Depamede dkk. (2024) Vol. 12 No. 3: 233-250

Chimica et Natura Acta

p-ISSN: 2355-0864 e-ISSN: 2541-2574

Homepage: http://jurnal.unpad.ac.id/jcena

In *Silico*-Predicted Cytokeratin Fragment 21-1 (Cyfra 21-1) Immunogenic Epitopes for The Early Detection of Nasopharyngeal Carcinoma

Brian Umbu Rezi Depamede[1], Bachti Alisjahbana[2], Sulaiman Ngongu Depamede[3], Sanny Hafidhoh Siti Nururrohmah[1], Ilham Tryawan[1], Muhammad Yusuf[4]

[1]Department of Biotechnology, Graduate School, Universitas Padjadjaran, Jl. Dipati Ukur No. 35, Bandung, 40132, West Java, Indonesia.
[2]Department of Internal Medicine, Faculty of Medicine, Universitas Padjadjaran, Hasan Sadikin General Hospital, Bandung, Jl. Pasteur No. 38 Bandung. Kel. Pasteur Kec. Sukajadi, 40161, West Java, Indonesia.
[3]Faculty of Animal Science, Mataram University, Jln. Majapahit No. 62 Mataram, NTB, 83125, Indonesia
[4]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jalan Raya Bandung – Sumedang km. 21, Sumedang 45363, West Java, Indonesia.
*Corresponding author: m.yusuf@unpad.ac.id

DOI: https://doi.org/10.24198/cna.v12.n3.51749

Abstract: Nasopharyngeal Carcinoma, an aggressive cancer in the head and neck region is mostly diagnosed at an advanced stage and thus has a poor prognosis. Cytokeratin Fragment 21-1 (CYFRA 21-1) is one of the oncomarkers discovered in saliva. For developing a diagnostic kit, CYFRA 21-1 antibodies and immunogens are needed. Natural CYFRA 21-1 immunogen is difficult to obtain, so it must be made in-silico using bioinformatics. We aim to predict CYFRA 21-1 immunogenic epitope to produce antibodies polyclonal against CYFRA 21-1, which can be used to develop NPC diagnostics. The immunogenic epitopes were predicted and chosen based on antigenicity, surface accessibility, and hydrophilicity, then characteristic was analyzed and evaluated. The epitope candidates were compared with other saliva protein biomarkers to find if there were cross-reactions. CYFRA 21-1 consists of 57 amino acids, where two immunogenic epitopes (C3 and D2) were chosen. The Ramachandran Plot of both epitopes shows that 100% of the amino acids were in the favoured area. Epitopes C3 and D2 have no cross-reaction with other protein biomarkers. The predicted immunogenic epitopes have the potential as antigen to produce antibodies for developing saliva-based immunodiagnostics to early diagnose NPC patients.

Keywords: antibody, cancer, cytokeratin fragment 21-1, epitope, NPC, saliva

Abstrak: *Karsinoma Nasofaring, kanker agresif di daerah kepala dan leher sebagian besar didiagnosis pada stadium lanjut sehingga memiliki prognosis yang buruk. Cytokeratin Fragment 21-1 (CYFRA 21-1) adalah salah satu penanda kanker yang ditemukan dalam saliva. Untuk mengembangkan kit diagnostik, diperlukan antibodi dan imunogen CYFRA 21-1. Imunogen CYFRA 21-1 alami sulit diperoleh, sehingga harus dibuat secara in-silico menggunakan bioinformatika. Kami bertujuan untuk memprediksi epitop imunogenik CYFRA 21-1 untuk menghasilkan antibodi poliklonal terhadap CYFRA 21-1, yang dapat digunakan untuk mengembangkan alat diagnostik Kanker Nasofaring (KNF). Epitop imunogenik diprediksi dan dipilih berdasarkan antigenisitas, aksesibilitas permukaan, dan hidrofilisitas, kemudian dianalisis dan dievaluasi karakteristiknya. Kandidat epitop dibandingkan dengan biomarker protein saliva lainnya untuk menemukan apakah ada reaksi silang. CYFRA 21-1 terdiri dari 57 asam amino, di mana dua epitop imunogenik (C3 dan D2) dipilih. Plot Ramachandran dari kedua epitop tersebut menunjukkan bahwa 100% asam amino berada di area yang disukai. Epitop C3 dan D2 tidak memiliki reaksi silang dengan biomarker protein lainnya. Epitop imunogenik yang diprediksi memiliki potensi sebagai antigen untuk memproduksi antibodi sebagai bahan pengembangan imunodiagnostik berbasis saliva untuk mendiagnosis pasien KNF secara dini.*

Kata kunci: *antibodi, kanker, cytokeratin fragment 21-1, epitop, KNF, saliva.*

INTRODUCTION

Nasopharyngeal Carcinoma (NPC) is an aggressive cancer in the head and neck region, endemic in Southeast Asia, South China and North Africa (Lin *et al*. 2023). NPC has an incidence rate of 4-25 cases per 100,000 people in these endemic areas (50-100 times more than in other regions) (Wong *et al*. 2021). In 2018, there were an estimated 129,079

diagnosed cases of NPC internationally (85% occurred in Asia), with a mortality rate of 72,987. This makes NPC the 23rd-highest cancer incidence and the 21st cause of death in the world. In Southeast Asia alone, NPC ranks 9th for all cancer incidences and 8th for the number of deaths from cancer (Chang *et al*. 2021). In Indonesia, in 2012, the prevalence of NPC was 6.2 per 100,000 population, with a total of 13,000 new cases (Kadriyan *et al*. 2022).

NPC is a malignant tumour that originated in the nasopharyngeal epithelial tissue, undergoing a complex multifactor regulatory process, appearing on the posterior and side walls of the nasopharyngeal area (Xu & Lou 2021). The occurrence of NPC is believed to be the result of an interaction between Epstein-Barr virus (EBV) infection, smoking and alcohol consumption, environmental factors, as well as genetics (Tang *et al*. 2021). Although the standard treatments are surgery, radiation and chemotherapy, the overall prognosis of NCP remains poor (Liu *et al*. 2021). This is because 80% of NPC cases are newly diagnosed at an advanced stage (stages III & IV). The delay in diagnosis is due to the asymptomatic NPC at an early stage, high metastatic rate, and difficult access for examination of local primary tumours in the structures of the nasal cavity (Siak *et al*. 2021).

The success of early detection of cancer plays a very important role in getting a better prognosis (Adeoye *et al*. 2022). However, at this time, there is no gold standard as to which test is best for the early detection of NPC (Liu *et al*. 2021). The most popular diagnostic test today is a tissue biopsy, but this procedure is invasive, complicated, painful, time-consuming, and can harm the patient (Patel *et al*. 2022). For this reason, an accurate non-invasive method is needed for the early detection and monitoring of NPC using specific biomarkers (Zambonin & Aresta 2022). An example of a protein in saliva that has a high potential to be a biomarker in epithelial cell cancer is Cytokeratin fragment 21-1 (CYFRA 21-1) (Tofighi *et al*. 2021). As a soluble fragment of Cytokeratin 19 (KRT19), CYFRA 21-1 works as a proteolytic (Xu et al., 2022). This biomarker is released in large quantities in saliva, dissolves and is released during the process of cell apoptosis. CYFRA 21-1 levels in normal individuals is around 3.06 ± 0.25 ng/mL, while in cancer individuals it increases to 17.46 ± 1.46 ng/mL (Rajkumar *et al*. 2015; Jafari & Hasanzadeh 2020).

