

GENERATION OF A NUCLEASE TARGETING PARALLEL G-QUADRUPLIX FOR SPECIFIC CUTTING OF DOUBLE STRANDED DNA

Dang Thanh Dung

Ho Chi Minh City Open University

ARTICLE INFO	ABSTRACT
Received: 07/3/2023	G-quadruplex is a secondary structure of DNA or RNA that plays an important role in many biological processes such as replication, transcription, translation and elongation of telomeres. Therefore, the G-quadruplex structure has emerged as a target molecule for drugs designation in biomedical applications. In this study, a novel nuclease which was developed can specifically target parallel G-quadruplex and play its catalytic function at a specific position. The novel nuclease was genetically generated by fusing a parallel G-quadruplex-recognized RHAU53 peptide motif with a catalytic protein domain of Fok1, resulting in the generation of RHAU53-Fok1. The fusion protein could be expressed in <i>E. coli</i> under IPTG inducer and purified by the his-tag affinity chromatography. Interestingly, RHAU53-Fok1 can selectively bind a parallel G-quadruplex and cut a double stranded DNA next to it. Cleavage of double stranded DNA by RHAU53-Fok1 showed multiple cutting sites on the substrate, resulting in a major fragment and several minor fragments. The novel nuclease provides a useful tool for recognizing and mapping the G-quadruplex structure in the genome.
Revised: 16/5/2023	
Published: 16/5/2023	
KEYWORDS	
G-quadruplex	
RHAU53-Fok1	
Nuclease	
DNA	
Specific cutting	

TẠO ENZYME NUCLEASE CÓ KHẢ NĂNG NHẬN DIỆN CẤU TRÚC G-QUADRUPLIX SONG SONG VÀ CẮT DNA TẠI VỊ TRÍ ĐẶC HIỆU

Đặng Thanh Dũng

Trường Đại học Mở Thành phố Hồ Chí Minh

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 07/3/2023	G-quadruplex là cấu trúc bậc 2 của DNA hay RNA mà đóng vai trò quan trọng trong các quá trình sinh học như sao chép, phiên mã, dịch mã và kéo dài các telomeres. Do đó, cấu trúc G-quadruplex được xem như một phân tử mục tiêu trong việc thiết kế các thuốc trúng đích ứng dụng trong y sinh. Trong nghiên cứu này, nuclease mới được phát triển có khả năng nhắm mục tiêu G-quadruplex song song và thực hiện chức năng cắt tại vị trí đặc hiệu. Enzyme mới được tạo ra bằng phương pháp kết hợp phân tử peptit RHAU53 có khả năng nhận diện G-quadruplex song song với vùng protein xúc tác cắt của Fok1, tạo nên protein dung hợp RHAU53-Fok1. Protein dung hợp này có thể được biểu hiện trong hệ thống <i>E. coli</i> dưới sự cảm ứng của IPTG và được tinh chế bằng sắc ký ái lực với cột His. Đặc biệt, RHAU53-Fok1 có thể nhận diện và bám đặc hiệu vào cấu trúc G-quadruplex song song và cắt DNA sợi kép bên cạnh cấu trúc đó. Enzyme cắt DNA ở nhiều vị trí cho ra 1 sản phẩm chính và các sản phẩm phụ. Enzyme mới này được xem là công cụ tiềm năng cho việc lập bản đồ vị trí cấu trúc G-quadruplex trong bộ gen.
Ngày hoàn thiện: 16/5/2023	
Ngày đăng: 16/5/2023	
TỪ KHÓA	
G-quadruplex	
RHAU53-Fok1	
Nuclease	
DNA	
Cắt đặc hiệu	

