. This approach might fail in finding a way to the top because there is no easy or obvious way to connect the reliable footholds and handholds together into a successful route.

The new approach involves identifying and including every potential handhold and foothold, regardless of their quality. Next, numerous climbers attempt to ascend the wall using various routes or combinations of these holds. The most promising routes are then identified, and subsequently, a different group of climbers examines these routes thoroughly.

From analogy to algorithms

This research is advancing quickly, and while it is possible to design high-affinity binding proteins using only the structural information of the target, the success rate has been low.[2] Recent developments in machine learning and artificial intelligence have augmented the design and prediction processes though. Using data from AlphaFold2 or RoseTTAFold, researchers saw a 10-fold increase in design success rate when combining probabilities of monomer structure sequence adoption with the probability of the structure binding to the target as designed.[2]

From the digital world to the laboratory

Taking the design out of the computer and into the laboratory to confirm that the predicted 'minibinders' have the desired functionality is a big task. In one study, they experimentally tested ~20,000 designs per target,[2] and another study reports up to 100,000 designs being tested per target.[1]

Each one of these starts with an oligo pool manufactured by Agilent on the SurePrint Oligo Library Synthesis platform. Baker lab protein design software generates DNA sequence files that can be directly uploaded to the SurePrint platform for the rapid production of completely custom oligo pools. As Baker Lab member Dr. Brian Coventry once said, "We have ordered enough DNA sequences for at least two million proteins from Agilent. They are a lot less expensive than people would think."[3]

Applications of minibinders

Minibinders have a wide range of applications due to their high selectivity, stability, and ease of production. These proteins can be tailored to fit their targets much more tightly than traditional monoclonal antibodies. The list of applications of minibinders has grown rapidly--here are a few examples:

- Neutralization of the SARS-CoV-2 spike protein, showing promise as antiviral agents.[4]
- Targeting of pattern recognition receptors for improving the performance of vaccines.[5]
- Neutralization of toxins from botulinum toxin and *Clostridioides difficile* toxin B.[6]
- Used as scaffolds for drug development, providing a platform for creating new therapeutics with high specificity and stability.[7]

Agilent's contribution to the success of protein engineering

Agilent's SurePrint oligo pools deliver cost-effective sequence diversity allowing researchers to screen thousands of sequence variants and perform low-cost gene assembly processes.

- **Length:** Minibinders typically require DNA sequences encoding for around 65 amino acids, which translates to 195 nucleotides. Additional sequence content for assembly and amplification bring the total length to 230 nucleotides.
- **High fidelity:** Agilent is renowned for its ability to produce long oligonucleotides with low error rates. This precision is crucial for research that demands accurate and reliable starting material for complex protein designs.
- **Customization and quality:** Agilent's SurePrint platform provides high-quality completely custom oligonucleotide libraries with no restrictions on sequence content.
- **Exceptional uniformity:** High-throughput functional genomics screening and protein engineering with fewer false positive or false negatives from over or under-represented sequences.
- **Efficiency and scalability:** Rapid turnaround times enabled by our high capacity continuously operating production facility keep your research on track and at scale.

The utilization of Agilent oligo pools in Nobel Prize-winning research highlights the importance of high-quality, custom oligonucleotide libraries in scientific investigations.

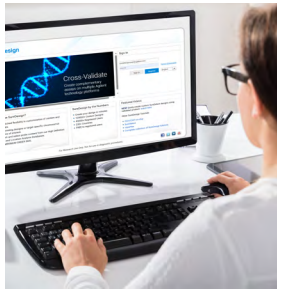
For researchers in protein, antibody, and genome engineering, Agilent's SurePrint oligo pools offer a valuable resource for achieving high fidelity, exceptional uniformity, and rapid turnaround times in all of their scientific endeavors.

Nobel Prize 2024, Chemistry laureates John Jumper, David Baker and Demis Hassabis. © Nobel Prize Outreach. Photo: Clément Morin.

Collection of articles

Explore a collection of articles from Nature highlighting the groundbreaking research and contributions of the awardees and their significant advancements in the field of protein design and engineering.

Read more



SureDesign

The SureDesign custom design tool creates custom designs for NGS, CGH, CRISPR, and FISH.

Learn more

SurePrint HiFi Oligo Pools

High Fidelity Oligo pools made for applications requiring the highest quality starting material.

