A Biophysical and Biomechanical Study of the Central Rod Domain of Dystrophin

T. R. Sizemore[1], D. E. B. Gomes[2], P. S. F. C. Gomes[2], R. C. Bernardi[3]

[1] Undergraduate Student, Physics Department , Auburn University
[2] Postdoctoral Researcher, Physics Department, Auburn University
[3] Assistant Professor, Physics Department, Auburn University

Abstract

The purpose of this project was to predict accurate, high resolution tridimensional structures for the spectrin-repeats (SR) of the central rod region of dystrophin. These structures can then be utilized in in-silico single molecule force spectroscopy experiments, such as Steered Molecular Dynamic Simulations to understand the mechanical properties of those mechanosensing proteins. The region of interest of the protein dystrophin is the central rod domain, which contains 24 SR regions that can be divided into four fragments: SR01-05, SR06-10, SR11-17, and SR18-24. Utilizing an artificial intelligence based molecular modeling software named AlphaFold, the individual structures of each SR were created. It was determined that these structures were accurate after careful inspection via sequence and structural alignment with the crystal structure of SR01. Accurate structures containing two adjacent SR regions were obtained using AlphaFold for SR01-02 through SR23-24. Superimposing these regions upon one another resulted in an accurate structure for each of the four fragments of the central rod domain of dystrophin. After combining the superimposed regions as a template with a sequence alignment in the software MODELLER, a single structure file was generated for each fragment, demonstrating that it is possible to obtain accurate structures of the entire central rod domain of dystrophin.

Key Words: Dystrophin, Spectrin-Repeat (SR), Molecular Dynamics, Molecular Modeling.

Introduction

Muscles consist of proteins that enable contraction to produce force and allow an organism to move. Almost all of the energy utilized in the cell is consumed in this contraction process, with the rest of the energy being[1] used by a significantly smaller portion of the cell that is reserved for preservation of the cellular integrity. The dystrophin glycoprotein complex (DCG) plays a vital role in cellular integrity as it links microtubules, thin and intermediate filaments with other key components of the extracellular matrix. Within the DCG, dystrophin (427 kDa) is the protein that attaches cytoskeletal components and the protein dystroglycan, stabilizing the sarcolemma (1).

The function of the DCG can be disrupted by means of mutations within the dystrophin gene, resulting in a mutated form of dystrophin. This is known as Duchenne's Muscular Dystrophy (DMD), which is the second most common genetic disease, as it affects one out of every 3,500 males born. The fatal dystrophinopathy can be caused by chromosomal rearrangement or deletion at the dystrophin locus by missense, point, or nonsense mutations. There is no known cure for DMD, and life expectancy ranges from mid twenties to early thirties (2-4).

Dystrophin has four main functional domains: an actin binding domain, a central rod domain, a cysteine-rich domain, and a carboxyl-terminus domain(5). While these aforementioned domains have been shown to be important in mechanical linkage of dystrophin, previous studies have demonstrated that presence of the C-terminus and N-terminal actin-binding domain is not necessary for the localization of dystrophin at the sarcolemma(6-9). Previous studies have found that the central rod domain of dystrophin is vital in rescuing the phenotype in the mouse model of Duchenne's Muscular Dystrophy, which is commonly referred to as the mdx

[1] Corresponding author: trs0044@auburn.edu

model. In those studies, microdystrophin that contained 8 spectrin-repeats (SR) more efficiently rescued the phenotype of mdx as compared to the microdystrophin that contained only four SR, thus demonstrating the impact of the central rod domain on the phenotype of dystrophinopathies (10-14).

Dystrophin's central rod domain consists of 24 SR, which are ~110 amino acid motifs of triple alpha helices folded into small rods that are approximately five nm in length(15,16). In order to understand more about the molecular mechanism by which dystrophin stabilizes the sarcolemma, a previous study broke up the 24 SR into four fragments: SR01-05: 338-938th, SR06-10: 939-1466th, SR11-17: 1464-2210th, and SR18-24: 2209-3044th (17). Upon breaking up the central rod domain, each fragment was stretched. The unfolding/refolding dynamics of these fragments was determined, and the mechanical properties of those fragments were quantified.

