Sketching Proteins with Bare Hands in VR

Donghyeok Ma
Department of Industrial Design, KAIST
donghyeok.ma@kaist.ac.kr

Joon Hyub Lee
Department of Industrial Design, KAIST
joonhyub.lee@kaist.ac.kr

Taegyu Jin
Department of Industrial Design, KAIST
taegyu.jin@kaist.ac.kr

Sang-Hyun Lee
Department of Industrial Design, KAIST
sang-hyun.lee@kaist.ac.kr

Ho Min Kim*
Graduate School of Medical Science and Engineering, KAIST
hm_kim@kaist.ac.kr

Seok-Hyung Bae
Department of Industrial Design, KAIST
seokhyung.bae@kaist.ac.kr

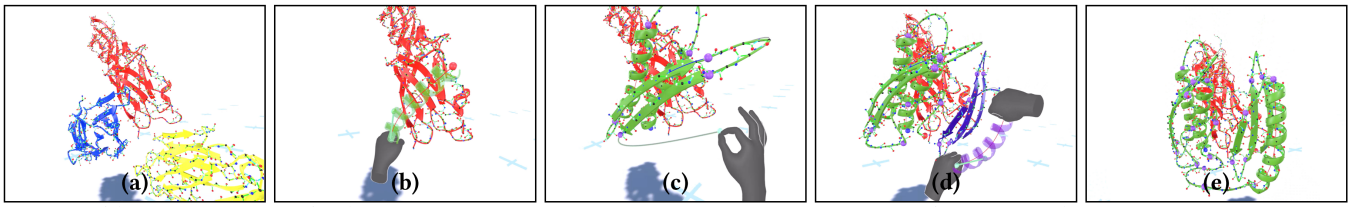

Figure 1: (a) A PD-L1 protein (red) on a cancer cell is known to bind to two PD-1 proteins (blue, yellow) on T cells and restrain the T cells' activation that is necessary for killing pathogens in an immune response [13]. (b-e) In our system, synthetic biologists designing a new protein that binds to PD-L1 in place of PD-1 to prevent this can explore a wide range of design options by intuitively authoring secondary structures that conform to the binding sites of PD-L1 with their bare hands in VR.

ABSTRACT

Recent developments in AI have made it possible to design new proteins that are crucial to meeting humanity's needs. However, tools for exploring the 3D structures of proteins in the early stages of AI-based protein design are lacking, leading to many preventable trials and errors and much wasted time and efforts in the design process. To address this, we propose a novel VR interaction system that enables synthetic biologists to intuitively author the 3D structures of proteins with their bare hands.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**.

KEYWORDS

Protein, sketch, bare hands, VR, AI-based protein design

ACM Reference Format:
Donghyeok Ma, Joon Hyub Lee, Taegyu Jin, Sang-Hyun Lee, Ho Min Kim, and Seok-Hyung Bae. 2023. Sketching Proteins with Bare Hands in VR. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), October 29–November 01, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3586182.3616675

*Also with the Center for Biomolecular & Cellular Structure, IBS, Republic of Korea.

UIST '23 Adjunct, October 29–November 01, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0096-5/23/10.
https://doi.org/10.1145/3586182.3616675

1 INTRODUCTION

Proteins, excluding water, are the most abundant components of the human body and act as bio-micromachines that take part in nearly all bodily functions such as respiration, digestion, movement, and immune response. Proteins accomplish these functions by binding with other proteins or chemicals, the occurrences and whereabouts of which are determined by the 3D structures of the proteins, which are in turn determined by the sequence of different amino acids that comprise them.

Each of the 20 types of amino acids found in the human body consists of a structural backbone that is identical across all types and a side chain that is unique to each type. When tens, hundreds, or even thousands of amino acids sequentially connect, the forces between the side chains and water molecules cause the sequence to fold, in turn causing the backbone molecules to locally form secondary structures (Figure 2), the architectural building blocks of the 3D structures of proteins [1].

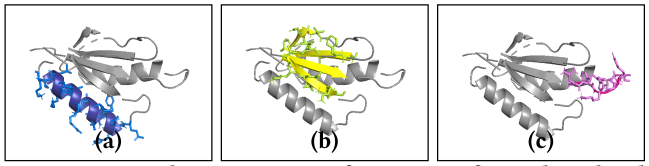

Figure 2: Secondary structures of proteins refer to three local patterns of folded amino acid sequences. (a) α-helices (blue) form due to forces between adjacent atoms in the backbone. (b) β-sheets (yellow) consist of an arrangement of nearby β-strands that form due to forces between non-adjacent atoms in the backbone. (c) Loops (pink) form due to the lack of such forces. These are commonly visualized as helical ribbons, arrow-shaped ribbons, and thin wires, respectively.

Synthetic biologists envision a future in which they can design and synthesize new proteins that are crucial to meeting humanity's needs, such as those that can cure cancer, fight against infectious diseases, and break down plastic wastes. Recent developments of AI-based techniques are pushing toward this future [5]: *AlphaFold* [10] and *RosettaFold* [2] quickly and rapidly predict 3D structures of proteins, given amino acid sequences; *ProteinMPNN* [7] predicts plausible amino acid sequences, given 3D structures of proteins; *RFDiffusion* [16] fills in missing gaps with suitable secondary structures, given incomplete 3D structures of proteins.