DOI: <https://doi.org/10.34238/tnu-jst.7490>

Email: dung.dthanh@ou.edu.vn

<http://jst.tnu.edu.vn>

80

Email: jst@tnu.edu.vn

1. Introduction

DNA or RNA G-quadruplexes are secondary structures formed from G-rich sequences that can be folded into four structural strands (Figure 1A) [1], [2]. The G-quadruplex structures are highly polymorphic: the four strands of the G-tetrad core can be parallel (the strands are oriented in the same direction) or nonparallel, with (i) three strands in one direction and one strand in the other direction or (ii) two strands in one direction and the other two in the opposite direction [3] (Figure 1B). Analysis of computational prediction shows that the human genome may contain more than 700,000 sequences that can form G-quadruplex structures [3]-[5]. G-quadruplexes are mainly present in the telomere region of the genome, which consists of 5 to 10,000 bps of G-rich repeats (TTAGGG) [6], [7]. G-quadruplexes also exist in the promoter where gene transcription is initiated [8], [9]. G-quadruplexes have also been found in the 5'-UTR region of the encoded mRNAs [10]-[12]. In addition, G-quadruplex also exists in bacteria, fungi and viruses [13], [14].

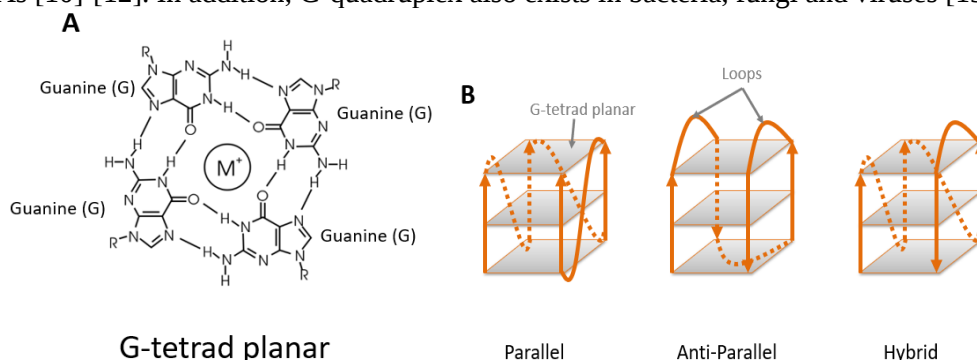


Figure 1. The formation of G-quadruplex. A) The G-quadruplex structure is formed in DNA or RNA in the presence of K^+ or Na^+ cations. B) G-quadruplexes with different topologies: parallel and non-parallel structures [15]

G-quadruplex formation plays an important role in many biological processes such as transcription, translation, and telomeric maintenance [3], [4]. During replication, single-stranded G-rich sequences are capable of forming G-quadruplexes, which inhibits replication, leading to genome instability. In human chromosomes, telomeres (approximately 100 to 280 nt) are favorable for the formation of G-quadruplexes and can inhibit telomerase activity, rendering telomeres unable to elongate [7]. Therefore, the presence of G-quadruplexes in the human genome is considered a novel molecular target for cancer therapeutic research [16]-[19].

A helicase, RHAU, has been shown to bind and specifically resolve G-quadruplex structures in the presence of ATP [20]. The mechanism of G-quadruplex recognition by RHAU has been identified from the structure of RHAU bound to G-quadruplex DNA, where RHAU recognizes G-tetrad via hydrophobic interactions, supplemented by electrostatic interactions between positively charged amino acids and negatively charged DNA phosphate groups [21]-[23]. G-quadruplex-specific function can be conjugated by binding RHAU to a functional protein [15], [24]-[27]. Previously, this method was applied to probe the G-quadruplex and develop a novel nuclease and ribonuclease for programmable DNA and RNA cleavage [24], [26]. In this study, a novel nuclease targeting G-quadruplex was genetically generated by combining the RHAU53 recognition motif with the Fok1 nuclease domain. RHAU53-Fok1 can specifically bind and cut double-stranded DNA next to it.