Molecular dynamic (MD) simulations can be utilized to determine the biophysical and biomechanical properties of proteins down to an atomistic level, at which it can provide unfolding/refolding dynamics of mechanosensing proteins(18- 20). A limitation of MD simulations, however, is that they are reliant upon accurate, high resolution structural data (21-23). As of this writing, there does not exist an accurate protein structure file of the central rod domain. The artificial intelligence based software AlphaFold could serve as the key to solving this problem, as previous studies have shown that the software was able to predict a near-native protein fold based upon its genetic sequence (24).

Within the scope of this project, AlphaFold can be utilized to create accurate protein structure files for the central rod domain of dystrophin. Upon ensuring the accuracy of these protein structures, they can then be used in MD simulations to obtain the mechanical properties of each fragment of the central rod domain, and the entire length as well.

Methods

Determination of Residues

In order to create accurate protein structures using AlphaFold, the residues of each SR must be selected and an appropriate fasta file created. The sequences for each individual SR of the central rod domain were retrieved from the protein sequence database UniProt (25).

Creation and Confirmation of Individual and Paired Spectrin-Repeats

Upon retrieving appropriate sequences per each single SR region and each paired SR region, AlphaFold version 2 was used through the Visual Molecular Dynamics (VMD) QwikFold plugin batch mode to create models for each SR region and paired region(27-28). An example of the single SR structures generated is seen in Fig. 1, and an example of a paired SR region is seen in Fig. 3. In order to ensure that the structures created were accurate, each individual SR created by AlphaFold was compared to the crystal structure of the first spectrin-repeat, SR01, deposited on the Protein Data Bank (ID:3UUN)(29). This superimposition is seen in Fig 2. Using the molecular modeling system VMD the crystal structure of SR01 was superimposed upon each individual SR to obtain structural and sequence alignment data (30).

Creation of Each Region: R1: SR01-05, R2: SR06-10, R3: SR11-17. R4: SR 18-24

Upon determining the accuracy of each individual SR as well as each paired adjacent SR's, the adjacent SR's were superimposed upon one another to form each region of the central rod domain. An example of the superimposed doublets forming a region is pictured in Fig. 4. While the generated superimposed structure is an accurate representation of the region, it is not a single structure file, which is necessary to run MD simulations. To turn each superimposed region into a single structural file, the individual PDB files that comprised the superimposed regions were saved in their specific coordinates as templates. Using the sequences of each one of the superimposed structures as well as the sequence of the entire region, a sequence alignment was run through the web-server PROMALS3D, which has been shown to outperform a number of existing methods for constructing multiple sequence or structural alignments using both reference dependent and reference-independent evaluation methods(31). The templates and sequence alignment were combined in MODELLER, which is an effective tool for comparative modeling of protein three-dimensional structures (32-35). After running MODELLER, the structure with the highest discrete optimized protein energy (DOPE)

score was selected and a single accurate, high resolution structure of each region was obtained. The resulting structures for Region 1, 2, 3, and 4 are pictured in Fig. 5, Fig. 6, Fig. 7, and Fig. 8 respectively.

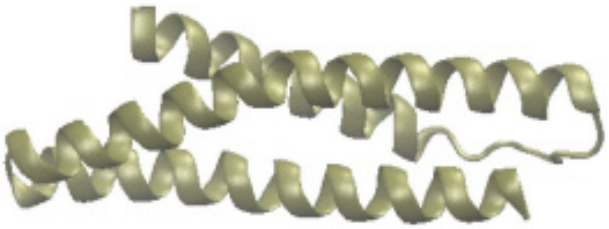

Fig. 1. Example of an AlphaFold generated structure for an individual spectrin-like repeat (SR). Represented in cartoon and colored in tan is the AlphaFold generated structure of SR06.

