Combining Artificial Intelligence and Virtual Reality to design new peptides inhibiting the ET domain of BRD4

Kalea Moore[1], Bhumika Singh[1], Alberto Perez*[1]
Quantum Theory Project, University of Florida
Counselor Ian Coffey[2], *UF CPET SSTP*

***Abstract*—Recent peptide therapeutics study the use of peptides in inhibiting the harmful functions of viral proteins. This study experimentally investigates creating a novel peptide derived from the latency-associated nuclear antigen (LANA-1 or LANA) that binds to the ET domain of Bromodomain-containing protein 4 (BRD4). Mutations were done to LANA in artificial intelligence hardware AlphaFold2 (AF2) and virtual reality (VR) software Nanome VR to improve binding interactions between LANA and BRD4. AlphaFold2 is a Linux coding software that uses multiple sequence alignments (MSA) to predict the 3D structures of biomolecular structures. Nanome utilizes various tools—scaling, rotating, coloring—to display the specific properties of 3D designs. These include hydrogen bonding, clashes between residues, and overall binding affinity. Results show that mutations contributed as many as six additional hydrogen bonds with less clashing. This study aims to use these implications to induce a decreased likelihood of tumor protection and the development of various cancerous and noncancerous diseases.**

***Key Terms*—BRD4, LANA, ET domain, RCSB PDB, AlphaFold2, Nanome.**

I. INTRODUCTION

Proteins, composed of multiple peptide chains, fuel thousands of intracellular processes. This code is created from amino acid sequences and determines the folding structure correlating to a protein's functions. These can range from immunological to respiratory regulation. When a virus enters the cell, it can hijack the protein into replicating its viral DNA by placing itself near the protein's transcription-start site. This process can create a viral protein, a modified polypeptide that can have harmful new functions.

A. Peptide therapeutics

There is an ongoing application of peptides as various therapeutics to combat this. Proteins have binding cavities in which small molecules can attach. However, peptides, or tiny proteins, lack the need for these sites and can instead bind through connected chains. Peptides can distract a protein by mimicking a protein's surface,

making it possible to hinder dangerous protein-protein interactions (PPI) [2], [3]. Their role spans from inhibiting the function of the proteins they bind to, to enhancing it. The extent to which they can do this depends on their size, specifically their surface area. A smaller peptide is unable to cover as many binding cavities [4] that may attach to acetylated histones (octameric proteins that decrease transcription via interaction) [7]. Peptides inhibiting viral proteins could reduce the likelihood of contracting several illnesses [3].

BRD4 of the bromodomain family The bromodomain and extra-terminal (BET) family is a group of four similar proteins, each with their beneficial roles in inflammation responses. Their name originates from their two N-terminal bromodomains and their varying C-terminal motifs. Their ET domains offer binding sites for viruses to copy and produce their genome. Under a virus's influence, the BET family can begin replicating/recruiting more viral proteins. BRD4 [5], containing two C-terminuses, is a commonly researched member of this family that, once infected by a virus, will begin to protect tumor tissue and express both cancerous and noncancerous diseased genes. The result of this unfortunate new role is the development of acute myeloid leukemia, aggressive carcinoma, human immunodeficiency virus (HIV), and several other destructive ailments [1].

LANA-1 inhibiting BRD4 A large nuclear peptide designated latency-associated nuclear antigen (LANA-1) [6] is the best-known tool for inhibiting BRD4. LANA is divided into an N-terminal, C-terminal, and central region, with its main roles being regulating viral episomes and transcription. Due to its nature of mimicry and its oligopeptide structure, LANA-1 successfully binds to and impedes BRD4's viral functions.

