

Computational Identification of Inhibitors Against *Porphyromonas gingivalis* Virulence Factors in Oral Squamous Cell Carcinoma

Nandini

M.Sc Student, Centre of Biotechnology (Bioinformatics)

Maharshi Dayanand University, Rohtak

dr.nandni2k1@gmail.com

Abstract: Oral squamous cell carcinoma (OSCC) is a major global malignancy, accounting for about 90% of oral cancer cases and ranking eighth in cancer incidence worldwide. While tobacco, alcohol, nicotine, and HPV are established risk factors, nearly 15% of OSCC cases occur without these, suggesting other contributors. Recent studies implicate bacterial infections, particularly *Porphyromonas gingivalis*, a periodontal pathogen, as an oncogenic factor in OSCC. *P. gingivalis* influences multiple cancer hallmarks, disrupting tumor suppression and apoptosis pathways while activating oncogenic signaling. Its virulence factors—Fimbriae (FimA), nucleoside diphosphate kinase (NDK), Lysine gingipain (Kgp), and Arginine gingipain (RgpB)—play crucial roles in OSCC pathogenesis. FimA promotes uncontrolled cell proliferation, NDK inhibits apoptosis and enables immune evasion, and gingipains (Kgp, RgpB) enhance metastasis. This in-silico study employed molecular modeling, high-throughput virtual screening, molecular docking, 3D-QSAR pharmacophore modeling, and molecular dynamics simulations to identify potential inhibitors of these virulence factors. Homology modeling was used to predict the 3D structure of NDK. For targets lacking known inhibitors (FimA and NDK), structure-based drug design was applied, while both structure- and ligand-based approaches were used for Kgp and RgpB. Natural and anticancer compound libraries were screened to identify potential leads. Docking scores, binding energies, and ADMET analyses guided the selection of top compounds with favorable pharmacokinetic properties. Molecular dynamics simulations provided insights into the stability and interactions of receptor-ligand complexes. Protocatechuic acid, Grossamide K, and shogaulsulfonic acid C emerged as promising inhibitors with potential therapeutic applications.

Keywords: Oral squamous cell carcinoma (OSCC); *Porphyromonas gingivalis*; Virulence factors; Molecular docking; 3D-QSAR modelling

I. INTRODUCTION

Cancer is an uncontrolled augmentation of abnormal cells in the body. Cancer developed from malignant tumors which can invade and spread the nearby cells. Oral cancer is among the most common cancer in the world. It is one of the most chronic head and neck malignancies and contributes significantly to the global cancer burden. According to the global cancer statistics report oral cancer has a rank of eight worldwide (Sung et al., 2021). Both the incidence and recurrence rates of oral cancer are high (Sarode et al., 2020). According to estimates, 377,713 new cases of oral cancer are reported each year, having a mortality rate of 177,75 (Sung et al., 2021). The progression of oral cancer is determined by the stages, which are I, II, III, and IV. In stage I, the tumor is under 2 cm in size and the cancer has not yet spread to the lymph nodes. In stage II, the tumor is between 2 and 4 cm in size and the cancer has not yet spread to the lymph nodes. In stage III, the tumor is either larger than 4 cm in size and has spread to one lymph node but not to the other parts. In stage IV, the tumor is any size and the cancer cells have spread to nearby tissues, the lymph node, and other parts of the body. Oral cancers are categorized into OSCC, verrucous carcinoma, Kaposi sarcoma, minor salivary gland carcinoma, and lymphomas. About 90% of cases of oral cancer are OSCC. Early stage OSCC usually