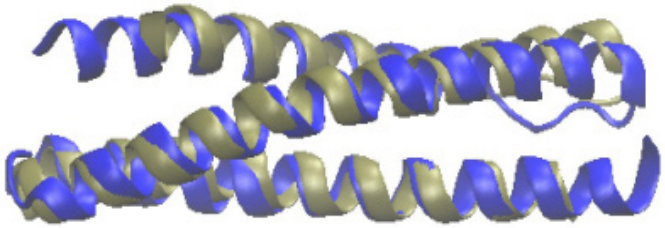

Fig. 2. Crystal structure of SR01 (blue) superimposed upon the AlphaFold generated structure of SR06 (tan). We notice the conservation of the triple alpha helical motif.

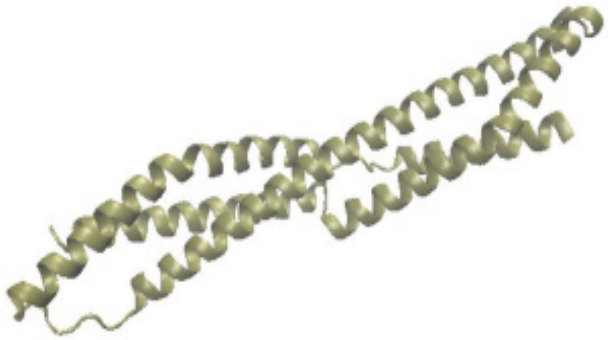

Fig. 3. Adjacent AlphaFold generated structure for SR06-07.

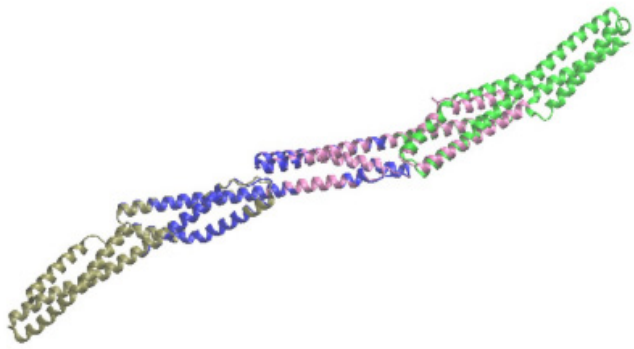

Fig. 4. Adjacent SR's superimposed structures that compose Region 2: SR06-07 (tan), SR07-08 (blue), SR08-09 (pink), and SR09-10 (green).

Fig. 5. MODELLER generated structure of Region 1:SR01-05.

Fig. 6. MODELLER generated structure of Region 2: SR-6-10.

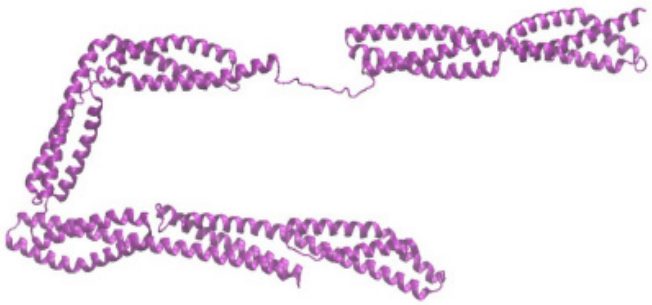

Fig. 7. MODELLER generated structure of Region 3: SR11-17

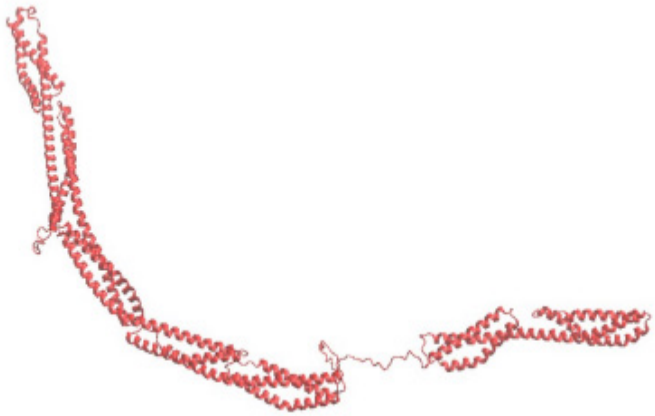

Fig. 8. MODELLER generated structure of Region 4: SR18-24.