CYFRA 21-1 is a marker recognized by two monoclonal antibodies against KRT19 fragments in the serum. The two antibodies' epitopes were KS 19.1 and BM 19.21, found to be located within helix 2B of the rod domain of KRT19. For the catcher antibody (KS 19.1), the epitope is located within the amino acid sequences 311-335, and for the detection antibody (BM 19.21), the epitope is located within the amino acid sequences 346-367 (Jose *et al*. 2013). NPC cases can be detected more quickly if there are practical diagnostic tools. One example of a practical diagnostic tool is the Immunoassay test, which needs antibodies for its development (Michel *et al*. 2020). To make CYFRA 21-1 antibody, the antigen is needed to be used as the vaccine. The problem is that CYFRA 21-1 antigen is hard to obtain naturally. So, it is necessary to use another method, namely with an antigens that uses peptide fragments to get the desired antibody (Toepak *et al*. 2022).

With the development of biotechnology, the usage of epitope vaccines has increased due to in-silico technology, which uses bioinformatic tools and databases to produce safe and encouraging results (Toepak *et al*. 2022). This approach has several advantages, including the ability to reduce the amount of time and money spent, build persistent immunity to the desired response, and eliminate the unwanted immune response by building specialized structures (Parvizpour *et al*. 2020). The appropriate technique to assess changes in biomarkers in saliva as an initial diagnostic that can produce a clinical impact on NPC disease has not been found (Guruduth *et al*. 2021). In this study, we aim to identify CYFRA 21-1 immunogenic epitope that can be used to produce antibodies polyclonal against CYFRA 21-1, which later on will be used in the development of NPC non-invasive diagnostics.

MATERIALS AND METHODS

Target protein sequence identification and retrieval

CYFRA 21-1 is part of KRT19, hence the first step of our study is retrieving the protein sequence of KRT19. The Universal Protein Resource (Uni-Prot) database (https://www.uniprot.org/) provided the FASTA formatted amino acid sequence for the protein KRT19 (Accession no. P08727) (Rezaei *et al*. 2023). The protein sequence of CYFRA 21-1 was chosen based on the specific epitope region starting from KS 19.1 to BM 19.21 (311-367) (Jose *et al*. 2013).

Epitope prediction and selection

The IEDB server (http://tools.iedb.org/bcell/) was used to predict immunogenic CYFRA 21-1 epitopes. Antigenicity, surface accessibility, and hydrophilicity were factors taken into consideration when choosing the epitope. The antigenicity, surface accessibility, and hydrophilicity scores were obtained using the Kolaskar and Tongaonkar antigenicity prediction, Emini surface accessibility prediction, and Parker hydrophilicity prediction methods in the IEDB server. The immunogenicity prediction algorithms' threshold output values were used to determine if a putative immunogenic epitope should be included. Bepipred linier epitope prediction (https://services.healthtech.dtu.dk/services/BepiPred-2.0/) were also used with the threshold of 0.4 (Guevarra *et al*. 2020; Rezaei *et al*. 2023).

Epitope physicochemical characterization

The Physicochemical characterization of the epitope construct was assessed using the ExPASyProtParam database (http://web.expasy.org/protparam/). ExPASyProtParam is used to calculate numerous physical and chemical parameters like the number of amino acids and their composition, theoretical isoelectric point (PI), molecular weight, instability index, grand average hydropathicity (GRAVY), extinction coefficient, estimated half-life, atomic composition, and the aliphatic index (Akash *et al*. 2023; Rezaei *et al*. 2023).

Secondary and tertiary structure evaluation, 3D structure modeling, refinement and validation

Using the GOR4 server (https://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), the epitopes' secondary structure was evaluated to predict the percentage of strand, helix and coil. The epitopes' tertiary structure was then modelled using the I-TASSER server (https://zhanggroup.org/I-TASSER/). YASARA Energy Minimization server (https://www.yasara.org/minimizationserver.htm) was utilized to refine the best model structure of the epitopes construct. Then, the model structure was analyzed by using Ramachandran plots (Swiss-Model: https://swissmodel.expasy.org/) (Rezaei *et al*. 2023).

Homology and protein structural comparison

We performed sequence alignment of the selected epitope (C3 and D2) with multiple protein biomarkers in saliva to prevent positive cross-reaction (Table 1). Pairwise and multiple sequence alignment of the protein sequence was performed using ClustalW. Phylogenetics was analyzed by the ETE3 and the BIONJ option. Both ClustaW and ETE3 were embedded in the Genome.jp website tools (https://www.genome.jp/tools-bin/clustalw). The I-TASSER server was also used to find proteins that are structurally similar to the chosen epitope by evaluating the TM-score and identity percentage. TM-score is a metric for evaluating the topological similarity of protein structures. The TM-score ranges from 0 to 1, with 1 denoting a perfect match between two structures. According to stringent PDB statistics, structures with scores below 0.17 correspond to randomly selected unrelated proteins, whilst those with scores above 0.5 are assumed to have the same fold in SCOP/CATH. Uniprot and the bgee.org website was used to analyze the origin of the organism, classification and function of each protein (Sharma & Dubey 2020; Dong *et al*. 2023; Rezaei *et al*. 2023).

Proteolytic cleavage analysis

Regarding the stability of peptide antigens, it is necessary to carry out proteolytic cleavage analysis. PeptideCutter server (https://web.expasy.org/peptide_cutter/) were used to analyze the frequencies of the amino acids in proteolytic cleavage sites. Out of a total of 38 proteases, PeptideCutter performs a digestion and offers detailed findings, including the locations of the cleavage sites, peptide sequences, lengths, and masses (Maillet 2020; Goh & Hahn 2021).

RESULT AND DISCUSSION

The study of biomarkers for NPC has seen significant advancements, one of which is cytokeratin (CK). CYFRA 21-1 is a fragment of KRT19, which is one of the CK proteins that make up the cell cytoskeleton. Although CYFRA 21-1 does not directly influence the development of NPC, it's important to note that NPC is a type of squamous cell carcinoma that forms in the epithelial layer of the nasopharynx, where all epithelial cells express CK protein. CK protein is a helpful indicator for identifying epithelial differentiation. When cancer cells lysis or necrosis, some CK components are released into the body. One of these proteins is KRT19, which is expressed and can be immunohistochemically detected in the cytoplasm of epithelial cancer cells. Therefore, KRT19 can be used as a tumor marker for many malignancies, including NPC. The soluble part of KRT19 that is released into body fluids is known as CYFRA 21-1 (Sulaiman 2022; Adusumilli *et al*. 2023). Based on numerous studies, CYFRA 21-1 is a promising biomarker for the detection, staging, and monitoring of various cancers using different techniques and samples (Lei *et al*. 2019; Liu *et al*. 2019; Rudhart *et al*. 2020; Tofighi *et al*. 2021; Zhao *et al*. 2021).

To detect cancer using CYFRA 21-1 as a biomarker, we need to create antibodies against CYFRA 21-1. To do this, we require a specific antigen. Antigen that is specific for cancer detection have been challenging to produce, and there is no one-size-fits-all method or tool for rational specific antigen production. The process of developing specific antigen requires several steps. By mapping thousands of biological components *in silico*, computational approaches can dramatically cut down on the time and expense of vaccine development. Specific antigen (vaccine) development and analysis using this method can start with proteome retrieval, epitope prediction, epitope selection, molecular interactions, and immune response simulation (Patronov & Doytchinova 2013; Parvizpour *et al*. 2020; Prawiningrum *et al*. 2022; Toepak *et al*. 2022).