2. Materials and methods

2.1. Construction of plasmid RHAU53-Fok1

A novel nuclease was generated by incorporating RHAU53 into the catalytic protein domain of Fok1. DNA encoding for RHAU53-Fok1 was amplified using the gene of *RHAU53-Fok1* (synthesized from IDT company, Singapore) with a pair of primer ON1: 5'-ctg cta tgg agg tca ggt

atg cag cgg ttt ctt tac cag act cg-3' and ON2: 5'-gcc agg atc cga tgt ttc agg aca acc cgc tgc ta-3'. The PCR product was treated with 2 restriction enzymes *NcoI* and *BamHI* and inserted into treated pET-Duet1 at *NcoI* and *BamHI* sites by T4 DNA ligase, resulting in the generation of plasmid pRHAU53-Fok1.

Amino acids sequence of RHAU53-Fok1 protein:
 MGHHHHHHHHSMHPGHLKGREIGMWYAKKQGQKNKEAERQERAVVHMDERREEQIV
 QLLNSVQAKDVSGGGSGGGSGGGSGGGQLVKSELEKKSELRHKLKYVPHEYIELIEIAR
 NSTQDRILEMKVMEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGG
 YNLPIGQADEMQRYVEENQTRNKHINPNEWVKVYPSSVTEFKFLFVSGHFKGNYKAQL
 TRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEEVRRKFNNGEINF

2.2. Protein expression and purification

The plasmid pRHAU53-Fok1 coding RHAU53-Fok1 was transformed into *E. coli* train BL21 (DE3). The bacteria were cultured in LB (Luria-Bertani) medium consisting of 50 mg/L of ampicillin and the *E. coli* cells were grown at 37°C, shaking 180 round per minute to reach an OD₆₀₀ of 0.5-1. IPTG (Sigma Aldrich, St. Louis, USA) was then added to a final concentration of 0.5 mM. The *E. coli* cells were subsequently incubated for 12 hours at 165°C, shaking at 180 rpm before being harvested by centrifugation. The pellet of *E. coli* was resuspended into the lysis buffer Bugbuster protein extraction reagent (Merck, Germany) and 10 µg/ml *DnaseI*. The insoluble cell debris was then removed by centrifugation at 20000 rpm for 2 hours at 4°C. The soluble fraction flowed through the His-tag column. Subsequently, the column was then washed with 20 volumes of column with 20 mM Tris-HCl, 50 mM NaCl and 10 mM imidazole buffer, pH 7. The proteins were eluted with a buffer of 20 mM Tris-HCl, 50 mM NaCl and 200 mM imidazole. The imidazole in the protein solution was removed by centrifugation using the Amicon Ultra-15 centrifugal filter (EMD Milipore). The pure protein was isolated and analyzed by SDS-PAGE.

2.3. Cleavage of double stranded DNA

The artificial substrate D1 was generated by adding G-quadruplex (**tt[gggt]₄**) to a single strand of 42 bp dsDNA (5'-**tt[gggt]₄** aag act tca agg acg acg gca aca tcc tgg cgc aca agc tgc-3'). D2 as the negative control was generated without G-quadruplex (5'-aag act tca agg acg acg gca aca tcc tgg cgc aca agc tgc-3'). Both D1 and D2 were labelled with FAM dye for detection. Cleavage reaction was carried out at 37°C in 1x CutSmart® buffer (New England Biolabs, UK). Cleavage products were analyzed by the native gel electrophoresis (18%).

3. Results and discussion

3.1. Construction of plasmid

The *RHAU53-Fok1* gene fragment was amplified by PCR with a pair of ON1 and ON2. The PCR product analyzed by gel electrophoresis with 2% agarose showed a single band between the 800 bps and 1000 bps lines of the DNA ladder (Figure 2A), corresponding to the size of the *RHAU53-Fok1* gene fragment (858 bps). This shows that the target gene segment has been amplified successfully. PCR products are treated with restriction enzymes and inserted into host vectors pET-Duet1 that are also uncirculated with restriction enzymes. The ligation product was then transformed into *E. coli* strain and spread on the agar plate (ampicillin antibiotic-containing medium). Colonies grown on the petri were selected to be tested for their ability to contain the target gene vector by colony PCR. Electrophoresis results showed that 2 colonies had a band between the 1000 bps and 1500 bps lines of the DNA ladder, corresponding to the theoretical design size of 1045 bps (Figure 2B). The recombinant DNA was then confirmed by DNA sequencing (1st Base, Singapore). DNA sequencing result showed 100% similarity with the

theoretical sequence (unshown data), indicating that the recombinant vector carried the target gene *RHAU140-Fok1*.