Conclusions

At the time of writing this paper, there does not exist an accurate protein structure of the central rod domain of dystrophin. Accurate structures of paired adjacent spectrin repeats were created upon determining the accuracy of each individual spectrin repeat. Each paired SR doublet maintains the consistent structure of a triple helical motif consisting of a rod-like shape. By superimposing each adjacent pair upon one another, an accurate structure of each of the regions was created. The full structure maintains the expected structure of the entire fragment, as it maintains the triple helical motif and forms the rod-shaped protein structure. Creating accurate protein structures using AlphaFold allows for the use of MD simulations to determine biophysical and biomechanical properties of proteins at an atomistic level. These findings demonstrate that our protocol allows us to obtain tridimensional structure models that are accurate. These structures are currently being utilized in Steered Molecular Dynamic simulations in order to obtain biophysical and biomechanical data of the central rod domain of dystrophin.

Acknowledgements

The authors thank Auburn University and the College of Sciences and Mathematics as well as the office of Undergraduate Research for jointly funding this project. This work was supported by the National Science Foundation under Grant MCB-2143787 (CAREER: In Silico Single-Molecule Force Spectroscopy).

Nomenclature

SR - Spectrin-like Repeat. MD - Molecular Dynamics. DOPE - discrete optimized protein energy. PDB - protein data bank.

References

1. Henderson, Davin Michael. (2011). A biochemical and biophysical study of dystrophin. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/120598.

2. Hoffman, E. P.; Monaco, A. P.; Feener, C. C.; Kunkel, L. M. Conservation of The Duchenne Muscular Dystrophy Gene in Mice and Humans. Science 1987, 238, 347– 350, DOI: 10.1126/science.3659917

3. Hoffman, E. P.; Brown, R. H., Jr.; Kunkel, L. M. Dystrophin: The Protein Product of The Duchenne Muscular Dystrophy Locus. Cell 1987, 51, 919– 928, DOI: 10.1016/0092-8674(87)90579-4

4. Koenig, M.; Beggs, A. H.; Moyer, M.; Scherpf, S.; Heindrich, K.; Bettecken, T.; Meng, G.; Muller, C. R.; Lindlof, M.; Kaariainen, H.; de la Chapellet, A.; Kiuru, A.; Savontaus, M. L.; Gilgenkrantz, H.; Recan, D.; Chelly, J.; Kaplan, J. C.; Covone, A. E.; Archidiacono, N.; Romeo, G.The Molecular Basis for Duchenne Versus Becker Muscular Dystrophy: Correlation of Severity with Type of Deletion. Am. J. Hum. Genet. 1989, 45, 498– 506

5. Gao QQ, McNally EM. The Dystrophin Complex: Structure, Function, and Implications for Therapy. Compr Physiol. 2015 Jul 1;5(3):1223-39. doi: 10.1002/cphy.c140048. PMID: 26140716; PMCID: PMC4767260.

6. Le Rumeur, E.; Winder, S. J.; Hubert, J. F. Dystrophin: More Than Just The Sum of Its Parts. Biochim. Biophys. Acta, Proteins Proteomics 2010, 1804, 1713–1722, DOI: 10.1016/j.bbapap.2010.05.001 7. Dunckley, M. G.; Wells, K. E.; Piper, T. A.; Wells, D. J.; Dickson, G. Independent Localization of Dystrophin N- and C-Terminal Regions to The Sarcolemma of Mdx Mouse Myofibres in Vivo. J. Cell Sci. 1994, 107, 1469– 1475

8. Helliwell, T. R.; Ellis, J. M.; Mountford, R. C.; Appleton, R. E.; Morris, G. E. A Truncated Dystrophin Lacking The C Terminal Domains Is Localized at The Muscle Membrane. Am. J. Hum. Genet. 1992, 50, 508– 514