Learn more

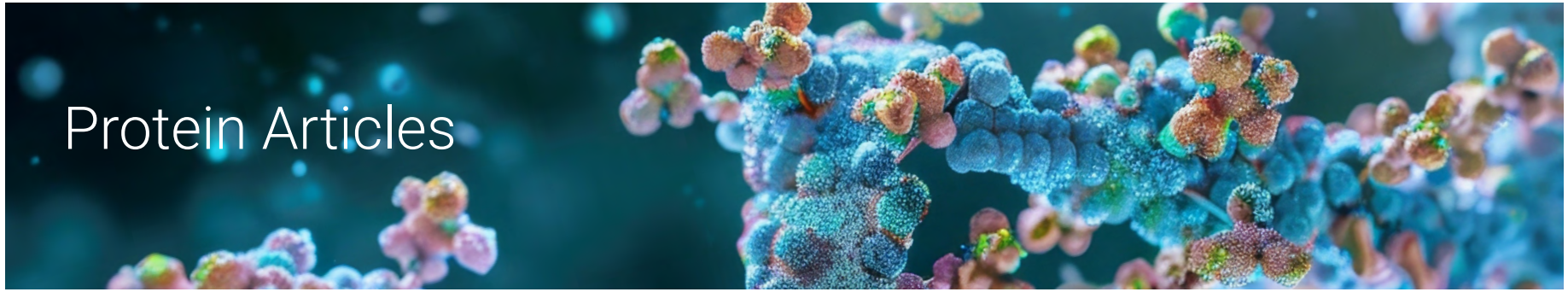
ProteoAnalyzer System

Automated SDS capillary electrophoresis system separates 12 protein samples in parallel.

Learn more

References

1. Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Park, J. S.; Jude, K. M.; Marković, I.; Kadam, R. U.; Verschueren, K. H. G.; Verstraete, K.; Walsh, S. T. R.; Bennett, N.; Phal, A.; Yang, A.; Kozodoy, L.; DeWitt, M.; Picton, L.; Miller, L.; Strauch, E.-M.; DeBouver, N. D.; Pires, A.; Bera, A. K.; Halabiya, S.; Hammerson, B.; Yang, W.; Bernard, S.; Stewart, L.; Wilson, I. A.; Ruohola-Baker, H.; Schlessinger, J.; Lee, S.; Savvides, S. N.; Garcia, K. C.; Baker, D. Design of Protein Binding Proteins from Target Structure Alone. *Nature* **2022**, *605* (7910), 551-560. DOI: 10.1038/s41586-022-04654-9
2. Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; DiMaio, F.; De Munck, S.; Savvides, S. N.; Baker, D. Improving de Novo Protein Binder Design with Deep Learning. *Nat. Commun.* **2023**, *14* (1), 38328. DOI: 10.1038/s41467-023-38328-5.
3. Coventry, B.; Baker, D. A New Way to Design Proteins. *Agilent Case Study* **2023**, 5994-5067EN. https://www.agilent.com/cs/library/casestudies/public/casestudy-new-way-design-proteins-5994-5067en-agilent.pdf. Last accessed 2025-04-11
4. Cao, L.; Goreshnik, I.; Coventry, B.; Case, J. B.; Miller, L.; Kozodoy, L.; Chen, R. E.; Carter, L.; Walls, A. C.; Park, Y.-J.; Strauch, E.-M.; Stewart, L.; Diamond, M. S.; Veesler, D.; Baker, D. De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors. *Science* **2020**, *370* (6515), 426-431. DOI: 10.1126/science.abd9909.
5. Adams, C. S.; Kim, H.; Burtner, A. E.; Lee, D. S.; Dobbins, C.; Criswell, C.; Coventry, B.; Tran-Pearson, A.; Kim, H. M.; King, N. P. De Novo Design of Protein Minibinder Agonists of TLR3. *Nat. Commun.* **2025**, *16* (1), 1234. DOI: 10.1038/s41467-025-56369-w.
6. Lv, X.; Zhang, Y.; Sun, K.; Yang, Q.; Luo, J.; Tao, L.; Lu, P. De Novo Design of Mini-Protein Binders Broadly Neutralizing Clostridioides difficile Toxin B Variants. *Nat. Commun.* **2024**, *15* (1), 52582. DOI: 10.1038/s41467-024-52582-1.
7. Figueiredo, M. L. Minibinders: Tiny Proteins with Huge Potential in Bioengineering. *Springer Nature Research Communities* **2025**. https://communities.springernature.com/posts/minibinders-tiny-proteins-with-huge-potential-in-bioengineering. Last accessed 2025-04-11



Protein Articles

Open Access

Design of Protein-Binding Proteins from the Target Structure Alone

Authors

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M. Jude, Iva Marković, Rameshwar U. Kadam, Koen H. G. Verschueren, Kenneth Verstraete, Scott Thomas Russell Walsh, Nathaniel Bennett, Ashish Phal, Aerin Yang, Lisa Kozodoy, Michelle DeWitt, Lora Picton, Lauren Miller, Eva-Maria Strauch, Nicholas D. DeBouver, Allison Pires, Asim K. Bera, Samer Halabiya, Bradley Hammerson, Wei Yang, Steffen Bernard, Lance Stewart, Ian A. Wilson, Hannele Ruohola-Baker, Joseph Schlessinger, Sangwon Lee, Savvas N. Savvides, K. Christopher Garcia and David Baker.