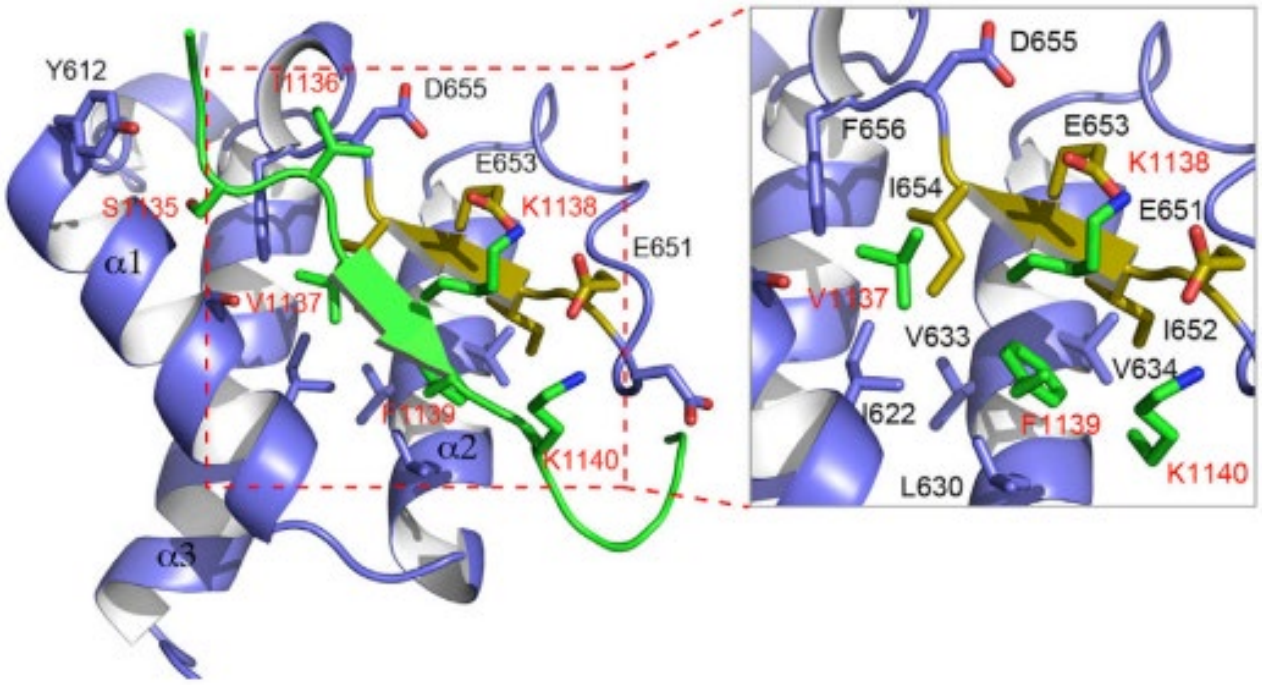


Figure 1: Binding Site of BRD4 and LANA. Residues in red depict key residues in LANA interacting with ET domain binding site residues [14].

Protein predictor AlphaFold2 with Nanome viewer AlphaFold2 [9], winner of the 2018 critical assessment of protein structure prediction (CASP), is an AI program capable of producing three-dimensional (3D) protein structures [10] using input amino acid sequences. Utilizing its ability to create multiple sequence alignments (MSA) and/or templates of proteins, we can predict the structure for given proteins in a few minutes. These structure predictions closely match other experimental methods, including X-ray crystallography, cryoEM, or NMR. The 3D structures of proteins obtained could then be used to study further how the protein binds with other biomolecules, such as peptides.

Viewing biomolecules with Nanome Once a protein-peptide complex's 3D folding structure is predicted, it can be loaded into a virtual reality program, Nanome, for further research. Nanome makes it possible to manipulate other residues, obtain a maximized view of the substitution, and see all perhaps masked amino acids within it.

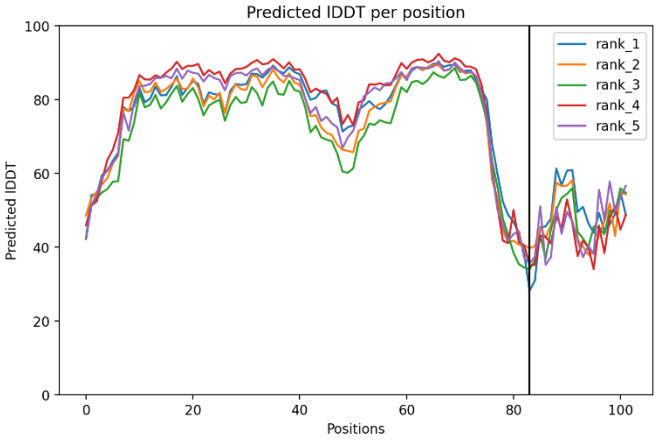
B. Utilization

The scope of this research lies within the ability to create new peptides (while referencing ones already deemed prominent binding partners) that will inhibit the disease-causing functions of viral proteins. The expected result is the reduced potential and spread of these diseases. Using AlphaFold2 and Nanome, we duplicate LANA-1 and design peptides that have better binding interactions with BRD4, inhibiting it for longer.