remains asymptomatic (Hassona et al., 2018). OSCC has a significant risk of local invasion and metastasis and the five year survival rate is about 50% (Lorenzo-Pouso et al., 2023). The treatment and prevention of OSCC have always been a major cause of cancer concern. This is mainly because OSCC is often detected late into metastasis due to the pain not developing until the later stages. The treatment options tend to be limited in OSCC at later stages due to cancer metastasis. Hence, OSCC prevention and care are complicated. Therefore, it is essential to identify OSCC at an early stage and also understand the precise molecular mechanism. Early detection of OSCC is thought to be the most efficient approach to preventing death and morbidity (Scully, 2014). Tumor location and diagnosis time are important components of the survival rate for patients diagnosed with OSCC (Siegel et al., 2009). Tobacco, alcohol, nicotine, and human papillomavirus (HPV) has been reported as risk factor for OSCC (Tenore et al., 2020). However, around 15% of cases of OSCC occurred each year without any of these risk factors (Perera et al., 2016). So, there is a need to investigate the potential risk factors for this 15% of cases. *Helicobacter pylori* (*H.pylori*) 1994 was recognized as a type I carcinogen by WHO and has been associated with a risk factor of gastric cancer (Correa, 1995). This remarkable evidence has opened up a new direction for investigating the role of bacteria in the etiology of cancer. This has also triggered the research for understanding the role of the oral microbiome in Oral Cancer. The role of the oral microbiome in modulating the carcinogenic process is an emerging concept in oral cancer biology. Understanding bacterial involvement and influence in cancer development may provide a new direction and therapeutic strategy for cancer prevention. The oral cavity is inhabited by a variety of microbes, which predominantly includes bacteria with fungi and viruses constituting a small proportion. Microbes in the oral cavity occupy different sub-habitats such as teeth, tongue, gingival crevices, saliva, buccal mucosa, etc. Over 700 bacterial species are present in the human oral cavity (Johny et al., 2022). These bacterial species coexist with other microbes in the host oral cavity, and their balance is critical for maintaining a healthy physiological environment of the oral cavity (Liu et al., 2019). A variety of external and internal events alter the balance of the oral microbiota, resulting in dysbiosis of the oral environment. The dysbiosis environment promotes pathogenic species dominance, which leads to diseases like dental caries and periodontitis (Michael A. Curtis et al., 2011).

II. LITERATURE REVIEW

P. gingivalis is a pathogenic bacteria found in the human oral cavity that is gram-negative, immobile, rod-shaped, anaerobic, and asaccharolytic. It is a significant pathogen in periodontal disease and has been linked to elevated cancer risk (X. B. Liu et al., 2019). *P. gingivalis* is a keystone pathogen, meaning it has a large impact even when it is in low abundance (Olsen & Yilmaz, 2019). *P. gingivalis* has a disproportionately strong effect on dysbiosis compared to its abundance. These bacteria modulate the composition of the microbiome as well as the host response (Darveau et al., 2012). *P. gingivalis* produces virulence factors such as fimbriae (FimA), cysteine proteases (Arg-gingipain (Rgp), Lys-gingipain (Kgp)), lipopolysaccharide (LPS), and nucleoside diphosphate kinase (NDK) to initiate an infection. Several studies have revealed the pathogenic pathways related to virulence factors that assist *P. gingivalis* to survive in the host cell (Gao et al., 2016; Lamont et al., 2022; X. B. Liu et al., 2019; Olsen & Yilmaz, 2019). Fimbriae help the invasion of the oral epithelial cell by engaging with the host cell's 1 integrin and triggering cytoskeleton rearrangement (Yilmaz et al., 2002). On the cell surface of *P. gingivalis*, there are two types of fimbriae. Mfa1 fimbriae and Fimbrillin (FimA). Based on nucleotide sequences, FimA is categorized into six types: I-V and Ib. fimA type I (ATCC 33277), fimA genotype Ib (HG1691), fimA types II (A7A1-28 or ATCC 53977), fimA types III (BH 6/26), and fimA genotype IV (strain W83 or ATCC BAA308) (Fujiwara et al., 1993; Zheng et al., 2011).

P. gingivalis infected OSCC-derived cell lines (SCC-25 and BHY) increased the expression of B7-H1 and B7-DC receptors, both of which impair T-cell survival (Groeger et al., 2011).