Journal: Nature

Year: 2022

DOI: doi.org/10.1038/s41586-022-04654-9

Affiliation: University of Washington, Seattle, USA

Abstract

The design of proteins that bind to a specific site on the surface of a target protein using no information other than the three-dimensional structure of the target remains a challenge1-5. Here we describe a general solution to this problem that starts with a broad exploration of the vast space of possible binding modes to a selected region of a protein surface, and then intensifies the search in the vicinity of the most promising binding modes. We demonstrate the broad applicability of this approach through the de novo design of binding proteins to 12 diverse protein targets with different shapes and surface properties. Biophysical characterization shows that the binders, which are all smaller than 65 amino acids, are hyperstable and, following experimental optimization, bind their targets with nanomolar to picomolar affinities.

We succeeded in solving crystal structures of five of the binder–target complexes, and all five closely match the corresponding computational design models. Experimental data on nearly half a million computational designs and hundreds of thousands of point mutants provide detailed feedback on the strengths and limitations of the method and of our current understanding of protein–protein interactions, and should guide improvements of both. Our approach enables the targeted design of binders to sites of interest on a wide variety of proteins for therapeutic and diagnostic applications.

Cao, L.; Coventry, B.; Goreshnik, I.; et al. Design of Protein-Binding Proteins from the Target Structure Alone. *Nature* **2022**, *605*, 551–560. DOI: 10.1038/s41586-022-04654-9.

Open Access

Improving De Novo Protein Binder Design with Deep Learning

Nathaniel R. Bennett, Brian Coventry, Inna Goreshnik, Buwei Huang, Aza Allen, Dionne Vafeados, Ying Po Peng, Justas Dauparas, Minkyung Baek, Lance Stewart, Frank DiMaio, Steven De Munck, Savvas N. Savvides and David Baker.

Journal: Nature Communications

Year: 2023

DOI: 10.1038/s41467-023-38328-5

Affiliation: University of Washington, Seattle, USA

Abstract

Recently it has become possible to de novo design high affinity protein binding proteins from target structural information alone. There is, however, considerable room for improvement as the overall design success rate is low. Here, we explore the augmentation of energy-based protein binder design using deep learning. We find that using AlphaFold2 or RoseTTAFold to assess the probability that a designed sequence adopts the designed monomer structure, and the probability that this structure binds the target as designed, increases design success rates nearly 10-fold. We find further that sequence design using ProteinMPNN rather than Rosetta considerably increases computational efficiency.

Bennett, N.R.; et al. Improving de novo protein binder design with deep learning. *Nat. Commun.* **2023**, *14*, 2625. DOI: 10.1038/s41467-023-38328-5.

Open Access

De Novo Design of Protein Minibinder Agonists of TLR3

Authors

Chloe S. Adams, Hyojin Kim, Abigail E. Burtner, Dong Sun Lee, Craig Dobbins, Cameron Criswell, Brian Coventry, Adri Tran-Pearson, Ho Min Kim and Neil P. King

Journal: Nature Communications

Year: 2025

DOI: 10.1038/s41467-025-56369-w

Affiliation: University of Washington, Seattle, USA

Abstract

Toll-like Receptor 3 (TLR3) is a pattern recognition receptor that initiates antiviral immune responses upon binding double-stranded RNA (dsRNA). Several nucleic acid-based TLR3 agonists have been explored clinically as vaccine adjuvants in cancer and infectious disease, but present substantial manufacturing and formulation challenges. Here, we use computational protein design to create novel miniproteins that bind to human TLR3 with nanomolar affinities. Cryo-EM structures of two minibinders in complex with TLR3 reveal that they bind the target as designed, although one partially unfolds due to steric competition with a nearby N-linked glycan. Multivalent forms of both minibinders induce NF-κB signaling in TLR3-expressing cell lines, demonstrating that they may have therapeutically relevant biological activity. Our work provides a foundation for the development of specific, stable, and easy-to-formulate protein-based agonists of TLRs and other pattern recognition receptors.