Target protein sequence identification and retrieval

In this study, the KRT19 protein sequence was retrieved in FASTA format from the Uniprot

database which consists of 400 amino acids. From the KRT19 protein sequence, CYFRA 21-1 protein sequence was identified starting from 311-367 which consists of 57 amino acids. The sequence is shown in Figure 1. Determination of this sequence was based on the results of previous studies that the sequence can be recognized by two monoclonal antibodies, namely KS 19.1 (known as catcher epitope) and BM 19.21 (known as detector epitope) (Jose *et al*. 2013). Based on this, it can be assumed that the specific antigen developed from this sequence can produce catcher and detector antibodies that are equivalent to the monoclonal antibodies developed by Jose *et al*. (2013). To prove this, it was necessary to carry out immunogenicity testing

Epitope prediction and selection

Initially, in predicting and selecting the catcher and detector epitope we used two approaches. The first approach was to use four servers to predict the immunogenic epitopes. The four servers are IEDB, ABCpred, BCEPS, and BepiPred (Jespersen *et al*. 2017; Ras-Carmona *et al*. 2021; Dar *et al*. 2022; Rezaei *et al*. 2023). Of the four servers, 7 epitopes were obtained which had a length of 16 amino acids. However, after looking at the antigenicity, surface accessibility, and hydrophilicity parameters, the seven epitopes had fewer immunogenic values (data not shown). Our next approach was to predict epitope one by one using the IEDB server (Kolaskar and Tongaonkar for antigenicity prediction, Emini for surface accessibility prediction, and Parker for hydrophilicity prediction) with the default threshold value (Guevarra *et al*. 2020). Assessment of these three parameters was very important in designing specific antigens because it increases the likelihood of producing antibodies specific to CYFRA 21-1.

In Table 2, six immunogenic sequences were identified. Further testing of the 6 candidates and based on the highest scores, epitopes C3 and D2 were determined (Table 3). When comparing the D2 epitope candidate with the original BM 19.21 epitope, the D2 epitope has higher antigenicity and meets all three immunogenic parameters. Meanwhile, the C3 epitope when compared to the original KS 19.1 epitope has a higher antigenicity/immunogenicity score. This high immunogenicity score plays an important role in producing antibodies, especially in diagnostic development. Bepipred analysis (IEDB), shows these two epitopes are unique. From the position of the immunogenic peptides (Figure 2), we can assume that both epitopes would be a good candidate for sandwich (one as a catcher and the other as the detection).

Epitope physicochemical characterization

The next step was to look at the physicochemical characterization of the two selected epitope candidates. The results using the ExPASyProtParam server showed that the epitopes C3 and D2 have a length of 27 amino acids, and have a relatively small

Table 1. Protein biomarkers in saliva (Dong *et al*. 2023).

No.	Protein biomarker	Related disease	Range
1.	Interleukin-1β	Periodontitis	$161.51^{(a)}$ - $1312.75^{(b)}$ (pg mL^{-1})
2.	Interleukin-8	Oral squamous cell carcinoma	$210.20^{(a)}$ - $1718.61^{(b)}$ (pg mL^{-1})
3.	Vascular endothelial growth factor (VEGF)	Oropharyngeal cancer	$280^{(a)}$ – $4321^{(b)}$ (pg mL^{-1})
4.	Triosephosphate isomerase (TPI1)	Gastric Cancer	$>800^{(a)}$ – $<400^{(b)}$ (U mL^{-1})
5.	α-amylase	Stress	$<50\text{-}100^{(a)}$ – $>150^{(b)}$ (U mL^{-1})

$^{(a)}$Value in normal condition; $^{(b)}$ Value in disease condition.

MTSYSYRQSSATSSFGGLGGGSVRFGPGVAFRAPSIHGGSGGRGVSVSSARFVSSSSSGA
YGGGYGGVLTASDGLLAGNEKLTMQNLNDRLASYLDKVRALEAANGELEVKIRDWYQKQG
PGPSRDYSHYYTTIQDLRDKILGATIENSRIVLQIDNARLAADDFRTKFETEQALRMSVE
ADINGLRRVLDELTLARTDLEMQIEGLKEELAYLKKNHEEEISTLRGQVGGQVSVEVDSA
PGTDLAKILSDMRSQYEVMAEQNRKDAEAWFTSRTEELNREVAGHTEQLQMSRSEVTDLR
RTLQGLEIELQSQLSMKAALEDTLAETEARFGAQLAHIQALISGIEAQLGDVRADSERQN
QEYQRLMDIKSRLEQEIATYRSLLEGQEDHYNNLSASKVL

Figure 1. CYFRA 21-1 sequence identification (underlined)

Table 2. Predicted immunogenic epitopes

Epitopes		Start	Peptide sequence	Length
Catcher	C1	311	QSQLSMKAALEDTLAETEARFGAQLA	26
	C2	312	SQLSMKAALEDTLAETEARFGAQLAH	26
	C3	311	QSQLSMKAALEDTLAETEARFGAQLAH	27
Detector	D1	338	IQALISGIEAQLGDVRADSERQNQEY	26
	D2	338	IQALISGIEAQLGDVRADSERQNQEYQ	27
	D3	340	ALISGIEAQLGDVRADSERQNQEYQRL	27

Table 3. Epitope candidates compared to the original CYFRA 21-1 epitope**.**

Epitopes	Antigenicity (> 1.011)	Accessibility (> 1.000)	Hydrophilicity (> 2.182)
C1	1.011	1.309	2.031
C2	1.014	1.029	1.881
C3	1.014	1.573	2.033
D1	1.013	1.889	2.519
D2	1.013	2.888	2.648
D3	1.011	3.843	2.537
KS 19.1	1.009	1.295	2.028
BM 19.21	0.983	8.249	3.418

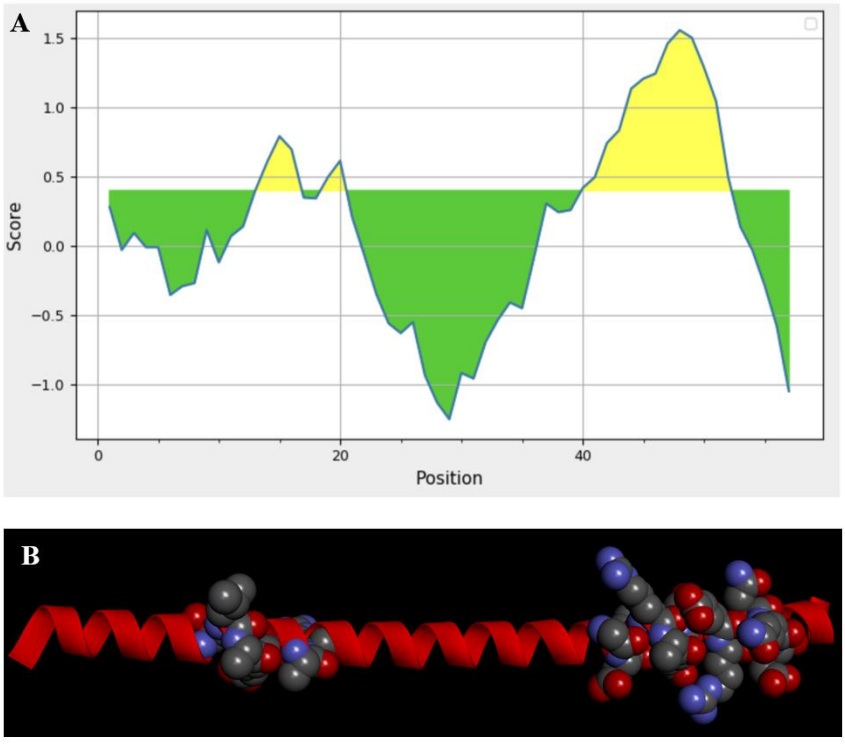

Figure 2. [A] Immunogenic parts of the CYFRA 21-1 epitope based on Bepipred linier epitope prediction (colored yellow, position 13-16 (TLAE), 19-20 (AR), and 40-52 (GDVRADSERQNQE). [B] 3D visualization of the alpha helix structure of CYFRA 21-1, and the immunogenic parts of CYFRA 21-1 epitopes (Red circles represent oxygen atoms, blue circles represent nitrogen atoms, and dark gray circles represent carbon atoms).

molecular weight of 2917.24 Da and 3032.27 Da respectively. If we assess the instability index, it can be seen that the C3 epitope is more stable than the original KS 19.1 epitope. However, the D2 epitope here was considered an unstable protein compared to the BM 19.21 epitope. In this case, if an epitope is unstable, the solution that can be done is to combine the epitope with another protein such as protein carrier to increase the stability of the protein (Tîrziu *et al*. 2023).