Inaba et al. investigated the effect of *P. gingivalis* on MMP9 synthesis in two OSCC cell lines (SAS and Ca9-22) because MMP9 is involved in tumor cell invasion and metastasis. *P. gingivalis* generates proMMP9 (an inactive form of MMP9) by activating the ERK1/2-Ets1, p38/HSP27, and NF- κ B pathways, according to the researchers. These pathways were activated by overexpression of PAR2 and PAR4 via gingipains. Furthermore, gingipains use proteolytic activity to convert proMMP9 to MMP9. Polyphenol inhibitors, which inhibit both gingipains and ProMMP9, were also found to reduce *P. gingivalis* invasiveness. (Inaba et al., 2014, 2015, 2016).



P.gingivalis infection enhanced the invasiveness and tumorigenic potential of OSCC cell lines (SCC25, OSC20, and SAS) by increasing the levels of MMP1, 2, 7, and IL-8, according to Ha et al. Another study conducted by the same group discovered that after being infected with *P.gingivalis*, OSCC (Ca922) cell lines changed shape and increased the expression of TNF, Snail, Slug, and Twist. Upregulation of MMP1, MMP10, and IL8 was also detected, all of which contributed to invasiveness. Stem cell features were also found, including overexpression of CD44 and CD133 (Ha et al., 2015, 2016). Cho et al. obtained similar results on the OSCC cell line (YD10B) (Cho et al., 2018).

According to Abdulkareem et al., *P.gingivalis* increased the levels of MMPs, vimentin, Twist, Snail, and Slug while decreasing the levels of E-cadherin in the OSCC (H400) cell line. EGF and TNF levels were also up considerably (Abdulkareem et al., 2018).

In 2017, Geng et al. discovered that prolonged *P.gingivalis* exposure boosted the proliferation, migration, and invasion of healthy gingival cells (HIOEC). MMP9, TNF, TLR, and NF-kB overexpression elevated carcinogenic properties (Geng et al., 2017).

In the OSCC cell line (SCC-25), overexpression of genes such as IKBKB, MAPK14, MAPK8, and JUN was detected. These genes are part of the NF-kB and MAPK downstream signaling pathways, which help cancer spread (Groeger et al., 2017).

Mir-203, AP-1, and cyclinD1 were found to be overexpressed in a *P.gingivalis* infected OSCC cell line (Tca8113). mir-21 (C. Chang, Wang, et al., 2019).

Kong et al. used IHC and PCR to determine the presence of *P.gingivalis* in cancer patient samples. *P.gingivalis* was found in high numbers in OSCC samples, followed by ESCC. The IHC and PCR results of OSCC demonstrated that *P.gingivalis* was present in 60% and 50% of the cases, respectively. Based on these findings, they discovered that the presence of *P. gingivalis* is related to severe OSCC lymph node metastases (Kong et al., 2021).

Using the RT-PCR approach, Kaliamoorthy and colleagues discovered that *P. gingivalis* was present in larger quantities in samples from OSCC patients, but not in normal samples (Kaliamoorthy et al., 2021).

According to Chen et al., OSCC patients had a higher level of *P.gingivalis* in their saliva than normal patients. They used PCR to assess FimA genotypes in OSCC samples and discovered that OSCC samples exhibit higher percentages of *P.gingivalis* genotypes I+Ib, II, and IV, with 21.1%, 21.1%, and 31.6%, respectively. (Chen et al., 2021).

III. RESEARCH METHODOLOGY

Determining the 3D structure of targets is one of the crucial steps in drug design. Homology modeling is one of the computational methods used to predict the 3D structure of the protein. Homology modeling has a wide range of uses in the drug discovery process. Homology modeling has made drug development faster, easier, less expensive, and more practical (Muhammed & Aki-Yalcin, 2019). The 3D structure of NDK has not yet been determined. Although the functional domain regions of both Arg-gingipain and Lys-gingipain were present the structure of the whole protein has not been reported. Two strains of *P.gingivalis* have been selected for this study i.e. ATCC33277 and W83. The sequence identity and similarity between NDK of strain ATCC33277 and W83 were 98.9% and 99.1%. The sequence similarity between Arg-gingipain was 99.1% and 99.3% for both strains. Furthermore, the sequence similarity between Lys-gingipain of both strains was 94.9% and 97.4%. Since the identity and similarity were >90%. Therefore to model the protein, sequences were taken only from the W83 strain. The 3D structures of these targets were prepared using two different approaches of homology modeling i.e. multi template and single template.