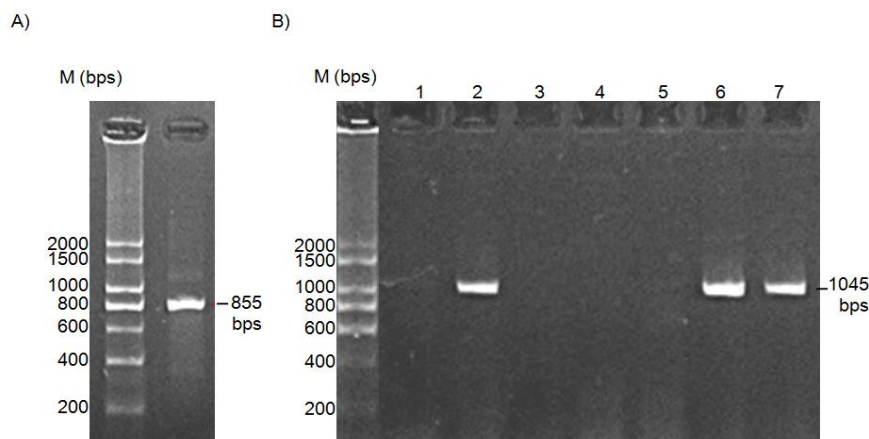


Figure 2. DNA electrophoresis: A) PCR product of *RHAU53-Fok1* gene (812 bps). (B) PCR products (1045 bps) of 7 random colonies from the culture plate, M (DNA Marker).

3.2. Protein expression and purification

The *RHAU53-Fok1* protein was expressed in *E. coli* strain BL21 (DE3) under the induction of 0.5 mM IPTG. Proteins were purified using the His column and the homogeneous proteins were then analyzed by SDS-PAGE. Electrophoresis result showed that the target protein band shifted near 30 kDa of the protein ladder, corresponding to a target protein of 31.6 kDa (Figure 3). The homogeneous protein was kept in -80 °C and ready for activity assay.

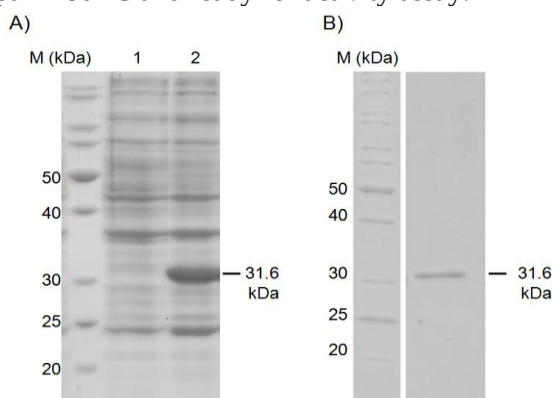


Figure 3. SDS-PAGE of *RHAU53-Fok1* protein: A) Protein expression in *E. coli* without (-) and with (+) 0.5 mM IPTG inducer. B) Homogeneous *RHAU53-Fok1* was collected by the affinity chromatography with His column, M (protein marker).

3.3. Cleavage of double stranded DNA

Cleavage of double stranded DNA by *RHAU53-Fok1* was performed with two artificial substrates D1 and D2 (negative control). A G-quadruplex TT(GGGT)₄ was formed into a single strand of 42-bp dsDNA substrate. Two substrates D1 and D2, with and without G4, were labeled with a fluorescent dye FAM at the 3' end, allowing DNA visualization. Nuclease *RHAU53-Fok1* selectively binds to D1 and cuts dsDNA next to it (Figure 4). *RHAU53-Fok1* showed multiple cleavage sites on the substrate providing a major fragment and several minor fragments, which is consistent with previously reported data for nuclease Fok1. In contrast, the addition of the same amount of *RHAU53-Fok1* to the negative control DNA (D2) showed small and nonspecific cleavage.