Adams, C.S.; et al. De novo design of protein minibinder agonists of TLR3. *Nat. Commun.* **2025**, *16*, 1234. DOI: 10.1038/s41467-025-56369-w

Open Access

De Novo Designed Hsp70 Activator Dissolves Intracellular Condensates

Authors

Jason Z. Zhang, Nathan Greenwood, Jason Hernandez, Josh T. Cuperus, Buwei Huang, Bryan D. Ryder, Christine Queitsch, Jason E. Gestwicki, and David Baker

Journal: Cell Chemical Biology

Year: 2025

DOI: 10.1016/j.chembiol.2025.01.006

Affiliation: University of Washington, Seattle, USA

Abstract

Cells constantly use and recycle their proteins in a process called protein quality control (PQC). Imbalance of PQC leads to accumulation of dysfunctional proteins, thus leading to a variety of diseases. Hsp70 is a central mediator of PQC by collaborating with J-domain proteins (JDPs) to regulate client protein folding and degradation. Despite the importance of this system, the molecular mechanisms governing these interactions remain unclear. This study employs computationally designed proteins to create synthetic Hsp70-binding proteins that can either enhance or inhibit its activity. One of these designs mimic native JDPs in promoting refolding of denatured proteins and can modulate intracellular condensates, revealing insights into the roles of these condensates. This work provides insight in three aspects: (1) computational and experimental methods to design synthetic binding proteins, (2) advances our understanding of Hsp70 interactions that regulate PQC, (3) introduces modular tools to manipulate Hsp70 activity and condensates. We anticipate that these results and tools will accelerate our understanding of PQC and manipulation of PQC in diseases.

Zhang, J. Z.; et al. De Novo Designed Hsp70 Activator Dissolves Intracellular Condensates. *Cell Chem. Biol.* **2024**, *32* (3), 463-473.e6. DOI: 10.1016/j.chembiol.2024.02.015

Open Access

Designed Miniproteins Potently Inhibit and Protect Against MERS-CoV

Authors

Robert J. Ragotte, M. Alejandra Tortorici, Nicholas J. Catanzaro, Amin Addetia, Brian Coventry, Heather M. Froggatt, Jimin Lee, Cameron Stewart, Jack T. Brown, Inna Goreshnik, Jeremiah N. Sims, Lukas F. Milles, Basile I.M. Wicky, Matthias Glögl, Stacey Gerben, Alex Kang, Asim K. Bera, William Sharkey, Alexandra Schäfer, Ralph S. Baric, David Baker, and David Veesler

Journal: BioRxiv

Year: 2024

DOI: 10.1101/2024.11.03.621760

Affiliation: University of Washington, Seattle, USA

Abstract

Middle-East respiratory syndrome coronavirus (MERS-CoV) is a zoonotic pathogen with 36% case-fatality rate in humans. No vaccines or specific therapeutics are currently approved to use in humans or the camel host reservoir. Here, we computationally designed monomeric and homo-oligomeric miniproteins binding with high affinity to the MERS-CoV spike (S) glycoprotein, the main target of neutralizing antibodies and vaccine development. We show that these miniproteins broadly neutralize a panel of MERS-CoV S variants, spanning the known antigenic diversity of this pathogen, by targeting a conserved site in the receptor-binding domain (RBD). The miniproteins directly compete with binding of the DPP4 receptor to MERS-CoV S, thereby blocking viral attachment to the host entry receptor and subsequent membrane fusion. Intranasal administration of a lead miniprotein provides prophylactic protection against stringent MERS-CoV challenge in mice motivating future clinical development as a next-generation countermeasure against this virus with pandemic potential.

Ragotte, R.J.; et al. Designed miniproteins potently inhibit and protect against MERS-CoV. *bioRxiv* **2024**, Nov 4, 2024.11.03.621760. DOI: 10.1101/2024.11.03.621760

Open Access

De Novo Design of Mini-Protein Binders Broadly Neutralizing *Clostridioides difficile* Toxin B Variants

Authors

Xinchen Lv, Yuanyuan Zhang, Ke Sun, Qi Yang, Jianhua Luo, Liang Tao and Peilong Lu