9. Corrado, K.; Rafael, J. A.; Mills, P. L.; Cole, N. M.; Faulkner, J. A.; Wang, K.; Chamberlain, J. S. Transgenic Mdx Mice Expressing Dystrophin with A Deletion in The Actin-Binding Domain Display A "Mild Becker" Phenotype. J. Cell Biol. 1996, 134, 873– 884, DOI: 10.1083/jcb.134.4.873

10. Warner, L. E.; DelloRusso, C.; Crawford, R. W.; Rybakova, I. N.; Patel, J. R.; Ervasti, J. M.; Chamberlain, J. S. Expression of Dp260 in Muscle Tethers The Actin Cytoskeleton to The Dystrophin-Glycoprotein Complex and Partially Prevents Dystrophy. Hum. Mol. Genet. 2002, 11, 1095– 1105, DOI: 10.1093/hmg/11.9.1095

11. Wang, B.; Li, J.; Xiao, X. Adeno-Associated Virus Vector Carrying Human Minidystrophin Genes Effectively Ameliorates Muscular Dystrophy in Mdx Mouse Model. Proc. Natl. Acad. Sci. U. S. A. 2000, 97, 13714–13719, DOI: 10.1073/pnas.240335297

12. Watchko, J.; O'Day, T.; Wang, B.; Zhou, L.; Tang, Y.; Li, J.; Xiao, X. Adeno-Associated Virus Vector-Mediated Minidystrophin Gene Therapy Improves Dystrophic Muscle Contractile Function in Mdx Mice. Hum. Gene Ther. 2002, 13, 1451– 1460, DOI: 10.1089/10430340260185085

13. Sakamoto, M.; Yuasa, K.; Yoshimura, M.; Yokota, T.; Ikemoto, T.; Suzuki, M.; Dickson, G.; Miyagoe-Suzuki, Y.; Takeda, S. Micro-Dystrophin Cdna Ameliorates Dystrophic Phenotypes When Introduced into Mdx Mice As A Transgene. Biochem. Biophys. Res. Commun. 2002, 293, 1265– 1272, DOI: 10.1016/S0006-291X(02)00362-5

14. Judge, L. M.; Haraguchiln, M.; Chamberlain, J. S. Dissecting The Signaling and Mechanical Functions of The Dystrophin-Glycoprotein Complex. J. Cell Sci. 2006, 119, 1537– 1546, DOI: 10.1242/jcs.02857.

15. Broderick MJ, Winder SJ. Spectrin, alpha-actinin, and dystrophin. Adv Protein Chem. 2005;70:203–246.

16. Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod–shaped cytoskeletal protein. Cell. 1988;53:219–228.

17. Dystrophin As a Molecular Shock Absorber Shimin Le, Miao Yu, Ladislav Hovan, Zhihai Zhao, James Ervasti, and Jie Yan ACS Nano 2018 12 (12), 12140-12148 DOI: 10.1021/acsnano.8b05721

18. Lee, E. H., Hsin, J., Sotomayor, M., Comellas, G., and Schulten, K. (2009). Discovery through the computational microscope. Structure 17, 1295–1306. doi:10.1016/J.STR.2009.09.001

19. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012). Biomolecular simulation: A computational microscope for molecular biology. Annu. Rev. Biophys. 41, 429–452. doi:10.1146/ANNUREV BIOPHYS-042910-155245

20. Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., et al. (2015). Molecular dynamics simulations of large macromolecular complexes. Curr. Opin. Struct. Biol. 31, 64–74. doi:10.1016/J.SBI.2015.03.007

21. Bernardi, R. C., and Pascutti, P. G. (2012). Hybrid QM/MM molecular dynamics study of benzocaine in a membrane environment: How does a quantum mechanical treatment of both anesthetic and lipids affect their interaction. J. Chem. Theory Comput. 8, 2197–2203. doi:10.1021/ct300213u

22. Vanommeslaeghe, K., and MacKerell, A. D. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochimica Biophysica Acta - General Subj. 1850, 861–871. doi:10.1016/j.bbagen.2014.08.004

23. Melo, M. C. R., Bernardi, R. C., Rudack, T., Scheurer,

M., Riplinger, C., Phillips, J. C., et al. (2018). NAMD goes quantum: An integrative suite for hybrid simulations. Nat. Methods 15, 351–354. doi:10.1038/nmeth.4638

24. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021b). Highly accurate protein structure prediction with AlphaFold. Nature 596 (596), 7873583–7873589. doi:10.1038/s41586-021-03819-2

25. Koenig M, Monaco AP, Kunkel LM. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. Cell. 1988 Apr 22;53(2):219-28. doi: 10.1016/0092- 8674(88)90383-2. PMID: 3282674.