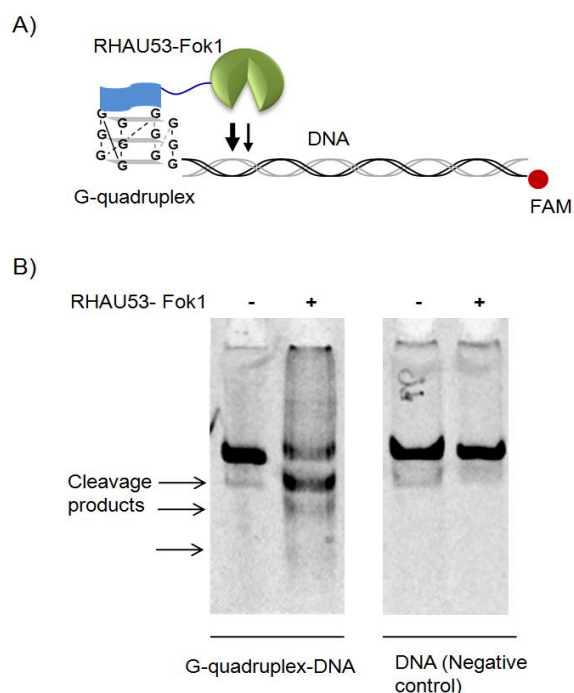


Figure 4. The recognition and cutting of double stranded DNA (with and without a G-quadruplex) by the RHAU53-Fok1 nuclease. A) Schematic representation of DNA recognition and cutting by enzyme. B) Cutting of double stranded DNA (D1) bearing G-quadruplex (5 μ M) and D2 (DNA without G-quadruplex as negative control) (5 μ M) by RHAU53-Fok1 (1 μ M) was carried out at 1 hour, 37°C in 1x CutSmart® buffer and analysed by native gel electrophoresis (18%).

Generation of a new nuclease has been considered as an alternative approach to study nucleic acid properties and functions, especially with long and intricate species. The type II restriction enzymes can recognize specific regions and cleave DNA at predetermined positions [28]. Furthermore, protein engineering technique are often straightforward owing to the gradual development of recombinant DNA technology. An increasing number of efforts are being made to generate artificially enzymes that can specifically cleave RNA and DNA such as combinations of specific DNA/RNA binding proteins known as a zinc finger into the catalytic domain of RNase HI or Fok1, generating fusion proteins that selectively cleaves the RNA strand or DNA at adjacent points of binding sites [29]-[31]. In this study, using the G-quadruplex-binding RHAU53 domain to generate a nuclease by incorporating it to Fok1. The fusion enzyme can recognize DNA substrates bearing G-quadruplex for cleavage at a specific site. This nuclease would provide a useful technique for mapping formation of G-quadruplex in the genome.

4. Conclusion

A novel nuclease RHAU53-Fok1 was generated by combining RHAU53 with the catalytic domain of Fok1. This fusion protein can selectively recognize and cut dsDNA bearing the parallel G-quadruplex motif at specific sites. RHAU53-Fok1 may provide a potential tool to probe or map the G-quadruplex formation in the genome. This work also proposes a general method to generate variant functional proteins to target G-quadruplex motifs in biological processes.

Acknowledgement

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 108.02-2017.305.