II. METHODOLOGICAL PROCEDURE

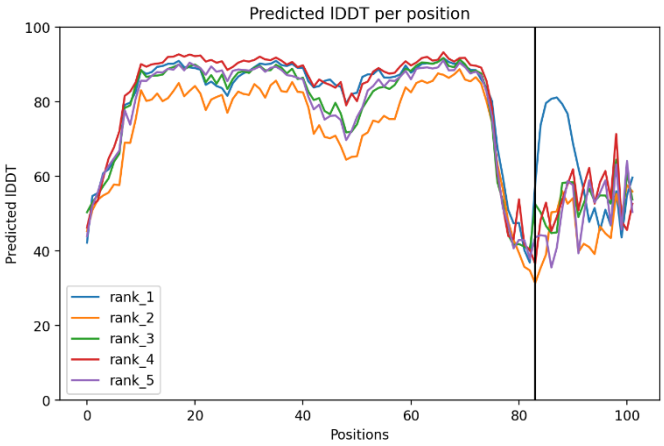
A. Materials

The protein-peptide complex composed of the ET domain of the BRD4 protein [1], [5] bound to LANA-1 [6] was taken from the RCSB Protein Databank (PDB, with code 2ND0) [11].



Hardware A MacBook Air M2 and Windows 11 Enterprise (both Mac and Windows serviceable for future replication of this study) were obtained from Apple Store. A Meta Quest 2 virtual reality headset (02) and two touch controllers (L&R) were purchased from Oculus/Meta Platform Technologies, LLC [12].

Software In the first phase, Google ColabFold was provided by Google to check that the AF2 predicted protein-peptide structure matched with the native structure (PDB ID 2ND0) [11]. In the following stages, where AF2 predicted the structures for the protein-peptide complexes with mutations, a Linux terminal was provided by MobaXTerm and WinSCP for file transfers. The HiPerGator supercomputing facility of the University of Florida was used to run AF2. Nanome Inc created the Nanome software used with Oculus to visualize and design novel peptides. Nanome [13] was downloaded onto the Meta Quest 2 headset before quantitative experimenting.



B. Procedure

AlphaFold2 The FASTA sequence of amino acids for the protein-peptide complex was taken as the input along with a template (PDB ID 2ND0). The MSA option was set to *mmseqs2_uniref_env*. The five top-ranked structures and pLDDT score were obtained for each complex. The pLDDT score measures per-residue confidence for each predicted structure, with a score between 0-100. The higher the confidence, the better the structure. For the initial analysis, the top-ranked structures were further analyzed in Nanome. AF2 structures run in phase two had the desired mutations added.

Figure 2: pLDDT Scores for the Native BRD4-LANA (PDB ID 2ND0) and mutated F9S complex, respectively.

Nanome The top-ranked BRD4-LANA predicted structure from AF2 was loaded into Nanome. A single-point mutation on each peptide residue was performed to design new peptides. The other 19 standard amino acids replaced each residue. The parameters, namely, hydrogen bonds, steric clashes, and orientation, were observed for

each mutation performed. The H-bonds were studied within a 0.4 Å radius distance while the clashes were set to an *Overlap Tolerance* of 0.6 Angstrom (Å). The mutation resulting in more H-bonds and fewer steric clashes was preferred.

In the second phase, structure prediction was again performed using AF2 on all the mutations with the desirable changes. Afterward, the top-ranked structures were again analyzed in Nanome. This time, the objective was to verify the prediction based on Nanome-based studies. For this, only the difference in H-bonds and steric clashes between the first and second phases was analyzed. Ball-and-stick representations were used during the entire study.

III. RESULTS

Viral protein BRD4 binds to latency peptide LANA through its ET domain (residues 1,133–1,144) with a current approximate binding affinity of K_D = 635 μM [14]. Various amino acid residue mutations can influence a change in binding affinity based on multiple parameters, including but not limited to hydrogen bonding, hydrophobic pockets, electrostatic interactions, and sulfhydryl or aromatic side chains. Mutations conducted by substituting each of the residues (20) into the ET domain that binds to the LANA peptide [14] yielded changes in most of these parameters.

Table 1: List of Successful Mutations Performed on LANA Peptide

Green: nonpolar, orange: polar (-ve), yellow: polar (0), pink: polar (+ve)

Alanine (A), Valine (V), Leucine (L), Glycine (G), Isoleucine (I), Methionine (M), Tryptophan (W), Cysteine (C), Proline (P), Phenylalanine (F), Aspartic Acid (D), Glutamic Acid (E), Tyrosine (T), Asparagine (N), Glutamine (Q), Serine (S), Threonine (T), Arginine (R), Lysine (K), Histidine (H).