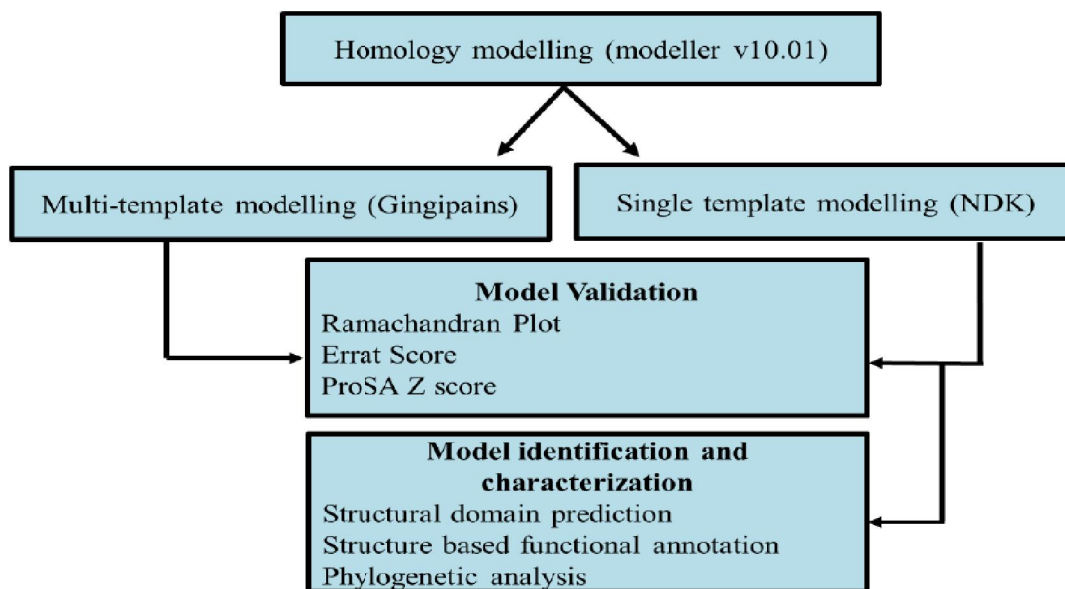


Figure 1. Workflow of Homology modelling

The amino acid sequences for Arg-gingipain and Lys-gingipain were derived from UniProt using UniProt IDs P95493 and Q51817. The sequence length of Arg-gingipain was 736 residues whereas the length of Lys-gingipain was 1732 residues. As the length of the whole proteins was long, blast results showed insufficient identity, similarity, and query coverage to a single protein from PDB. However, the Blast analyses showed that certain fragments of the protein had a lot of similarities with their domain structures and other proteins. Therefore multi-template approach was used to generate the 3D structure of proteins. The amino acid sequences of whole Arg-gingipain Lys-gingipain were taken to search the homologs from PDB. A total of three homologs templates were identified for Arg-gingipain from the blast search which covers the entire target sequences. Similarly for Lys-gingipain total of six homologs templates were identified but these templates didn't cover the entire sequences of the query. Therefore the regions that were not covered by the templates were trimmed and modeled using the I-TASSER server, and the resulting model structures were used as a template together with the other homologs to model the full structure of Lys- gingipain. Therefore, nine templates were employed to create the 3D structure of Lys-gingipain. The 3D structures of all the templates were downloaded from PDB. After retrieving the templates for both targets, multiple sequence alignments were carried out between targets and templates. Furthermore, a total of 50 models were generated for both targets, and the best models were selected based on the lowest dope score. The loop modeling tool of modeller v10.1 was used to further optimize the chosen models. The Yasara energy minimization tool was used to minimize the energy of the model structures after loop refinement to eliminate steric conflicts and stabilize the model structure (Land & Humble, 2018)., Finally, the reliability of the models was evaluated using Errat, Ramachandran plot, and the ProSa web server.