Journal: Nature Communications

Year: 2024

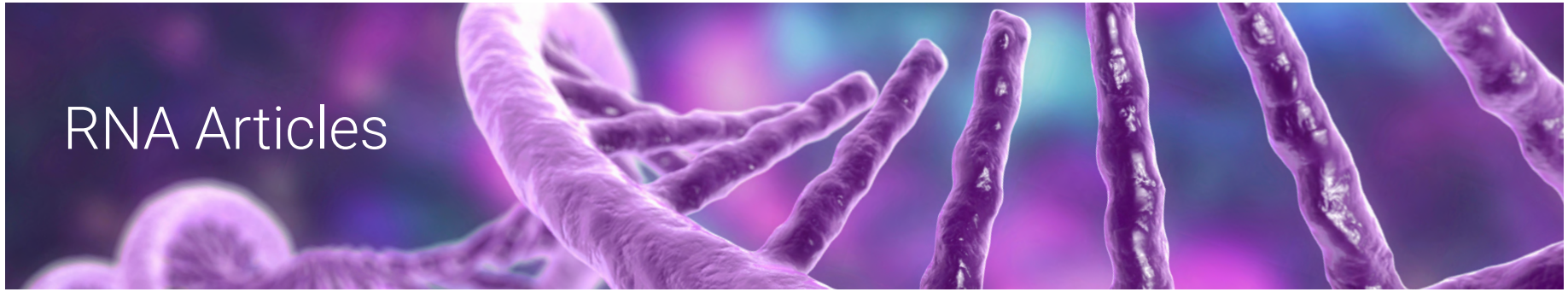
DOI: 10.1038/s41467-024-52582-1

Affiliation: Westlake University, Hangzhou, China

Abstract

Clostridioides difficile toxin B (TcdB) is the key virulence factor accounting for *C. difficile* infection-associated symptoms. Effectively neutralizing different TcdB variants with a universal solution poses a significant challenge. Here we present the de novo design and characterization of pan-specific mini-protein binders against major TcdB subtypes. Our design successfully binds to the first receptor binding interface (RBI-1) of the varied TcdB subtypes, exhibiting affinities ranging from 20 pM to 10 nM. The cryo-electron microscopy (cryo-EM) structures of the mini protein binder in complex with TcdB1 and TcdB4 are consistent with the computational design models. The engineered and evolved variants of the mini-protein binder and chondroitin sulfate proteoglycan 4 (CSPG4), another natural receptor that binds to the second RBI (RBI-2) of TcdB, better neutralize major TcdB variants both in cells and in vivo, as demonstrated by the colon-loop assay using female mice. Our findings provide valuable starting points for the development of therapeutics targeting *C. difficile* infections (CDI).

Lv, X.; et al. De Novo Design of Mini-Protein Binders Broadly Neutralizing Clostridioides Difficile Toxin B Variants. *Nat. Commun.* **2024**, *15*, 8521. DOI: 10.1038/s41467-024-52582-1



RNA Articles

Preprint

Characterizing 3D RNA Structural Features from DMS Reactivity

Authors

D. H. Sanduni Deenalattha, Chris P. Jurich, Bret Lange, Darren Armstrong, Kaitlyn Nein, and Joseph D. Yesselman

Journal: BioRxiv

Year: 2024

DOI: 10.1101/2024.11.21.624766

Affiliation: University of Nebraska, Lincoln, USA

Abstract

Dimethyl sulfate (DMS) chemical mapping probes RNA structure, where low reactivity is generally interpreted as Watson-Crick (WC) base pairs and high reactivity as unpaired nucleotides. Studies examining DMS reactivity of RNAs with known 3D structures have identified nucleotides that deviate from this interpretation with distinct solvent accessibility and hydrogen bonding patterns. Understanding the frequency of these outliers and their recurring structural 3D features remains incomplete. To address this knowledge gap, we systematically analyzed DMS reactivity patterns across a library of 7,500 RNA constructs containing two-way junctions with known 3D structures. We observe DMS reactivity exists on a continuum over four orders of magnitude with approximately 10% overlap in reactivity between WC and non-WC nucleotides. We find that non-WC bases with WC-like DMS protection exhibit increased hydrogen bonding and decreased solvent accessibility, whereas WC pairs exhibiting greater DMS reactivity tend to flank junctions, correlating with weaker base stacking and greater junction dynamics. Furthermore, we discover that DMS reactivity values in non-canonical pairs correlate with atomic distances and base pair geometry, enabling discrimination between different 3D conformations. These DMS reactivity patterns indicate that DMS reactivity provides atomic-scale information about RNA 3D conformations, which can be used to model RNA structures and dynamics.