As with products, when developing new proteins, it is crucial to explore a wide range of options in the early stages to achieve desired outcomes with fewer trials and errors later on. In product design, "sketching" is a quintessential tool for rough, but quick and easy ideation during these stages. In this study, by extending *WireSketch* [12], a system of bimanual interactions for intuitively authoring 3D curve networks, we propose a novel VR interaction system for helping synthetic biologists explore 3D structures of proteins in the early stages of AI-based protein design (Figure 1).

2 EXPERT INTERVIEWS

To understand the needs of the target users, we observed and recorded the workflows of four synthetic biologists working on AI-based protein design and conducted structured interviews. Their research experience in protein and protein design averaged 6.25 years. We collected an average of 2.5 hours of data from each expert.

According to them, the current AI-based protein design process takes four steps. First, conceptualize 3D structures of new proteins that can bind to their targets. Second, predict amino acid sequences that best satisfy the desired structures and bindings [7, 10]. Third, fine-tune the amino acid sequences to maximize the strength and stability of the bindings [9]. Fourth, validate the results through DNA synthesis, protein expression, and structural analysis [6].

During the first step, however, exploring a wide range of design options for the 3D structures of proteins presents significant challenges: The irregular shapes of intricately intertwined secondary structures make it difficult to grasp their spatial relations through a 2D screen. In addition, authoring the 3D structures of proteins involves moving, rotating, lengthening, shortening, bending, twisting, cutting, and joining secondary structures in 3D space, which can be cumbersome with 2D input devices such as a mouse. Moreover, as not all manipulations are physicochemically valid, guesswork leads to many trials and errors. Finally, the lack of an integrated system necessitates frequent and distracting jumps between various specialized tools [3, 8, 11, 14].

3 INTERACTIVE SYSTEM

Inspired by the synthetic biologists waving their hands in midair and wielding whatever objects that lie around in a desperate attempt to express ideas regarding the 3D structures of new proteins, we propose a novel VR interaction system that enables quick and easy sketching of new proteins through intuitive bare hand interactions.

To achieve this, we extend *WireSketch* [12], a VR interaction system for authoring 3D curve networks during the early stages of product design. Because the intuitive bare hand interactions of the system that emulate wire crafting make it quick and easy to roughly create 3D curves, precisely edit them, and weave a network from them, we believe that it is an ideal platform upon which to implement functionalities needed for designing new proteins.

Extending *WireSketch*, our system supports the creation of three types of secondary structures through hand gestures that mimic holding physical objects (Figure 3a-c). It also automatically converts the 3D curve networks into physicochemically valid 3D structures of proteins to the best approximation (Figure 3d-f). Finally, our system exports the results in the standard PDB (Protein Data Bank) file format [4] to be used as input of AI-based tools [7, 10] for the subsequent stages of protein design.

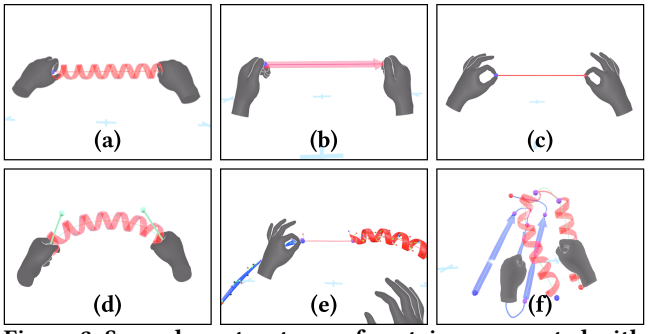

Figure 3: Secondary structures of proteins are created with midair hand gestures that resemble holding (a) a tube (α-helix), (b) a paper strip (β-strand), or (c) a wire (loop). With the interactions of *WireSketch* [12], users can freely (d) bend and twist, (e) join, and (f) arrange the secondary structures.

4 IMPLEMENTATION

Our system was implemented using the Unity 3D game engine and the Meta Quest Pro VR headset with the built-in hand tracking capability. We referred to the work of Priestle [15] to calculate physicochemically valid 3D structures of new proteins. To showcase the user scenario of our system, we used it to design a new protein that can potentially mitigate the adverse effects of cancer cells on the healthy functioning of the immune system [13] (Figure 1).

5 CONCLUSION & FUTURE WORK

Based on our expert interviews with synthetic biologists, we designed and implemented a novel VR interaction system for quickly and easily expressing design ideas on 3D structures and the bindings of new proteins with bare hands during the early stages of AI-based protein design. Our system enables the authoring of physicochemically valid 3D structures consisting of multiple secondary structures such as α-helices, β-strands, and loops, and exporting the design outcomes in the standard PDB file format [4]. In future work, we will present a novel workflow for AI-based protein design by integrating a generative AI, such as *RFDiffusion* [16], into our system. We will also conduct formal evaluations with synthetic biologists to validate the system's usefulness and usability.