Secondary & tertiary structure evaluation and 3D structure refinement & validation

The second structure was evaluated via GOR4, and the tertiary structure was built using the I-Tasser server. Epitope C3 secondary evaluation showed a 62.96% helix, 7.41% strand, and 29.63% coil. Whereas for epitope D2 had 55.56% helix, 7.41% strand, and 37.04% coil. From tertiary structure evaluation, the best model of the epitope construct was selected based on the highest C-score (range from -5 to 2). The C-score indicates the confidence score of the protein model evaluated by the I-TASSER server. The best antigen model C-score for C3 was 0.32, with an estimated TM-score 0.76±0.10 and RMSD 1.0±1.0A, whereas the C-score for D2 was 0.30, with an estimated TM-score 0.75±0.10 and RMSD 1.0±1.0A. Both epitopes C3 and D2 had a good confidence score which means the model has a good quality structure. The best 3D structure model was then refined using YASARA server, which is shown in Figure 3. After being refined using the YASARA server, both epitopes were validated using the Ramachandran Plot. Based on the findings, 100% of the amino acids were in the favoured area for epitopes C3 and D2 (Figure 3).

Homology and protein structural comparison

The specific antigen design aims to produce antibodies that can be used to develop a saliva-based diagnostic tool to detect CYFRA 21-1 in NPC patients. For this reason, it is necessary to carry out a comparative analysis between the C3 and D2 epitopes against several biomarker proteins that are commonly found in human saliva. From the sequence alignment (Supplementary Figure 1), the gap open penalty value used in the ClustalW was set to 10.0, the gap extension penalty was set to 0.1, and the weight matrix used was BLOSUM. The minimum pairwise alignment score for both epitope was 11, and the highest score was 14 for epitope C3 and 18 for epitope D2. The homology between epitope C3 and D2 with other protein saliva were analyze using ETE3, and were presented in a phylogenetic tree (Figure 4). From the homology and the overall results of the pairwise alignment score that is below 50%,

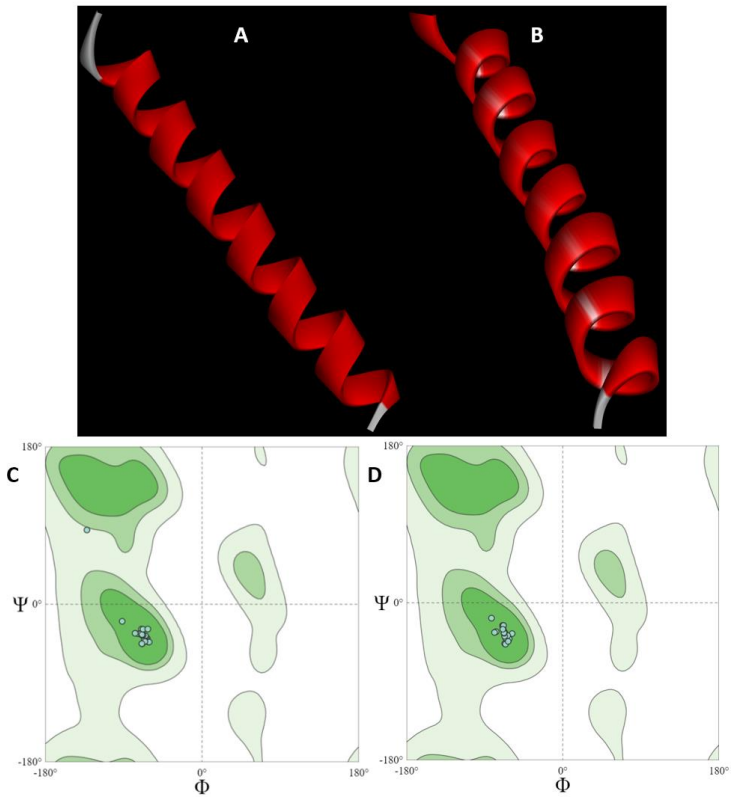

Figure 3. (A) Refined model of epitope C3; (B) Refined model of epitope D2; (C) C3 Ramachandran Plots; (D) D2 Ramachandran Plots.

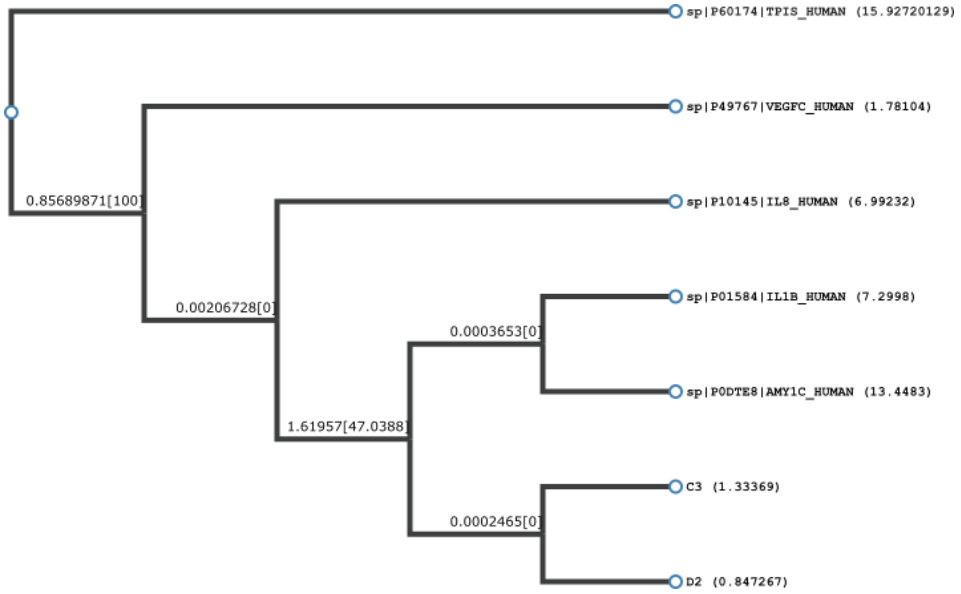

Figure 4. Homology between epitope C3 and D2 with other protein saliva.

Table 4. Sequence alignment with other protein saliva.

	Pairwise alignment score (%)	
	Epitope C3	**Epitope D2**
Interleukin-1β	14	11
Interleukin-8	11	14
Vascular endothelial growth factor (VEGF)	14	18
Triosephosphate isomerase (TPI1)	14	14
α-amylase	14	14

means there would be minimal positive cross-reaction between the epitopes with the other protein biomarker found in saliva (Table 4).

The results of analysis using I-TASSER, it shows that there are proteins that are structurally similar to the C3 and D2 epitopes. Based on Uniprot ID, in both C3 and D2, ten types of proteins were obtained that were structurally similar. The origin of the organism, classification and function of each protein are presented in Table 5 for epitope C3 and Table 6 for epitope D2.

Based on the results of homology analysis and comparison of protein structures using ClustalW, it can be seen that the average pairwise alignment score between epitopes C3 and D2 with other protein biomarkers is below 50%. The analysis results indicate that epitopes C3 and D2 are not closely related to the five biomarker proteins in saliva, so the likelihood of cross-reaction is low. Cross-reaction can occur when epitopes have identical sequences or conformations with the same structure as other proteins or antigens (Mitchell *et al*. 2018). Determining if a cross-reaction has occurred based on sequence similarity is not highly reliable, as proteins form a structure. Assessing cross-reactions based on structure can enhance accuracy and validity (Buraphaka *et al*. 2024). Therefore, homology between the three-dimensional structures of the two epitopes with other proteins was conducted using the I-TASSER server. Based on the structure, the five biomarker proteins in saliva did not appear in the results of the three-dimensional structure homology analysis from the I-TASSER server.