REFERENCES

- [1] M. Gellert, M. N. Lipsett, and D. R. Davies, "Helix Formation by Guanylic Acid," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 48, no. 12, pp. 2013-2018, 1962, doi: 10.1073/pnas.48.12.2013.
- [2] D. Sen and W. Gilbert, "Formation of Parallel 4-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis," *Nature*, vol. 334, no. 6180, pp. 364-366, 1988, doi: 10.1038/334364a0.
- [3] S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, and S. Neidle, "Quadruplex DNA: sequence, topology and structure," *Nucleic Acids Research*, vol. 34, no. 19, pp. 5402-5415, 2006, doi: 10.1093/nar/gkl655.
- [4] H. J. Lipps and D. Rhodes, "G-quadruplex structures: in vivo evidence and function," *Trends in Cell Biology*, vol. 19, no. 8, pp. 414-422, 2009, doi: 10.1016/j.tcb.2009.05.002.
- [5] N. Maizels and L. T. Gray, "The G4 genome," *PLoS Genet*, vol. 9, no. 4, 2013, Art. no. e1003468, doi: 10.1371/journal.pgen.1003468.
- [6] G. N. Parkinson, M. P. H. Lee, and S. Neidle, "Crystal structure of parallel quadruplexes from human telomeric DNA," *Nature*, vol. 417, no. 6891, pp. 876-880, 2002, doi: 10.1038/nature755.
- [7] D. J. Patel, A. T. Phan, and V. Kuryavii, "Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: Diverse higher order DNA and RNA targets for cancer therapeutics," *Nucleic Acids Research*, vol. 35, no. 22, pp. 7429-7455, 2007, doi: 10.1093/nar/gkm711.
- [8] S. Balasubramanian, L. H. Hurley, and S. Neidle, "Targeting G-quadruplexes in gene promoters: a novel anticancer strategy?," *Nat Rev Drug Discov*, vol. 10, no. 4, pp. 261-275, 2011, doi: 10.1038/nrd3428.
- [9] D. Lin *et al.*, "A benzindole substituted carbazole cyanine dye: a novel targeting fluorescent probe for parallel c-myc G-quadruplexes," *Analyst*, vol. 140, no. 16, pp. 5772-5780, 2015, doi: 10.1039/c5an00866b.
- [10] C. K. Kwok, Y. Ding, S. Shahid, S. M. Assmann, and P. C. Bevilacqua, "A stable RNA G-quadruplex within the 5'-UTR of Arabidopsis thaliana ATR mRNA inhibits translation," *Biochem J*, vol. 467, no. 1, pp. 91-102, 2015, doi: 10.1042/BJ20141063.
- [11] S. P. P. Pany, M. Sapra, J. Sharma, V. Dhamodharan, S. Patankar, and P. I. Pradeepkumar, "Presence of Potential G-Quadruplex RNA-Forming Motifs at the 5'-UTR of PP2A α mRNA Repress Translation," *Chembiochem*, 2019, doi: 10.1002/cbic.201900336.
- [12] R. Simone, P. Fratta, S. Neidle, G. N. Parkinson, and A. M. Isaacs, "G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome," *Febs Lett*, vol. 589, no. 14, pp. 1653-1668, 2015, doi: 10.1016/j.febslet.2015.05.003.
- [13] D. Ji, M. Juhas, C. M. Tsang, C. K. Kwok, Y. Li, and Y. Zhang, "Discovery of G-quadruplex-forming sequences in SARS-CoV-2," *Brief Bioinform*, vol. 22, no. 2, pp. 1150-1160, 2021, doi: 10.1093/bib/bbaa114.
- [14] C. Zhao *et al.*, "Targeting RNA G-Quadruplex in SARS-CoV-2: A Promising Therapeutic Target for COVID-19?," *Angew Chem Int Ed Engl*, vol. 60, no. 1, pp. 432-438, 2021, doi: 10.1002/anie.202011419.
- [15] L. T. A. Nguyen and D. T. Dang, "RHAU Peptides Specific for Parallel G-Quadruplexes: Potential Applications in Chemical Biology," *Mol Biotechnol*, vol. 65, no. 3, pp. 291-299, 2023, doi: 10.1007/s12033-022-00552-7.
- [16] J. L. Mergny and C. Helene, "G-quadruplex DNA: A target for drug design," *Nature Medicine*, vol. 4, no. 12, pp. 1366-1367, 1998, doi: 10.1038/3949.
- [17] H. Y. Han and L. H. Hurley, "G-quadruplex DNA: a potential target for anti-cancer drug design," *Trends in Pharmacological Sciences*, vol. 21, no. 4, pp. 136-142, 2000, doi: 10.1016/S0165-6147(00)01457-7.
- [18] S. M. Kerwin, "G-quadruplex DNA as a target for drug design," *Current Pharmaceutical Design*, vol. 6, no. 4, pp. 441-471, 2000, doi: 10.2174/1381612003400849.
- [19] S. M. Kerwin, B. Mamiya, C. Brian, T. Fletcher, J. T. Kern, and P. W. Thomas, "G-quadruplex DNA as a target for drug discovery: Design of telomerase inhibitors based on G-quadruplex DNA structure and dynamics," *Abstracts of Papers of the American Chemical Society*, vol. 219, pp. U6-U6, Mar 26 2000.