ACKNOWLEDGMENTS

This research was supported by the KAIST Convergence Research Institute Operation Program and grants from the Institute for Basic Science (IBS-R030-C1 to H.M.K.). We thank Hyo Jin Kim, Hyun Joo Ro, Jin Sook Ahn, and Jeong Won Yun from IBS for the interviews.

REFERENCES

[1] Christian B. Anfinsen. 1973. Principles that govern the folding of protein chains. *Science* 181, 4096 (1973), 223–230. https://doi.org/10.1126/science.181.4096.223

[2] Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 6557 (2021), 871–876. https://doi.org/10.1126/science.abj8754

[3] Simon Bennie, Martina Maritan, Jonathon Gast, Marc Loschen, Daniel Gruffat, Roberta Bartolotta, Samuel Hessenauer, Edgardo Leijja, and Steven McCloskey. 2023. A virtual and mixed reality platform for molecular design & drug discovery - Nanome version 1.24. In *5th Workshop on Molecular Graphics and Visual Analysis of Molecular Data*. https://doi.org/10.2312/molva.20231114

[4] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Research* 28, 1 (2000), 235–242. https://doi.org/10.1093/nar/28.1.235

[5] Ewen Callaway. 2022. Scientists are using AI to dream up revolutionary new proteins. *Nature* 609 (2022), 661–662. https://doi.org/10.1038/d41586-022-02947-7

[6] Daniel Cressey and Ewen Callaway. 2017. Cryo-electron microscopy wins chemistry Nobel. *Nature* 550 (2017), 167. https://doi.org/10.1038/nature.2017.22738

[7] Justas Dauparas, Ivan Anishchenko, Nathaniel R. Bennett, Hua Bai, Robert Ragotte, Lukas F. Milles, Basile I. M. Wicky, Alexis Courbet, Robbert Jan de Haas, Neville Bethel, Phil J. Y. Leung, Timothy F. Huddy, Sam Pellock, Doug Tischer, Frederick Chan, Brian Koepnick, Hannah Nguyen, Alex Kang, Banumathi Sankaran, Asim K. Bera, Neil P. King, and David Baker. 2022. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* 378, 6615 (2022), 49–56. https://doi.org/10.1126/science.add2187

[8] Warren L. DeLano. 2002. PyMOL: an open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography* 40, 1 (2002), 82–92. https://legacy.ccp4.ac.uk/newsletters/newsletter40/11_pymol.pdf

[9] Jiayi Dou, Anastassia A. Vorobieva, William Sheffler, Lindsey A. Doyle, Hahnbeom Park, Matthew J. Bick, Binchen Mao, Glenna W. Foight, Min Yen Lee, Lauren A. Gagnon, Lauren Carter, Banumathi Sankaran, Sergey Ovchinnikov, Enrique Marcos, Po-Ssu Huang, Joshua C. Vaughan, Barry L. Stoddard, and David Baker. 2018. De novo design of a fluorescence-activating β-barrel. *Nature* 561 (2018), 485–491. https://doi.org/10.1038/s41586-018-0509-0

[10] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (2021), 583–589. https://doi.org/10.1038/s41586-021-03819-2

[11] Robert Kleffner, Jeff Flatten, Andrew Leaver-Fay, David Baker, Justin B. Siegel, Firas Khatib, and Seth Cooper. 2017. Foldit Standalone: a video game-derived protein structure manipulation interface using Rosetta. *Bioinformatics* 33, 17 (2017), 2765–2767. https://doi.org/10.1093/bioinformatics/btx283

[12] Sang-Hyun Lee, Taegyu Jin, Joon Hyub Lee, and Seok-Hyung Bae. 2022. WireSketch: bimanual interactions for 3D curve networks in VR. In *UIST '22 Adjunct*. Article 37, 3 pages. https://doi.org/10.1145/3526114.3558726

[13] David Yin-wei Lin, Yoshimasa Tanaka, Masashi Iwasaki, Apostolos G. Gittis, Hua-Poo Su, Bunzo Mikami, Taku Okazaki, Tasuku Honjo, Nagahiro Minato, and David N. Garboczi. 2008. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. *Proceedings of the National Academy of Sciences* 105, 8 (2008), 3011–3016. https://doi.org/10.1073/pnas.0712278105

[14] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. 2021. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Science* 30, 1 (2021), 70–82. https://doi.org/10.1002/pro.3943

[15] John P. Priestle. 1988. *RIBBON*: a stereo cartoon drawing program for proteins. *Journal of Applied Crystallography* 21, 5 (1988), 572–576. https://doi.org/10.1107/S0021889888005746

[16] Joseph L. Watson, David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, Andrew J. Borst, Robert J. Ragotte, Lukas F. Milles, Basile I. M. Wicky, Nikita Hanikel, Samuel J. Pellock, Alexis Courbet, William Sheffler, Jue Wang, Preetham Venkatesh, Isaac Sappington, Susana Vázquez Torres, Anna Lauko, Valentin De Bortoli, Emile Mathieu, Sergey Ovchinnikov, Regina Barzilay, Tommi S. Jaakkola, Frank DiMaio, Minkyung Baek, and David Baker. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* (2023). https://doi.org/10.1038/s41586-023-06415-8