IV. RESULTS AND DISCUSSION

The BLAST server provided ten hits for RgpB out of which three hits were chosen as a template based on the query coverage, percent identity, and E-value. Identities of the hits were 99.77, 98.65 and 97.01 and query coverage of 58%, 59%, and 21%. The structures have the following PDB IDs -4IEF, 1CVR, and 5AG8. The three-dimensional structure of RgpB was predicted using the above-mentioned PDB IDs as templates, as illustrated in (Figure 10A). For Kgp a total of twelve hits provided by the blast server out of these six hits were chosen as templates having PDB IDs- 5MUN, 4RBM, 4ITC, 3KM5, 3M1H and 5HFS based on the query coverage, percent identity and E-value. The identity percentage of these templates was 99.06, 99.13, 100, 99.13, 99.43, and 51 and query coverage was 12%, 26%, 30%, 29%, and 4%.



These templates were not enough to model the entire protein because all these templates were not able to cover the entire sequence of Kgp. The missing sequences were trimmed and modeled using I-TASSER and they were named template3, template6, and template7. Using these templates the Kgp was modeled (Figure 2A). A total of ten models were predicted for both Kgp and RgpB, and the best models for each were chosen based on the lowest dope score. Furthermore, loop modeling of the selected models was performed to refine the loop regions of the model structure. Moreover, the accuracy of the predicted models was evaluated using the following: Errat score of RgpB was 95% (2C) whereas for Kgp the Errat score was 53.29% (Figure 2C) the scores for both the models were >50 and considered a high-quality model (Li & Wang, 2007), The Ramachandran plot analysis of RgpB showed that 94.7% residues were there in in the allowed regions and 0.6% residues were present in the disallowed region (Figure 1B). Since more than 90% of the residues of the model were found in the allowed region, therefore the model structure comes under the good quality structure (Yadav et al., 2013). The model Kgp showed 81.9% in the allowed regions and 3% residues in the disallowed regions (Figure 2B). Furthermore, the ProSA Z score of the RgpB model was -9.59 which falls within the range of the X-ray structures score (Figure 2D).