Further analysis using I-TASSER, we figured out that both epitopes C3 and D2 have some similar structure proteins. We retrieve ten proteins that have the closest structure for each epitope. From this analysis, the origin, TM-score, identity score, classification, function and whether it is found in saliva can be seen. As mentioned in Table 5, of the ten proteins that are structurally similar to the C3 epitope, four are found in the Homo Sapiens species. Meanwhile, from Table 6, it can be seen that there are four proteins structurally similar to the D2 epitope found in human (*Homo sapiens)* species. The eight proteins found in human (*Homo Sapiens)* that are similar to the C3 and D2 epitopes can be found in saliva (produced by minor salivary glands). However, if we look at the TM-score and identity score that is quite small, hence the cross-reaction between the

Table 5. Protein structurally close to epitope C3

Rank	Protein PDB Hit	Uniprot ID	TM-score*	IDEN* (%)	Organism	Classification	Function	Presence in saliva
1	7CGPC (Mitochondrial import inner membrane translocase subunit Tim29)	Q9BSF4	0.942	0.074	*Homo sapiens*	Translocase	A part of the TIM22 complex, which facilitates the entry and insertion of multi-pass transmembrane proteins into the inner membrane of the mitochondria. The membrane potential serves as the external driving factor for the twin-pore translocase that the TIM22 complex generates. Functions in the assembly of the TIMM22 protein into the TIM22 complex and is necessary for the stability of the TIM22 complex. may interact with TOMM40 to encourage cooperation between TIM22 and TOM complexes (Callegari *et al.*, 2016).	Yes
2	6RW8A (A component of insecticidal toxin complex (Tc))	D3VHH9	0.942	0.111	*Xenorhabdus nematophila (strain ATCC 19061 / DSM 3370 / CCUG 14189 / LMG 1036 / NCIMB 9965 / AN6)*	Toxin	N/A	No
3	3A68C (Ferritin-4, chloroplastic)	Q948P5	0.939	0.111	*Glycine max*	Oxidoreducta-se	Keeps iron in a soluble, safe, and accessible state. Iron is taken up in the ferrous state and deposited as ferric hydroxides after oxidation, which is necessary for iron homeostasis (By similarity) (Masuda *et al*. 2010).	No
4	7O41D (TrwG protein)	O50335	0.939	0.037	*Escherichia coli*	Membrane protein	N/A	No

5	7T3UA (Inositol 1,4,5-trisphosphate receptor type 3)	Q14573	0.938	0.148	*Homo sapiens*	Metal transport	It is a receptor for inositol 1,4,5-trisphosphate, a second messenger that, by analogy, mediates the release of intracellular calcium and is involved in maintaining the balance of calcium ions in cells (Rönkkö *et al*. 2020).	Yes
6	5MRCCC (uL3m)	P31334	0.938	0.037	*Saccharomyces cerevisiae (strain ATCC 204508 / S288c)*	Ribosome	A part of the mitochondrial ribosome (mitoribosome), a special translational apparatus in charge of producing proteins from the mitochondrial genome, including at least some of the crucial transmembrane components of the mitochondrial respiratory chain. Translation products are cotranslationally integrated into the inner membrane of the mitochondria by the mitoribosomes, which are connected to it (Pfeffer *et al*. 2015).	No
7	7JGDA1 (Erythrocyte membrane protein 1)	Q6UDW7	0.937	0.037	*Plasmodium falciparum*	Sugar binding protein	N/A	No
8	7EU7A (Glutamate receptor ionotropic, NMDA 1)	Q05586	0.932	0.074	*Homo sapiens*	Membrane protein	It is a part of NMDA receptor complexes, heterotetrameric ion channels that are ligand-gated and have significant calcium permeability as well as voltage-dependent sensitivity to magnesium. Glycine must bind to the zeta subunit of the channel, glutamate must bind to the epsilon subunit, and membrane depolarization is necessary to remove Mg2+-induced channel blockage (Chen *et al*. 2021).	Yes
9	7RPMA (Neuronal acetylcholine receptor subunit	P36544	0.930	0.148	*Homo sapiens*	Membrane protein	AChR undergoes a significant conformational shift after binding acetylcholine, which affects all subunits and causes the opening of an ion-conducting channel across the plasma membrane. Alpha-bungarotoxin shuts down the channel (Kalashnyk *et al*.	Yes

	alpha-7)						2023).	
10	6M6ZA (TMH4C4)		0.930	0.074	*Escherichia coli*	De novo protein	N/A	No

* TM-score: a metric for assessing the topological similarity of protein structures.
* IDEN: the percentage sequence identity in the structurally aligned region.

Table 6. Protein structurally close to epitope D2

Rank	Protein PDB Hit	Uniprot ID	TM-score*	IDEN* (%)	Organism	Classification	Function	Presence in saliva
1	6SO5C (Tail-anchored protein insertion receptor WRB)	O00258	0.919	0.074	*Homo sapiens*	Membrane protein	Required for the endoplasmic reticulum (ER) to receive tail-anchored (TA) proteins post-translationally (Vilardi *et al.* 2011).	Yes
2	6MZCM (Transcription initiation factor TFIID subunit 9)	Q16594	0.899	0.074	*Homo sapiens*	Transcription	A significant part of the beginning of RNA polymerase II (Pol II)-dependent transcription is played by the TFIID basal transcription factor complex (Chen *et al*. 2021)	Yes
3	2NPSC (Vesicle transport through interaction with t-SNAREs homolog 1A)	Q9JI51	0.890	0.000	*Rattus norvegicus*	Transport protein	By interacting with t-SNAREs on the target membrane, the V-SNARE mediates vesicle transport pathways. It is suggested that these interactions mediate certain elements of the specificity of vesicle trafficking and encourage the fusing of the lipid bilayers. engaged in the movement of vesicles from late endosomes to the trans-Golgi network. KCNIP1 and KCND2 use an unconventional RAB1-dependent traffic pathway to the cell surface in conjunction with VAMP7 (By similarity). Concerned about cytokine output that	No

							has risen due to cellular aging (Zwilling *et al.* 2007).	
4	6Z6FD (HDA1 complex subunit 2)	Q06629	0.888	0.111	*Saccharomyces cerevisiae (strain ATCC 204508 / S288c))*	Gene regulation	Necessary for the HDA1 histone deacetylase complex to function. The lysine residues on the N-terminal portion of the core histones (H2A, H2B, H3 and H4) are deacetylated by the HDA1 histone deacetylase complex. In transcriptional control, cell cycle progression, and developmental processes, histone deacetylation provides a tag for epigenetic repression (Wu *et al.* 2001).	No
5	5B3DA (Flagella synthesis protein FlgN)	P0A1J7	0.884	0.111	*Salmonella typhimurium (strain LT2 / SGSC1412 / ATCC 700720)*	Protein transport	N/A	No
6	7O3VA (TrwJ protein)	O50331	0.884	0.185	*Escherichia coli*	Membrane protein	N/A	No
7	7T3UA (Inositol 1,4,5-trisphosphate receptor type 3)	Q14573	0.883	0.111	*Homo sapiens*	Metal transport	Inositol 1,4,5-trisphosphate receptor, a second messenger that mediates the release of intracellular calcium (By similarity). Homeostasis of calcium ions within cells is affected (Rönkkö *et al.* 2020).	Yes
8	7OPCS (RNA polymerase-associated protein LEO1)	Q8WVC0	0.879	0.148	*Homo sapiens*	Transcription	It is a part of the PAF1 complex (PAF1C), which is involved in the control of embryonic stem cell pluripotency and has a variety of roles during transcription by RNA polymerase II (Kim *et al.* 2010).	Yes
9	6FKFB (ATP synthase	P00825	0.879	0.037	*Spinacia oleracea*	Membrane protein	when a proton gradient is present across the membrane, it converts ADP to ATP. In general, the beta subunits host the catalytic sites	No

	subunit beta, chloroplastic)						(Yang *et al*. 2020).	
10	6VQ6I (V-type proton ATPase subunit E 1)	Q6PCU2	0.879	0.074	*Rattus norvegicus*	Proton transport	Subunit of the vacuolar(H+)-ATPase (V-ATPase), a multisubunit enzyme made up of a membrane-integral complex (V0) that transports protons and a peripheral complex (V1) that hydrolyzes ATP. In some cell types, V-ATPase is directed to the plasma membrane, where it is in charge of acidifying the extracellular environment in addition to acidifying and regulating the pH of internal compartments (Abbas *et al*. 2020).	No

* TM-score: a metric for assessing the topological similarity of protein structures.
* IDEN: the percentage sequence identity in the structurally aligned region.

epitopes with other protein saliva possibly low. In terms of protein structure, the C3 and D2 epitopes are unique. Since the aim of this research is to produce antibody polyclonal against CYFRA 21-1, the possibility of cross-reaction can be suppressed or abolished using some blocking agent (Lv *et al*. 2020).