Residue	Mutation	H-bond	Clashing	Changes in Secondary Structure
Serine 5 (S5)	S5D	+2	same	antiparallel
	S5N	+2	less	antiparallel
Isoleucine 6 (I6)	I6C	+5	same	antiparallel
	I6Q	+3	same	antiparallel
	I6S	+4	same	antiparallel
	I6R	+3	same	antiparallel
	I6H	+4	less	antiparallel
Phenylalanine 9 (F9)	F9S	+6	less	antiparallel
	F9R	+5	less	antiparallel
	F9K	+5	less	antiparallel
Lysine 11 (K11)	K11L	+2	less	antiparallel
	K11I	+3	less	antiparallel
	K11M	+1	same	parallel
	K11W	+4	less	antiparallel
	K11F	+2	less	antiparallel
	K11Y	+1	less	antiparallel
	K11R	+2	less	antiparallel
Leucine 13 (L13)	L13N	+2	same	antiparallel
Proline 14 (P14)	P14T	+3	same	antiparallel

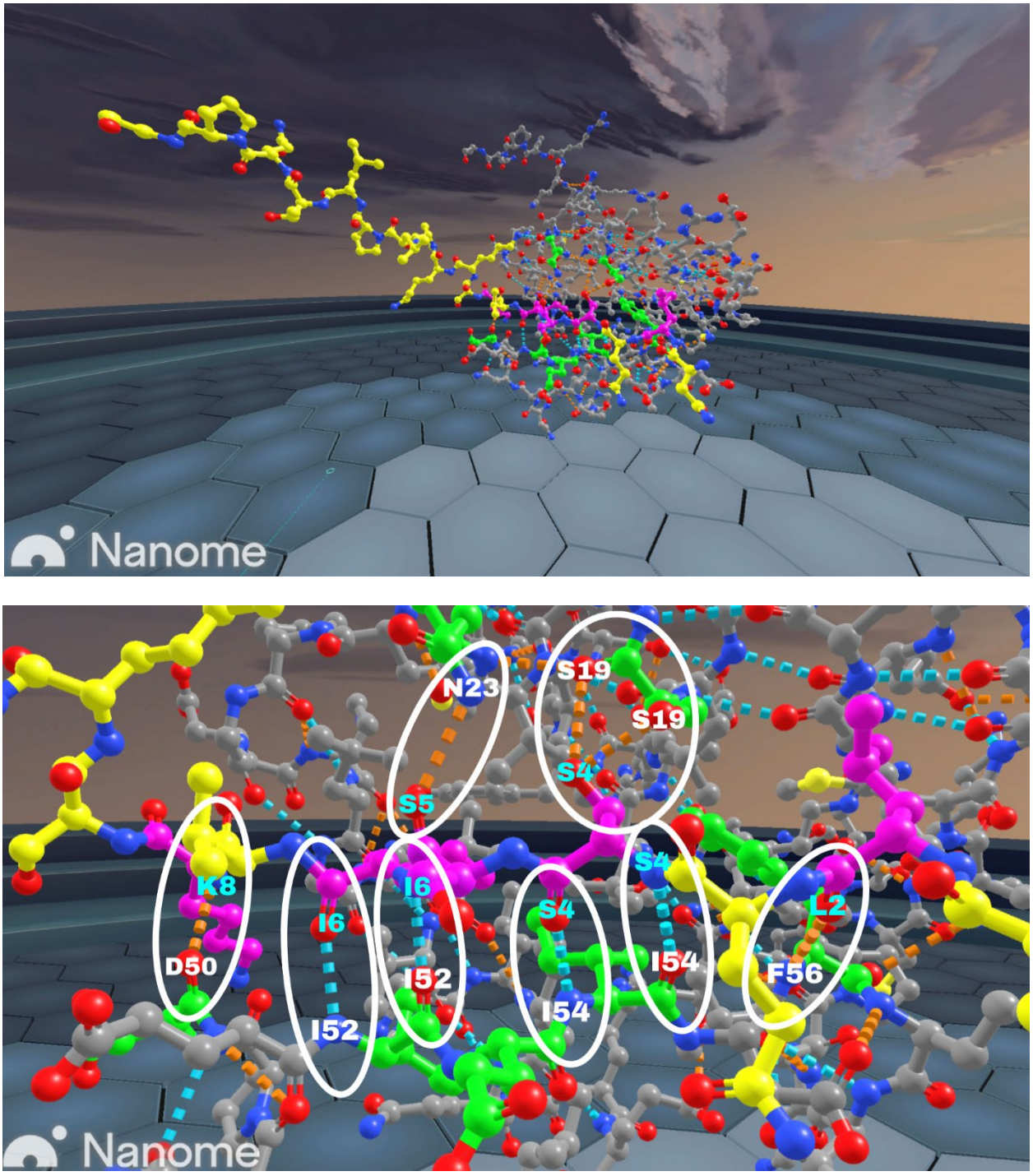


Figure 3: Phenylalanine 9 (F9) Mutated with Serine (F9S). **(a)** The entire system as it is seen inside Nanome. (**b)** Hydrogen bonds between the peptide (blue text) and protein (white text) residues are represented by the blue and red dotted lines. The yellow structure is the mutated peptide, and the grey structure is the protein BRD4. The interacting residues of the peptide are highlighted in pink, while the interacting residues of the protein are in green.

Final mutations after analyzing AlphaFold2 predictions Mutations predicted by AF2 were put back into Nanome to compare the new structure to the previous mutation done on the LANA peptide. Hydrogen bonds between BRD4 and the model peptide were counted and recorded. Four hydrogen bonds were present in the original complex.

The final mutations were put into tables for each residue based on several conditions spanning from more hydrogen bonding and less clashing. There were six key residues and six non-key residues for which data was recorded, making up the eleven total residues interacting within LANA's bound ET domain (residues 1,133 to 1,144, for simplicity, they are referred to as 3-14 in this report). Key residues valine 7 (V7), lysine 8 (K8), lysine 10 (K10, and non-key residues glutamine 3 (Q3), serine 4 (S4), and proline 12 (P12) had no better mutations as described by their low pLDDT score.

Table 1 is organized into three columns per residue. The *Mutations* column shows the mutated residue. The *H-Bond* column displays a number value with a plus sign (+) representing the number of additional bonds compared to the native complex. *The Clashing* column depicts the difference in the number of clashes in the actual mutation as opposed to the practice mutations, described as less or the same. The last column, *Changes in Secondary Structure,* shows whether there was structural variation in the peptide or protein structure. Mutations with no additional bonds or fewer bonds than before (four before mutations) and more clashing were eliminated. Similarly, mutations with pLDDT graphs that fall under 75 for the predicted IDDT were also omitted.