Deenalattha, D.H.S.; et al. Characterizing 3D RNA Structural Features from DMS Reactivity.
bioRxiv **2024**, DOI: 10.1101/2024.11.21.624766

Open Access

Large-Scale Analysis of Small Molecule-RNA Interactions Using Multiplexed RNA Structure Libraries

Authors

Ryosuke Nagasawa, Kazumitsu Onizuka, Kaoru R. Komatsu, Emi Miyashita, Hirotaka Murase, Kanna Ojima, Shunya Ishikawa, Mamiko Ozawa, Hirohide Saito and Fumi Nagatsugi

Journal: Communications Chemistry

Year: 2024

DOI: 10.1038/s42004-024-01181-8

Affiliation: Tohoku University, Japan

Abstract

The large-scale analysis of small-molecule binding to diverse RNA structures is key to understanding the required interaction properties and selectivity for developing RNA-binding molecules toward RNA-targeted therapies. Here, we report a new system for performing the large-scale analysis of small molecule–RNA interactions using a multiplexed pull-down assay with RNA structure libraries. The system profiled the RNA-binding landscapes of G-clamp and thiazole orange derivatives, which recognizes an unpaired guanine base and are good probes for fluorescent indicator displacement (FID) assays, respectively.

We discuss the binding preferences of these molecules based on their large-scale affinity profiles. In addition, we selected combinations of fluorescent indicators and different ranks of RNA based on the information and screened for RNA-binding molecules using FID. RNAs with high- and intermediate-rank RNA provided reliable results. Our system provides fundamental information about small molecule–RNA interactions and facilitates the discovery of novel RNA-binding molecules.

Nagasawa, R.; et al. Large-Scale Analysis of Small Molecule-RNA Interactions Using Multiplexed RNA Structure Libraries. *Commun. Chem.* **2024**, *7*, 98. DOI: 10.1038/s42004-024-01181-8.

Open Access

Compact CRISPR Genetic Screens Enabled by Improved Guide RNA Library Cloning

Authors

Seok-Jin Heo, Lauren D Enriquez, Scot Federman, Amy Y Chang, Rachel Mace, Kaivalya Shevade, Phuong Nguyen, Adam J Litterman, Shawn Shafer, Laralynne Przybyla, and Eric D Chow

Journal: Genome Biology

Year: 2024

DOI: 10.1186/s13059-023-03132-3

Affiliation: University of California, San Francisco, USA

Abstract

CRISPR genome editing approaches theoretically enable researchers to define the function of each human gene in specific cell types, but challenges remain to efficiently perform genetic perturbations in relevant models. In this work, we develop a library cloning protocol that increases sgRNA uniformity and greatly reduces bias in existing genome-wide libraries.

We demonstrate that our libraries can achieve equivalent or better statistical power compared to previously reported screens using an order of magnitude fewer cells. This improved cloning protocol enables genome-scale CRISPR screens in technically challenging cell models and screen formats.

Heo, S. J.; et al. Compact CRISPR Genetic Screens Enabled by Improved Guide RNA Library Cloning. *Genome Biol.* **2024**, *25* (1), 25. DOI: 10.1186/s13059-023-03132-3.

Open Access

Future Directions for High-Throughput Splicing Assays in Precision Medicine

Authors

Christy L Rhine, Christopher Neil, David T Glidden, Kamil J Cygan, Alger M Fredericks, Jing Wang, Nephi A Walton, William G Fairbrother

Journal: Human Mutations

Year: 2019

DOI: 10.1002/humu.23866

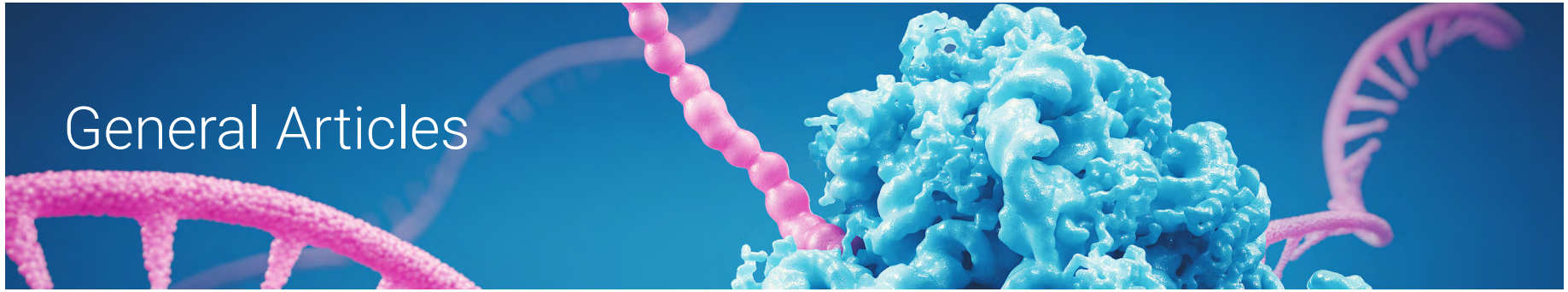
Affiliation: Brown University, Providence, USA