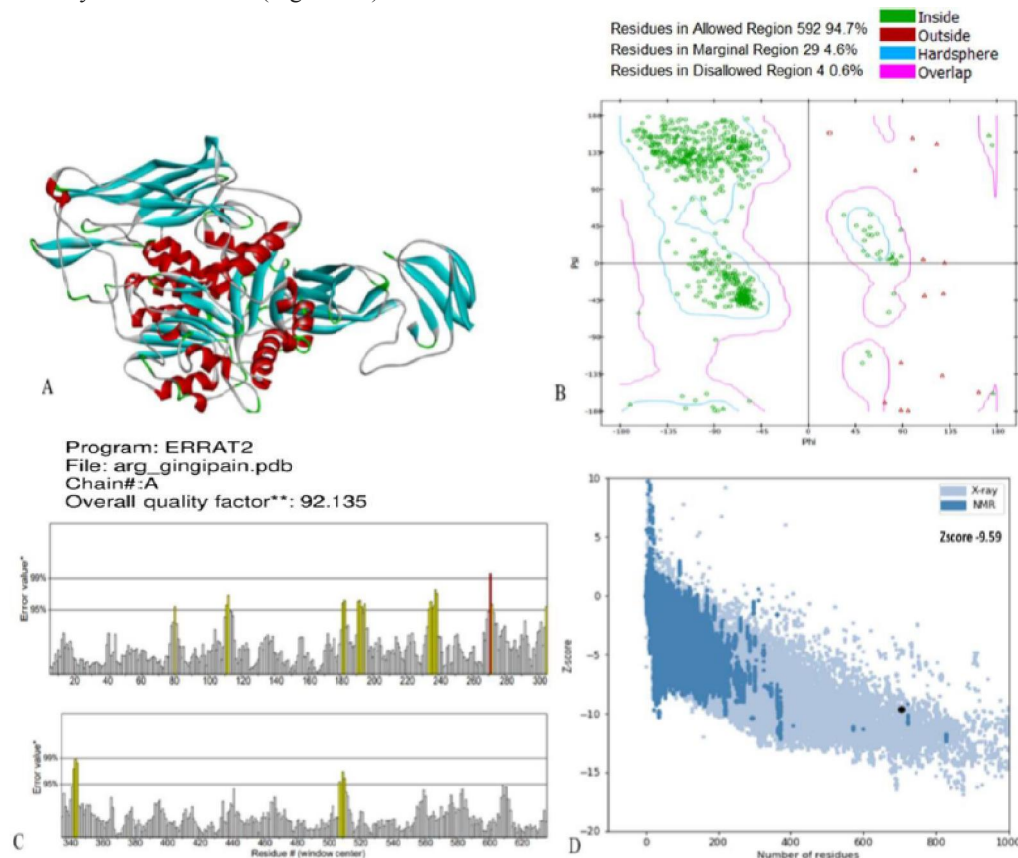


Figure 2. 3D structure and validation of protein model (A) RgPB model predicted by using modellerv10.1 (B) Ramachandran plot generated through Discovery Studio (C) Overall quality by Errat (D) Z score by ProSAWeb obtained from ProSA webserver.