IV. CONCLUSION

The difference in additional hydrogen bonding ranged from one additional bond with fewer clashes (K11M and K11Y, Table 1) to six bonds with fewer clashes (F9S). Some mutation tables, including F9 and I6, had patterns of similar clashing results (less, same). For LANA to bind to BRD4, electrostatic anchoring by two arginine or lysine residue between two glutamates is needed [14]. F9's aromatic side chain enters a hydrophobic pocket within the three-helix bundle [14]. Moreover, replacing a nonpolar residue such as F9 with something polar positive, like arginine or lysine, is more likely to attract other negative residues and form more hydrogen bonds (+5 for both F9R and F9K). This is the opposite of what occurs in K11, as it is a positive polar residue in which the most successful mutations are of the nonpolar group. Polar neutral residues (tyrosine, asparagine, glutamine, serine, and threonine) were the most present group, recorded as successful in at least one out of the remaining successful six mutated residues (ranging from +1 K11Y to +5 F9R). This may be due to their neutral charge attracting other positive or negative residues; additionally, their polarity would yield more hydrogen bonds. All mutations except for K11M resulted in an antiparallel beta-strand like the native complex. K11M became parallel.

V. DISCUSSION

Initially, smaller hypotheses surrounded the interactions between specific amino acid residues involved in BRD4-LANA binding. Cysteine was expected to be more present in the final mutations because of the presence of sulfur; however, there is

only one (I6C, Table 1). The long structures of arginine, histidine, and lysine were expected to be less likely to enter the final rounds of mutations due to a higher likelihood that they would cause clashes. This was incorrect, with arginine contributing additional hydrogen bonds and less than or the same clashes for three of the six final residues (I6R, F9R, and K11R). What other studies label as critical for BRD4-LANA recognition [14], F9 unsurprisingly had the largest hydrogen bonding increase of six (ten total between BRD4 and LANA, F9S). This relates to its nonpolar phenyl group of phenylalanine, which is more likely to interact with other nonpolar amino acids.