Abstract

Classification of variants of unknown significance is a challenging technical problem in clinical genetics. As up to one third of disease-causing mutations are thought to affect pre-mRNA splicing, it is important to accurately classify splicing mutations in patient sequencing data. Several consortia and healthcare systems have conducted large-scale patient sequencing studies, which discover novel variants faster than they can be classified. Here we compare the advantages and limitations of several high-throughput splicing assays aimed at mitigating this bottleneck, and describe a dataset of ~5000 variants that we analyzed using our Massively Parallel Splicing Assay (MaPSy).

The Critical Assessment of Genome Interpretation group (CAGI) organized a challenge, in which participants submitted machine learning models to predict the splicing effects of variants in this dataset. We discuss the winning submission of the challenge (MMSplice) which outperformed existing software. Finally, we highlight methods to overcome the limitations of MaPSy and similar assays, such as tissue-specific splicing, the effect of surrounding sequence context, classifying intronic variants, synthesizing large exons, and amplifying complex libraries of minigene species. Further development of these assays will greatly benefit the field of clinical genetics, which lack high-throughput methods for variant interpretation.

Rhine, C.L; et al. Future Directions for High-Throughput Splicing Assays in Precision Medicine. *Hum. Mutat*. **2019**, *40*(9), 1225-1234. DOI: 10.1002/humu.23866



General Articles

Open Access

Transcription Factor Interactions Explain the Context-Dependent Activity of CRX Binding Sites

Authors

Kaiser J. Loell, Ryan Z. Friedman, Connie A. Myers, Joseph C. Corbo, Barak A. Cohen, and Michael A. White

Journal: PLoS Computational Biology

Year: 2024

DOI: 10.1371/journal.pcbi.1011802

Affiliation: Washington University, St. Louis, USA

Abstract

The effects of transcription factor binding sites (TFBSs) on the activity of a cis-regulatory element (CRE) depend on the local sequence context. In rod photoreceptors, binding sites for the transcription factor (TF) Cone-rod homeobox (CRX) occur in both enhancers and silencers, but the sequence context that determines whether CRX binding sites contribute to activation or repression of transcription is not understood. To investigate the context-dependent activity of CRX sites, we fit neural network-based models to the activities of synthetic CREs composed of photoreceptor TFBSs. The models revealed that CRX binding sites consistently make positive, independent contributions to CRE activity, while negative homotypic interactions between sites cause CREs composed of multiple CRX sites to function as silencers. The effects of negative homotypic interactions can be overcome by the presence of other TFBSs that either interact cooperatively with CRX sites or make independent positive contributions to activity. The context-dependent activity of CRX sites is thus determined by the balance between positive heterotypic interactions, independent contributions of TFBSs, and negative homotypic interactions. Our findings explain observed patterns of activity among genomic CRX-bound enhancers and silencers, and suggest that enhancers may require diverse TFBSs to overcome negative homotypic interactions between TFBSs.

Loell, K. J.; et al. Transcription Factor Interactions Explain the Context-Dependent Activity of CRX Binding Sites. *PLoS Comput. Biol.* **2024**, *20*(1), e1011802. DOI:10.1371/journal.pcbi.1011802.

Open Access

User-Defined Single Pot Mutagenesis Using Unamplified Oligo Pools

Authors

Angélica V Medina-Cucurella, Paul J Steiner, Matthew S Faber, Jesús Beltrán, Alexandra N Borelli, Monica B Kirby, Sean R Cutler, and Timothy A Whitehead

Journal: Protein Engineering, Design and Selection

Year: 2019

DOI: 10.1093/protein/gzz013

Affiliation: Michigan State University, East Lansing, USA

Abstract

User-defined mutagenic libraries are fundamental for applied protein engineering workflows. Here we show that unamplified oligo pools can be used to prepare site saturation mutagenesis libraries from plasmid DNA with near-complete coverage of desired mutations and few off-target mutations. We find that oligo pools yield higher quality libraries when compared to individually synthesized degenerate oligos. We also show that multiple libraries can be multiplexed into a single oligo pool, making preparation of multiple libraries less expensive and more convenient. We provide software for automatic oligo pool design that can generate mutagenic oligos for saturating or focused libraries.