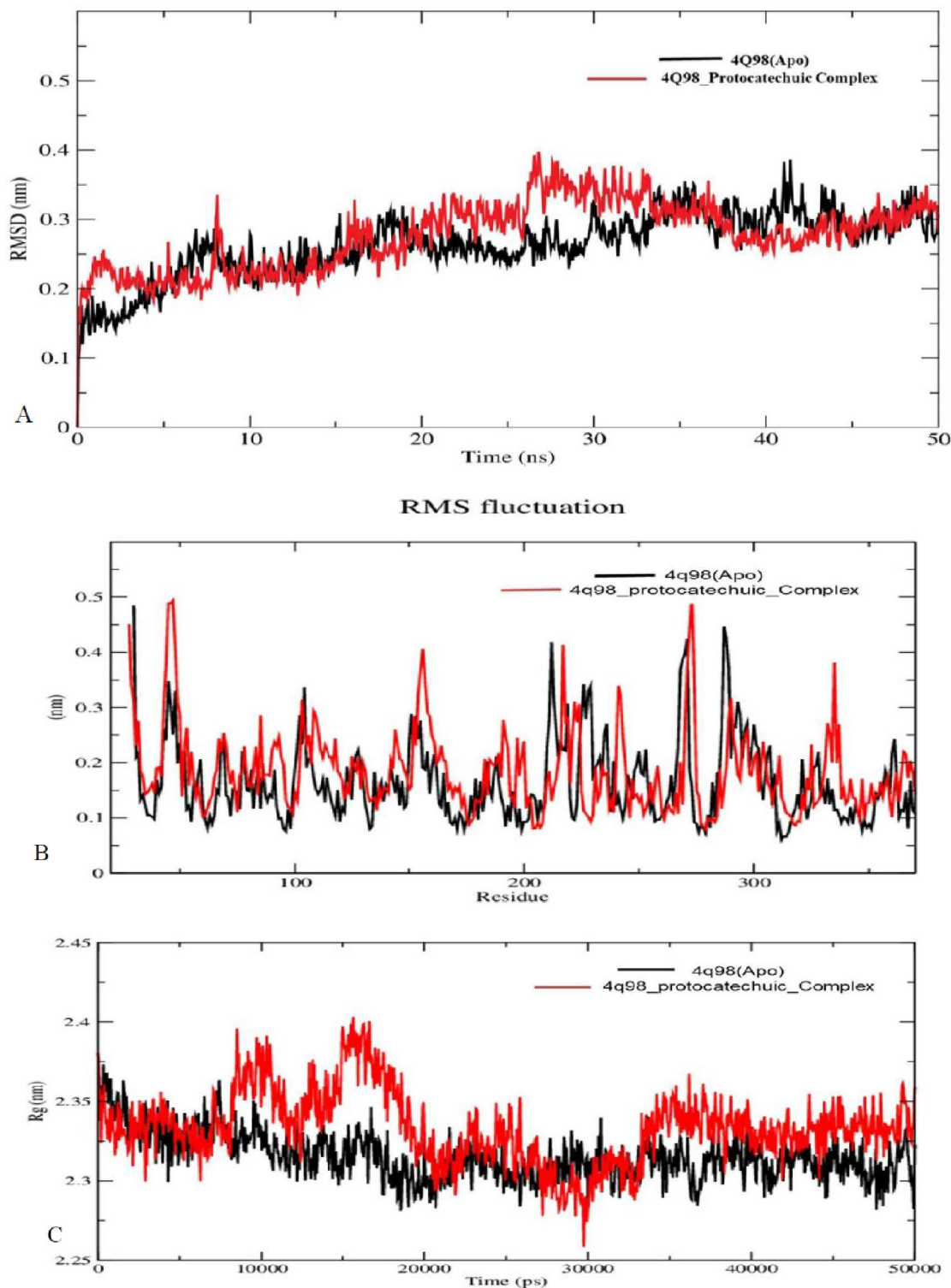


Figure 3 (A) Comparative RMSD of apo-protein (FimA) (4Q98) and protein-ligand complex (B) Comparative RMSF of apo-protein and protein-ligand complex (C) Comparative RG of apo-protein and protein-ligand complex.



Root-Mean-Square Deviation (RMSD) plots for Apo protein FimA (4Q98) and FimA- protocatechuic complex (Figure 3A) revealed intriguing insights into the behavior of the protein-ligand complex during a 50ns molecular dynamics simulation. Throughout the simulation, it was evident that the complex exhibited more significant deviations in its structure when compared to the Apo protein until approximately 38ns. Subsequently, a notable phenomenon occurred as the complex reached a plateau and achieved equilibrium, indicating a stabilization of the protein-ligand complex over time. After the ligand binding the active site residues showed less fluctuation (Figure 3B), suggesting the formation of strong and persistent interactions between the protein and the ligand. The Rg plot (Figure 3C) further contributed to our understanding of the complex's behavior. Fluctuations in the Rg curve were observed until approximately 39ns, indicating conformational changes in the protein structure following ligand binding. However, after this point, the Rg curve remained stable for the remainder of the 50ns simulation, suggesting that the protein achieved a relatively consistent and energetically favourable conformation.

V. CONCLUSION

Oral squamous cell carcinoma (OSCC) remains a major global health challenge, often diagnosed at advanced stages with high mortality rates. Recent studies have revealed the oncogenic role of *Porphyromonas gingivalis*, a periodontal pathogen, in OSCC progression through its virulence factors—FimA, NDK, Kgp, and RgpB—which influence key cancer hallmarks such as uncontrolled cell growth, apoptosis inhibition, and metastasis. This in-silico investigation aimed to identify natural inhibitors targeting these virulence factors using structure-based and ligand-based drug design approaches. Homology modeling of NDK (PG1018) revealed its structural similarity to known NDK proteins, establishing it as a viable therapeutic target. Virtual screening, molecular docking, and 3D-QSAR pharmacophore modeling were employed to identify potent inhibitors with strong binding affinities and favorable pharmacokinetic properties. ADMET analysis confirmed their drug-likeness, while molecular dynamics simulations validated the stability of protein-ligand interactions. Among screened compounds, Protocatechuic acid, Grossamide K, and Shogasulfonic acid C demonstrated strong inhibitory potential against FimA, NDK, and gingipains (Kgp and RgpB). These findings suggest that natural anticancer compounds can serve as safe, effective therapeutic agents against *P. gingivalis*-associated OSCC. Overall, this computational study provides a strong foundation for future experimental validation and the development of targeted, nature-derived treatments for OSCC.