- [20] S. D. Creacy, E. D. Routh, F. Iwamoto, Y. Nagamine, S. A. Akman, and J. P. Vaughn, "G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates," *J Biol Chem*, vol. 283, no. 50, pp. 34626-34634, 2008, doi: 10.1074/jbc.M806277200.
- [21] B. Heddi, V. V. Cheong, H. Martadinata, and A. T. Phan, "Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 31, pp. 9608-9613, 2015, doi: 10.1073/pnas.1422605112.
- [22] M. C. Chen *et al.*, "Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36," *Nature*, vol. 558, no. 7710, pp. 465-469, 2018, doi: 10.1038/s41586-018-0209-9.
- [23] B. Heddi, V. V. Cheong, E. Schmitt, Y. Mechulam, and A. T. Phan, "Recognition of different base tetrads by RHAU (DHX36): X-ray crystal structure of the G4 recognition motif bound to the 3'-end tetrad of a DNA G-quadruplex," *J Struct Biol*, vol. 209, no. 1, 2020, Art. no. 107399, doi: 10.1016/j.jsb.2019.10.001.
- [24] D. T. Dang, L. T. A. Nguyen, T. T. T. Truong, H. D. Nguyen, and A. T. Phan, "Construction of a G-quadruplex-specific DNA endonuclease," *Chem Commun (Camb)*, vol. 57, no. 37, pp. 4568-4571, 2021, doi: 10.1039/d0cc05890d.
- [25] D. T. Dang and A. T. Phan, "Development of Fluorescent Protein Probes Specific for Parallel DNA and RNA G-Quadruplexes," *Chembiochem*, vol. 17, no. 1, pp. 42-45, 2016, doi: 10.1002/cbic.201500503.
- [26] D. T. Dang and A. T. Phan, "Development of a ribonuclease containing a G4-specific binding motif for programmable RNA cleavage," *Sci Rep*, vol. 9, no. 1, p. 7432, 2019, doi: 10.1038/s41598-019-42143-8.
- [27] T. T. Thi Truong, C. Cao, and D. T. Dang, "Parallel G-quadruplex-mediated protein dimerization and activation," *RSC Adv*, vol. 10, no. 50, pp. 29957-29960, 2020, doi: 10.1039/d0ra06173e.
- [28] K. Y. Hu, J. A. Wu, H. C. Kao, Y. T. Liu, and S. H. Pai, "Isolation and characterization of a newly identified type II restriction endonuclease from a local *Streptomyces* sp. in Taiwan," *Applied Biochemistry and Biotechnology*, vol. 73, no. 2-3, pp. 231-241, 1998, doi: 10.1007/Bf02785658.
- [29] Y. G. Kim, J. Cha, and S. Chandrasegaran, "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain," *Proc Natl Acad Sci U S A*, vol. 93, no. 3, pp. 1156-60, 1996, doi: 10.1073/pnas.93.3.1156.
- [30] F. D. Urnov, J. C. Miller, Y. L. Lee, C. M. Beausejour, M. J. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, and M. C. Holmes, "Highly efficient endogenous human gene correction using designed zinc-finger nucleases," *Nature*, vol. 435, no. 7042, pp. 646-651, 2005, doi: 10.1038/nature03556.
- [31] J. P. Guilinger, D. B. Thompson, and D. R. Liu, "Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification," *Nat Biotechnol*, vol. 32, no. 6, pp. 577-582, 2014, doi: 10.1038/nbt.2909.