Medina-Cucurella, A. V.; et al. User-Defined Single Pot Mutagenesis Using Unamplified Oligo Pools. *Protein Eng. Des. Sel.* **2019**, *32* (1), 41-45. DOI: 10.1093/protein/gzz013.

Open Access

Oligo Pools as an Affordable Souce of Synthetic DNA for Cost-Effective Library Construction in Protein- and Metabolic Pathway Engineering

Authors

Bastiaan P Kuiper, Rianne C Prins, and Sonja Billerbeck

Journal: ChemBioChem

Year: 2021

DOI: 10.1002/cbic.202100507

Affiliation: University of Groningen, Groningen, NL

Abstract

The construction of custom libraries is critical for rational protein engineering and directed evolution. Array-synthesized oligo pools of thousands of user-defined sequences (up to ~350 bases in length) have emerged as a low-cost commercially available source of DNA. These pools cost ≤10 % (depending on error rate and length) of other commercial sources of custom DNA, and this significant cost difference can determine whether an enzyme engineering project can be realized on a given research budget. However, while being cheap, oligo pools do suffer from a low concentration of individual oligos and relatively high error rates.

Several powerful techniques that specifically make use of oligo pools have been developed and proven valuable or even essential for next-generation protein and pathway engineering strategies, such as sequence-function mapping, enzyme minimization, or de-novo design. Here we consolidate the knowledge on these techniques and their applications to facilitate the use of oligo pools within the protein engineering community.

Kuiper, B.P.; et al. Oligo Pools as an Affordable Source of Synthetic DNA for Cost-Effective Library Construction in Protein- and Metabolic Pathway Engineering. *Chembiochem* **2022**, *23*(7), e202100507. DOI: 10.1002/cbic.202100507

Subscription

Gene Assembly from Chip-synthesized Oligonucleotides

Authors

Nikolai Eroshenko, Sriram Kosuri, Adam H. Marblestone, Nicholas Conway, and George M. Church

Journal: Current Protocols in Chemical Biology

Year: 2012

DOI: 10.1002/9780470559277.ch110190

Affiliation: Harvard, Cambridge, USA

Abstract

De novo synthesis of long double-stranded DNA constructs has a myriad of applications in biology and biological engineering. However, its widespread adoption has been hindered by high costs. Cost can be significantly reduced by using oligonucleotides synthesized on high-density DNA chips. However, most methods for using off-chip DNA for gene synthesis have failed to scale due to the high error rates, low yields, and high chemical complexity of the chip-synthesized oligonucleotides. We have recently demonstrated that some commercial DNA chip manufacturers have improved error rates, and that the issues of chemical complexity and low yields can be solved by using barcoded primers to accurately and efficiently amplify subpools of oligonucleotides. This unit includes protocols for computationally designing the DNA chip, amplifying the oligonucleotide subpools, and assembling 500- to 800-bp constructs.

Eroshenko, N; et al. Gene Assembly from Chip-Synthesized Oligonucleotides. *Curr. Protoc. Chem. Biol.* **2012**, *4*, 1-17. DOI: 10.1002/9780470559277.ch110190

Further reading and resources

Customer spotlight

Precision Genome Editing Puts Cancer Mutations Under the Spotlight

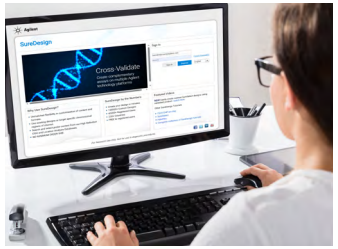
Case study

University of Washington's David Baker Lab Develops a New Way to Design Proteins

Flyer

SurePrint Oligonucleotide Libraries

Product details



SureDesign

The SureDesign custom design tool creates custom designs for NGS, CGH, CRISPR, and FISH.

Product details

SurePrint HiFi Oligo Pools

High Fidelity Oligo pools made for applications requiring the highest quality starting material.

Product details

SurePrint Oligo Pools

Oligo pools ideal for accelerated protein engineering, directed evolution, gene assembly, pooled CRISPR screens and single cell profiling.

Learn more:
www.agilent.com/en/solutions/genomics-applications-solutions/ols

Get answers to your technical questions and access resources in the Agilent Community:
community.agilent.com

Contact us:
genome.engineering@agilent.com

U.S. and Canada
1-800-227-9770
agilent_inquiries@agilent.com

Europe
info_agilent@agilent.com

Asia Pacific
inquiry_lsca@agilent.com

Table of contents

PR7001-4704

Published in the USA, August 1, 2025
5994-8512EN

Agilent
Trusted Answers