26. Drozdetskiy A, Cole C, Procter J & Barton GJ. (first published online April 16, 2015) JPred4: a protein secondary structure prediction server, Nucleic Acids Res. Web Server issue [doi:10.1093/nar/gkv332]

27. Barreto Gomes, D. E.; Gomes, P. S.; Bernardi, R. C. QwikMD 2.0: Bridging the Gap between Sequence, Structure, and Protein Function. Biophys. J.2022, 121 (3), 132a. https://doi.org/10.1016/j.bpj.2021.11.2041.

28. Gomes, D. E. B., Melo, M. C. R., Gomes, P. S. F. C., and Bernardi, R. C. (2022). Bridging the gap between in vitro and in silico single-molecule force spectroscopy. bioRxiv. doi:10.1101/2022.07.14.500151

29. Muthu M, Richardson KA, Sutherland-Smith AJ. The crystal structures of dystrophin and utrophin spectrin repeats: implications for domain boundaries. PLoS One. 2012;7(7):e40066. doi: 10.1371/journal.pone.0040066. Epub 2012 Jul 20. PMID: 22911693; PMCID: PMC3401230.

30. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. 14, 33– 38. doi:10.1016/0263-7855(96)00018-5

31. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res. 2008 Apr;36(7):2295-300. doi: 10.1093/nar/gkn072. Epub 2008 Feb 20. PMID: 18287115; PMCID: PMC2367709.

32. B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.

33. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.

34. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779- 815, 1993.

35. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

Authors Biography

Toby Sizemore is a junior-year student pursuing a B.A. degree in Chemistry and a minor in Biological Sciences at Auburn University. He is an Undergraduate Research Fellow investigating the protein involved in Duchenne Muscular Dystrophy (DMD). He is inspired by his brother, who was born with DMD, to pursue an MD/PhD to become a medical scientist.

Dr. Diego E. B. Gomes is a postdoctoral researcher at Auburn university. Diego holds a PhD in Biophysics from the Universidade Federal do Rio de Janeiro (Brazil). He is currently working on developing QwikFold, a VMD interface for AI based structural biology software. QwikFold will be integrated into a new version of QwikMD, which is also under development by Dr. Gomes. Besides that, he is working on mechanical properties of nanobodies, proteins that are under development for diagnosing and fighting cancer.

Dr. Priscila S. F. C. Gomes is a postdoctoral researcher at Auburn University. Priscila holds a dual-PhD in Biophysics from the École Normale Supérieure de Cachan (France) and the Universidade Federal do Rio de Janeiro (Brazil). Her main research interest is on the use of bioinformatics and molecular dynamics tools to investigate bacteria and virus adhesion. Dr. Gomes has recently contributed to understanding how mechanical properties of the COVID spike protein has evolved since the SARS-CoV-1 outbreak in early 2000s

Dr. Rafael C. Bernardi is an Assistant Professor of Biophysics at the Department of Physics at Auburn University. Before joining Auburn's faculty, Dr. Bernardi was a Postdoc (2012-2017), and then a Research Scientist (2017-2020), at the Beckman Institute at the University of Illinois. Dr. Bernardi is co-Investigator of the NIH Center for Macromolecular Modeling and Visualization, which is known worldwide for the development of NAMD and VMD software. Dr. Bernardi's main research interest is on protein mechanics, and how protein complexes behave under mechanical load. Such problems are very important in many areas of biomedicine, including bacterial adhesion during infection, viruses interaction with human proteins, and enzymatic activity in the gut microbiome.