Proteolytic cleavage analysis

From the proteolytic cleavage analysis using PeptideCutter, there were several enzymes that cleave with both epitope (Shown in Table 7). Data shows the enzymes that causes the most cleavages is Proteinase K, Thermolysin, and Pepsin (pH>2) with 30, 20 and 15 number of cleavages.

The proteolytic cleavage analysis shows that there were several enzymes that may cause cleavage to the epitopes. In general, the cleavage is relatively normal according to the function of each cleavage enzyme. As can be seen, most prominently, the cleavage is carried out by Proteinase K. Proteinase K is known as the most destructive proteinase. Based on this description of the cleavage site, handling the C3 and D2 epitopes when used as antigen must pay attention to several aspects so that the protease enzymes do not have time to cut the epitope sequence. This can be done by using adjuvants as stabilizers and also paying attention to the buffer used regarding pH and storage temperature (Eladawy *et al*. 2020).

CONCLUSION

In this study, we have successfully designed two specific antigen (epitopes) that have the potential to generate antibodies against CYFRA 21-1 *in silico*. The epitopes are C3 and D2 with the sequences QSQLSMKAALEDTLAETEARFGAQLAH and IQALISGIEAQLGDVRADSERQNQEYQ respectively. Both do not have cross-reactions with other salivary biomarker proteins and have a unique protein structure that has the potential to be used as antigen to produce antibodies for the development of saliva-based immunodiagnostics to diagnose NPC patients. To the best of our knowledge, this is the first reported *in silico* antigen design for salivary biomarkers in detecting NPC. To approve this antigen design, an *in vivo* study needs to be carried out.

Table 7. Enzymes that cleave CYFRA 21-1 sequence

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	4	20 43 48 55
Asp-N endopeptidase	3	11 40 44
Asp-N endopeptidase + N-terminal Glu	9	10 11 15 17 35 40 44 46 51
CNBr	2	6 57
Chymotrypsin-high specificity	2	21 53
Chymotrypsin-low specificity	12	4 6 10 14 21 25 27 31 39 53 56 57
Clostripain	4	20 43 48 55
Formic acid	3	12 41 45
Glutamyl endopeptidase	6	11 16 18 36 47 52
LysC	1	7
LysN	1	6
Pepsin (pH 1,3)	13	3 4 10 13 14 21 24 25 30 31 38 39 56
Pepsin (pH >2)	15	3 4 10 13 14 21 24 25 30 31 38 39 52 53 56
Proteinase K	30	4 8 9 10 11 13 14 15 16 17 18 19 21 23 25 26 28 30 31 32 35 36 37 39 42 44 47 52 53 56
Staphylococcal peptidase I	6	11 16 18 36 47 52
Thermolysin	20	3 5 7 8 9 13 14 20 22 24 25 27 29 30 31 34 38 43 55 56
Trypsin	5	7 20 43 48 55

ACKNOWLEDGMENT

The authors would like to thank The Ministry of Research, Technology and Higher Education Indonesia and Research Center for Molecular Biotechnology and Bioinformatics, Padjadjaran University. Funding: This work was supported by The Ministry of Research, Technology and Higher Education Indonesia (with the contract number: 3089/UN6.3.1/PT.00/2023). It was done related to the master project of Brian Umbu Rezi Depamede, MD. under the supervision of Muhammad Yusuf, Ph.D and Prof. dr. Bachti Alisjahbana, SpPD-KPTI, Ph.D.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abbas, Y.M., Wu, D., Bueler, S.A., Robinson, C.V. & Rubinstein, J.L. (2020). Structure of V-ATPase from the mammalian brain. *Science*. **367(6483)**: 1240-1246.

Adeoye, J., Wan, C.C.J., Zheng, L.W., Thomson, P., Choi, S.W. & Su, Y.X. (2022). Machine learning-based genome-wide salivary DNA methylation analysis for identification of noninvasive biomarkers in oral cancer diagnosis. *Cancers*. **14(19)**: 4935.

Adusumilli, P., Babburi, S., Venigalla, A., Benarji, K.A., Sai, S.K. & Soujanya, P. (2023). Estimation of salivary and serum CYFRA 21-1 levels in patients with oral squamous cell carcinoma. *Journal of Oral and Maxillofacial Pathology*. **27(1)**: 98-102.

Akash, S.R., Hossain, M.I. & Ali, M.S. (2023). Developing a multiepitope vaccine against dengue virus in Bangladesh using immunoinformatics approach. *Journal of Advanced Biotechnology and Experimental Therapeutics*. **6(1): 44–57**.

Buraphaka, H., Dobutr, T., Wiese, M.D., Lopata, A. L. & Daduang, S. (2024). Structure-based epitope prediction and assessment of cross-reactivity of Myrmecia pilosula venom-specific IgE and recombinant Sol g proteins (Solenopsis geminata). *Scientific Reports*. **14(1)**: 11145.

Callegari, S., Richter, F., Chojnacka, K., Jans, D. C., Lorenzi, I., Pacheu-Grau, D., Jakobs, S., Lenz, C., Urlaub, H., Dudek, J. Chacinska, A. & Rehling, P. (2016). TIM29 is a subunit of the human carrier translocase required for protein transport. *Febs Letters*. **590(23)**: 4147-4158.

Chang, E.T., Ye, W., Zeng, Y.X. & Adami, H. O. (2021). The evolving epidemiology of nasopharyngeal carcinoma. *Cancer Epidemiology, Biomarkers & Prevention*. **30(6)**: 1035-1047.

Chen, X., Qi, Y., Wu, Z., Wang, X., Li, J., Zhao, D., Hou, H, Li, Y., Yu, Z., Liu, W., Ren, Y., Li, Z., Yang, H. & Xu, Y. (2021). Structural insights into preinitiation complex assembly on core promoters. *Science*. **372(6541)**: eaba8490.

Dar, M.A., Kumar, P., Kumar, P., Shrivastava, A., Dar, M.A., Chauhan, R., Trivedi, V., Singh, A., Khan, E., Velayutham, R. & Dhingra, S. (2022). Designing of peptide based multi-epitope vaccine construct against gallbladder cancer using immunoinformatics and computational approaches. *Vaccines*. **10(11)**: 1850.

Dong, T., Matos Pires, N. M., Yang, Z. & Jiang, Z. (2023). Advances in electrochemical biosensors based on nanomaterials for protein biomarker detection in saliva. *Advanced Science*. **10(6)**: 2205429.

Eladawy, M., El-Mowafy, M., El-Sokkary, M.M.A. & Barwa, R. (2020). Effects of lysozyme, proteinase K, and cephalosporins on biofilm formation by clinical isolates of Pseudomonas aeruginosa. *Interdisciplinary Perspectives on Infectious Diseases*. **2020(1)**: 6156720.

Goh, C.J. & Hahn, Y. (2021). Analysis of proteolytic processing sites in potyvirus polyproteins revealed differential amino acid preferences of NIa-Pro protease in each of seven cleavage sites. *PLoS One*. **16(1)**: e0245853.