Additional hydrogen bonds correlate to a stronger protein and peptide interaction. Fewer clashes liken to a higher likelihood of the mutation being possible and stable. In this study, a novel peptide was created from the successful results shown in Figures 2 and 3—one that can inhibit viral protein BRD4's dangerous functions better than LANA. An extended time in which BRD4 is "distracted" by the peptide corresponds to fewer opportunities for BRD4 to protect tumor cells and express disease-carrying genes. The implications of further study within this area of peptide therapeutics could lead to an overall decrease in developing related diseases—myeloid leukemia, aggressive carcinoma, and HIV.

VI. ACKNOWLEDGEMENTS

I want to thank Dr. Alberto Perez, Singh Bhumika, Reza Esmaeeli, and the rest of the Molecular Recognition Group at the University of Florida for guiding this project.

REFERENCES

[1] Z. Liu *et al.*, "Drug Discovery Targeting Bromodomain-Containing Protein 4," *Journal of Medicinal Chemistry*, vol. 60, no. 11, pp. 4533–4558, Mar. 2017, doi: https://doi.org/10.1021/acs.jmedchem.6b01761.

[2] A. D. Cunningham, N. Qvit, and D. Mochly-Rosen, "Peptides and peptidomimetics as regulators of protein–protein interactions," *Current Opinion in Structural Biology*, vol. 44, pp. 59–66, Jun. 2017, doi: https://doi.org/10.1016/j.sbi.2016.12.009.

[3] B. O. Villoutreix, C. M. Labbé, D. Lagorce, Guillaume Laconde, and O. Sperandio, "A Leap into the Chemical Space of Protein-Protein Interaction Inhibitors," *Current Pharmaceutical Design*, vol. 18, no. 30, pp. 4648–4667, Sep. 2012, doi: https://doi.org/10.2174/138161212802651571.

[4] L. Wang *et al.*, "Therapeutic peptides: current applications and future directions," *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, p. 48, Feb. 2022, doi: https://doi.org/10.1038/s41392-022-00904-4.

[5] T. Kanno *et al.*, "BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones," vol. 21, no. 12, pp. 1047–1057, Dec. 2014, doi: https://doi.org/10.1038/nsmb.2912.

[6] G. Stuber *et al.*, "HHV-8 encoded LANA-1 alters the higher organization of the cell nucleus," vol. 6, no. 1, Apr. 2007, doi: https://doi.org/10.1186/1476-4598-6-28.

[7] B. N. Devaiah *et al.*, "BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin," *Nature Structural & Molecular Biology*, vol. 23, no. 6, pp. 540–548, May 2016, doi: https://doi.org/10.1038/nsmb.3228.

[8] D. L. Nelson, A. L. Lehninger, and M. M. Cox, *Lehninger Principles of Biochemistry*, 8th ed. Macmillan Higher Education: Basingstoke, 2021.
[9] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, no. 7873, pp. 583–589, Jul. 2021.
[10] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult, "Critical assessment of methods of protein structure prediction (CASP)—Round XIII," *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 12, pp. 1011–1020, Oct. 2019, doi: https://doi.org/10.1002/prot.25823.
[11] R. P. D. Bank, "RCSB PDB - 2ND0: Solution NMR structures of BRD4 ET domain with LANA peptide," *www.rcsb.org*. https://www.rcsb.org/structure/2ND0 (accessed Jul. 12, 2023).
[12] "Research | Oculus," *www.oculus.com*. https://www.oculus.com/research/
[13] K. Funakawa, "New Harvard study shows the efficacy of Nanome & VR for Chemistry Education," *Medium*, Jan. 13, 2021. https://blog.matryx.ai/new-harvard-study-shows-the-efficacy-of-nanome-vr-for-chemistry-education-cb45da304ea2 (accessed Jul. 12, 2023).
[14] Q. Zhang *et al.*, "Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4," vol. 24, no. 7, pp. 1201–1208, Jul. 2016, doi: https://doi.org/10.1016/j.str.2016.04.019.
[15] N.-Z. Xie, Q.-S. Du, J.-X. Li, and R.-B. Huang, "Exploring Strong Interactions in Proteins with Quantum Chemistry and Examples of Their Applications in Drug Design," *PLOS ONE*, vol. 10, no. 9, p. e0137113, Sep. 2015, doi: https://doi.org/10.1371/journal.pone.0137113.