REFERENCES

- [1] H. Inaba, H. Sugita, M. Kuboniwa, S. Iwai, M. Hamada, T. Noda, I. Morisaki, R. J. Lamont, and A. Amano, "Involvement of protease-activated receptor 4 in over-expression of matrix metalloproteinase-9 induced by *Porphyromonas gingivalis*," *Cell. Microbiol.*, vol. 17, no. 4, pp. 533–545, Apr. 2015.
- [2] N. H. Ha, J. H. Oh, S. H. Kim, H. J. Lee, H. J. Park, and S. H. Yeo, "Prolonged and repetitive exposure to *Porphyromonas gingivalis* increases aggressiveness of oral cancer cells by promoting acquisition of cancer stem cell properties," *Tumour Biol.*, vol. 36, no. 12, pp. 9947–9960, Dec. 2015.
- [3] S. Groeger, F. Jarzina, E. Domann, and J. Meyle, "*Porphyromonas gingivalis* activates NF- κ B and MAPK pathways in human oral epithelial cells," *BMC Immunol.*, vol. 18, no. 1, p. 1, Jan. 2017.
- [4] F. Geng, X. Zhang, M. Dong, Y. Jin, S. Liu, and X. Wang, "Persistent exposure to *Porphyromonas gingivalis* promotes proliferative and invasion capabilities and tumorigenic properties of human immortalized oral epithelial cells," *Tumour Biol.*, vol. 39, no. 12, pp. 1–13, 2017.
- [5] B. H. Woo, H. J. Park, J. H. Kim, et al., "Oral cancer cells sustainedly infected with *Porphyromonas gingivalis* exhibit resistance to Taxol and have higher metastatic potential," *Oncotarget*, vol. 8, no. 50, pp. 86558–86572, 2017.
- [6] Q. Chen, J. Shan, J. Zhu, et al., "Salivary *Porphyromonas gingivalis* predicts outcome in oral squamous cell carcinomas: a cohort study," *BMC Oral Health*, vol. 21, p. 456, 2021.
- [7] J. Kong, Y. Yuan, H. Wang, et al., "Frequencies of *Porphyromonas gingivalis* detection in oral-digestive tract tumors and association with clinicopathological features," *Pathol. Oncol. Res.*, 2021.
- [8] B. K. Sobocki, C. A. Basset, B. Bruhn-Olszewska, et al., "Molecular mechanisms leading from periodontal disease to cancer," *Int. J. Mol. Sci.*, vol. 23, art. 970, 2022.



- [9] T.-J. Li, J. Wang, and coauthors, "Periodontal pathogens: a crucial link between periodontal disease and systemic cancers — mechanisms and implications," *Front. Microbiol.*, 2022.
- [10] S. Singh, "Porphyromonas gingivalis in oral squamous cell carcinoma," *Cancer Treat. Res. Commun.*, vol. 31, p. 100590, 2022.
- [11] A. K. Kylmä, "Prognostic role of Porphyromonas gingivalis gingipain expression in head and neck cancers," *Anticancer Res.*, vol. 42, no. 11, pp. 5415–5422, Nov. 2022.
- [12] A. Radaic, P. Kamarajan, A. Cho, et al., "Biological biomarkers of oral cancer," *Periodontol. 2000*, vol. 96, no. 1, pp. 250–280, Oct. 2024.
- [13] A. Balapure, A. et al., "Early detection of oral cancer biomarkers using microfluidic point-of-care devices," *Anal. Methods*, 2024.
- [14] B. K. Sobocki et al., "Molecular mechanisms and pathways linking periodontal pathogens to carcinogenesis: review and perspectives," *MDPI Int. J. Mol. Sci.*, 2022.
- [15] C. Zenobia and R. J. Hajishengallis, "Porphyromonas gingivalis virulence factors involved in manipulation of the host immune response," *Front. Immunol.*, review (2015)