Guevarra Jr, L.A., Boado, K.J.O., Ceñidoza, F.B.B., Imbao, M.R.L.M., Sia, M.J.G. & Dalmacio, L. M. M. (2020). A synthetic peptide analog of in silico-predicted immunogenic epitope unique to dengue virus serotype 2 NS1 antigen specifically binds immunoglobulin G antibodies raised in rabbits. *Microbiology and Immunology*. **64(2)**: 153-161.

Guruduth, B., Oyetunji, O., Ryan, T., Sathyapriya, R., Yen, K.L., Kai, T., Francine C., Pedro, J.T., Stephanie, G., Matthew P., Liz, K., Ally, P., Hal, T., Nevenka, D., Salomon, Momchilo, V. & Chamindie, P. (2021). The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer. *NPJ Genomic Medicine*. **6(1)**: 1-11.

Jafari, M. & Hasanzadeh, M. (2020). Non-invasive bioassay of Cytokeratin Fragment 21.1 (Cyfra 21.1) protein in human saliva samples using immunoreaction method: An efficient platform for early-stage diagnosis of oral cancer based on biomedicine. *Biomedicine & Pharmacotherapy*. **131**: 110671.

Jespersen, M.C., Peters, B., Nielsen, M. & Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*. **45(W1)**: W24-W29.

Jose, J., Sunil, P,M., Nirmal, R.M. & Varghese, S.S. (2013) CYFRA 21-1: An overview. *Oral &Maxillofacial Pathology Journal*. 4(2): 368-371.

Kadriyan, H., Yudhanto, D., Cahyawati, T.D. & Wedayani, N. (2022). Pengenalan gejala

nasopharyngeal cancer pada populasi beresiko di Lombok Utara 2021 pada saat pandemi Covid19. *Jurnal Pengabdian Magister Pendidikan IPA*. **5(3)**: 231-234.

Kalashnyk, O., Lykhmus, O., Koval, L., Uspenska, K., Obolenskaya, M., Chernyshov, V. Komisarenko, S. & Skok, M. (2023). α7 Nicotinic acetylcholine receptors regulate translocation of HIF-1α to the cell nucleus and mitochondria upon hypoxia. *Biochemical and Biophysical Research Communications*. **657**: 35-42.

Kim, J., Guermah, M. & Roeder, R.G. (2010). The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*. **140(4)**: 491-503.

Lei, Q., Zhao, L., Ye, S., Sun, Y., Xie, F., Zhang, H., Zhou, F. & Wu, S. (2019). Rapid and quantitative detection of urinary Cyfra21-1 using fluorescent nanosphere-based immunochromatographic test strip for diagnosis and prognostic monitoring of bladder cancer. *Artificial Cells, Nanomedicine, and Biotechnology*. **47(1)**: 4266-4272.

Lin, J., Lin, D., Qiu, S., Huang, Z., Liu, F., Huang, W., Xu, Y., Zhang, X. & Feng, S. (2023). Shifted-excitation Raman difference spectroscopy for improving in vivo detection of nasopharyngeal carcinoma. *Talanta*. **257**: 124330.

Liu, L., Xie, W., Xue, P., Wei, Z., Liang, X. & Chen, N. (2019). Diagnostic accuracy and prognostic applications of CYFRA 21-1 in head and neck cancer: A systematic review and meta-analysis. *PLoS One*. **14(5)**: e0216561.

Liu, W., Chen, G., Gong, X., Wang, Y., Zheng, Y., Liao, X., Liao, W., Song, L., Xu, J. & Zhang, X. (2021). The diagnostic value of EBV-DNA and EBV-related antibodies detection for nasopharyngeal carcinoma: a meta-analysis. *Cancer Cell International*. **21**: 1-13.

Lv, H., Wu, N.C., Tsang, O.T.Y., Yuan, M., Perera, R.A., Leung, W.S., So, R.T.Y., Chan, J.M.C.C., Yip, G.K., Chik, T.S.H.C., Wang, Y., Choi, C.Y.C.C., Lin, Y., Ng, W.W., Zhao, J., Poon, L.L.M., Peiris, J.S.M., Wilson, I.A. & Mok, C. K. (2020). Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections. *Cell Report.* **31(9)**: 1-6.

Maillet, N. (2020). Rapid Peptides Generator: fast and efficient in silico protein digestion. *NAR Genomics and Bioinformatics*. **2(1)**: lqz004.

Masuda, T., Goto, F., Yoshihara, T. & Mikami, B. (2010). Crystal structure of plant ferritin reveals a novel metal binding site that functions as a transit site for metal transfer in ferritin. *Journal of Biological Chemistry*. **285(6)**: 4049-4059.

Michel, M., Bouam, A., Edouard, S., Fenollar, F., Di Pinto, F., Mège, J. L., Drancourt, M. & Vitte, J. (2020). Evaluating ELISA, immunofluorescence, and lateral flow assay for SARS-CoV-2 serologic assays. *Frontiers in Microbiology*. **11**: 597529.

Mitchell, A.M., Srikumar, T. & Silhavy, T. J. (2018). Cyclic enterobacterial common antigen maintains the outer membrane permeability barrier of Escherichia coli in a manner controlled by YhdP. *MBio*. **9(4)**: 10-1128.

Parvizpour, S., Pourseif, M.M., Razmara, J., Rafi, M. A. & Omidi, Y. (2020). Epitope-based vaccine design: a comprehensive overview of bioinformatics approaches. *Drug Discovery Today*. **25(6)**: 1034-1042.

Patel, A., Patel, S., Patel, P. & Tanavde, V. (2022). Saliva based liquid biopsies in head and neck cancer: how far are we from the clinic?. *Frontiers in Oncology*. **12**: 828434.

Patronov, A. & Doytchinova, I. (2013). T-cell epitope vaccine design by immunoinformatics. *Open Biology*. **3(1)**: 120139.

Pfeffer, S., Woellhaf, M.W., Herrmann, J.M. & Förster, F. (2015). Organization of the mitochondrial translation machinery studied in situ by cryoelectron tomography. *Nature Communications*. **6(1)**: 6019.

Prawiningrum, A.F., Paramita, R.I. & Panigoro, S.S. (2022). Immunoinformatics approach for epitope-based vaccine design: Key steps for breast cancer vaccine. *Diagnostics*. **12(12)**: 2981.

Rajkumar, K., Ramya, R., Nandhini, G., Rajashree, P., Kumar, A.R, & Anandan, S.N (2015). Salivary and serum level of CYFRA 21-1 in oral precancer and oral squamous cell carcinoma. *Oral Diseases*. **21(1)**: 90-96.

Ras-Carmona, A., Pelaez-Prestel, H. F., Lafuente, E. M. & Reche, P. A. (2021). BCEPS: A web server to predict linear B cell epitopes with enhanced immunogenicity and cross-reactivity. *Cells*. **10(10)**: 2744.

Rezaei, M., Esmaeili, F., Karam, M.R.A., Ehsani, P., Farsangi, Z.A. & Bouzari, S. (2023). In silico design and in vivo evaluation of two multi-epitope vaccines containing build-in adjuvant with chitosan nanoparticles against uropathogenic *Escherichia coli*. *International Immunopharmacology*. **117**: 109999..

Rönkkö, J., Molchanova, S., Revah-Politi, A., Pereira, E., Auranen, M., Toppila, J. Kvist, J., Ludwig, A., Neumann, J. Bultynck, G., Humblet-Baron, S., Liston, A., Paetau, A., Rivera, C., Harms, M.B. Tyynismaa, H. & Ylikallio, E. (2020). Dominant mutations in ITPR3 cause Charcot-Marie-Tooth disease. *Neuromuscular Disorders*. **30**: S167-S168.

Rudhart, S.A., Gehrt, F., Birk, R., Schultz, J.D., Stankovic, P., Georgiew, R., Wilhelm, T., Stuck, B.A. & Hoch, S. (2020). Clinical relevance of CYFRA 21-1 as a tumour marker

in patients with oropharyngeal squamous cell carcinoma. *European Archives of Oto-Rhino-Laryngology*. **277**: 2561-2571.

Sharma, R. & Dubey, S. K. (2020). Computational management of alignment of multiple protein sequences using ClustalW. *In First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019*. Singapore. January 2020. pp. 177-186.

Siak, P.Y., Khoo, A.S.B., Leong, C.O., Hoh, B.P. & Cheah, S. C. (2021). Current status and future perspectives about molecular biomarkers of nasopharyngeal carcinoma. *Cancers*. **13(14)**: 3490.

Sulaiman, E. (2022) Nasopharyngeal Carcinoma patient characteristics at sayang general hospital Cianjur West Java. *Proceedings of the 2nd Global Health and Innovation in conjunction with 6th ORL Head and Neck Oncology Conference (ORLHN 2021)*. Cianjur. 21 February 2022. pp 312-314.

Tang, L.L., Chen, Y.P., Chen, C.B., Chen, M.Y., Chen, N.Y., Chen, X.Z., Du, X.J., Fang, W.F., Feng, M., Gao, J., Han, F., He, X., Hu, C.S., Hu, D.S., Hu, G.Y., Jiang, H., Jiang, W., Jin, F., Lang, J.Y., Li, J.G., Lin, S.J., Liu, X., Liu, Q.F., Ma, L., Mai, H.Q., Qin, J.Y., Liu, Q.F., Ma, L., Mai, H.Q., Qin, J.Y., Shen, L.F., Sun, Y., Wang, P.G., Wang, R.S., Wang, R.Z., Wang, X.S., Wang, Y., Wu, H., Xia, Y.F., Xiao, S.W., & Yang, K.Y., Li, J.L., Zhu, X.D. & Ma, J. (2021). The Chinese Society of Clinical Oncology (CSCO) clinical guidelines for the diagnosis and treatment of nasopharyngeal carcinoma. *Cancer Communications*. **41(11)**: 1195-1227.

Tîrziu, A., Avram, S., Mada, L., Crișan-Vida, M., Popovici, C., Popovici, D., Popovici, D., Faur, C., Duda-Seiman, C., Păunescu, V. & Vernic, C. (2023). Design of a synthetic long peptide vaccine targeting HPV-16 and-18 using immunoinformatic methods. *Pharmaceutics*. **15(7)**: 1798.

Toepak, E.P., Simarmata, S.N., Rahman, S. & Angga, S. C. (2022). Design of vaccines candidate based on ebola virus epitop with in-silico approach perancangan kandidat vaksin berbasis epitop virus ebola dengan pendekatan in-silico. *Jurnal Ilmiah Berkala: Sains dan Terapan Kimia*. **16(1)**: 57–63.

Tofighi, F.B., Saadati, A., Kholafazad-kordasht, H., Farshchi, F., Hasanzadeh, M. & Samiei, M. (2021). Electrochemical immunoplatform to assist in the diagnosis of oral cancer through the determination of CYFRA 21.1 biomarker in human saliva samples: Preparation of a novel portable biosensor toward non-invasive diagnosis of oral cancer. *Journal of Molecular Recognition*. **34(12)**: e2932.

Vilardi, F., Lorenz, H. & Dobberstein, B. (2011). WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *Journal of Cell Science*. **124(8)**: 1301-1307.

Wong, K.C., Hui, E.P., Lo, K.W., Lam, W.K.J., Johnson, D., Li, L., Tao, Q., Chan, K.C.A., To, K.F., King, A.D., Ma, N.B.Y. & Chan, A. T. (2021). Nasopharyngeal carcinoma: an evolving paradigm. *Nature Reviews Clinical Oncology*. **18(11)**: 679-695.

Wu, J., Carmen, A.A., Kobayashi, R., Suka, N. & Grunstein, M. (2001). HDA2 and HDA3 are related proteins that interact with and are essential for the activity of the yeast histone deacetylase HDA1. *Proceedings of the National Academy of Sciences*. California. 10 April 2001. pp 4391-4396.

Xu, L., Wang, S., Wu, Z., Xu, C., Hu, X., Ding, H., Zhang, Y., Shen, B. & Wu, K. (2022). Development of a colloidal gold immunochromatographic strip for rapid detection of CYFRA 21-1 in lymph node metastasis of thyroid cancer. *Frontiers in Bioengineering and Biotechnology*. **10**: 871285.

Xu, S. & Lou, Z. (2021). Ultrasensitive detection of nasopharyngeal carcinoma-related miRNA through garland rolling circle amplification integrated catalytic hairpin assembly. *ACS Omega*. **6(9)**: 6460-6465.

Yang, J.H., Williams, D., Kandiah, E., Fromme, P., & Chiu, P.L. (2020). Structural basis of redox modulation on chloroplast ATP synthase. *Communications Biology*. **3(1)**: 482.

Zambonin, C. & Aresta, A. (2022). MALDI-TOF/MS analysis of non-invasive human urine and saliva samples for the identification of new cancer biomarkers. *Molecules*. **27(6)**: 1925.

Zhao, T., Mao, G., & Chen, M. (2021). [Retracted] The role of change rates of CYFRA21-1 and CEA in predicting chemotherapy efficacy for non-small-cell lung cancer. *Computational and Mathematical Methods in Medicine*. **2021(1)**: 1951364.

Zwilling, D., Cypionka, A., Pohl, W.H., Fasshauer, D., Walla, P.J., Wahl, M.C. & Jahn, R. (2007). Early endosomal SNAREs form a structurally conserved SNARE complex and fuse liposomes with multiple topologies. *The EMBO Journal*. **26(1)**: 9-18.

Supplementary-1

```
CLUSTAL 2.1 multiple sequence alignment


C3                      -------------------------------------------------QS
D2                      -------------------------------------------------IQA
sp|P49767|VEGFC_HUMAN   -------------------------MHLLGFFSVACSLLAAALLPGPREA
sp|P01584|IL1B_HUMAN    -----------------------------------------MAEVPELASE
sp|P10145|IL8_HUMAN     ----------------------------------------MTSKLAVA
sp|P60174|TPIS_HUMAN    ----------------------------------------MAPSRKFFV
sp|P0DTE8|AMY1C_HUMAN   MKLFWLLFTIGFCWAQYSSNTQQGRTSIVHLFEWRWVDIALECERYLAPK

C3                      QLS-MKAALEDTLAETEARFGAQLAH-----------------------
D2                      LISGIEAQLGDVRADSERQNQEYQ-------------------------
sp|P49767|VEGFC_HUMAN   PAAAAAFESGLDLSDAEPDAGEATAYASKDLEEQLRSVSSVDELMTVLYP
sp|P01584|IL1B_HUMAN    MMAYYSGNEDDLFFEADGPKQMKCSFQDLDLCPLDGGIQLRISDHHYSKG
sp|P10145|IL8_HUMAN     LLAAFLISAALCEGAVLPRSAKELRCQCIKTYSKPFHPKFIKELRVIESG
sp|P60174|TPIS_HUMAN    GGNWKMNGRKQSLGELIGTLNAAKVPADTEVVCAPPTAYIDFARQKLDPK
sp|P0DTE8|AMY1C_HUMAN   GFGGVQVSPPNENVAIHNPFRPWWERYQPVSYKLCTRSGNEDEFRNMVTR

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   EYWKMYKCQLRKGGWQHNREQANLNSRTEETIKFAAAHYNTEILKSIDNE
sp|P01584|IL1B_HUMAN    FRQAASVVVAMDKLRKMLVPCPQTFQENDLSTFFPFIFEEEPIFFDTWDN
sp|P10145|IL8_HUMAN     PHCANTEIIVKLSDGRELCLDPKENWVQRVVEKFLKRAENS---------
sp|P60174|TPIS_HUMAN    IAVAAQNCYKVTNGAFTGEISPGMIKDCGATWVVLGHSERRHVFGESDEL
sp|P0DTE8|AMY1C_HUMAN   CNNVGVRIYVDAVINHMCGNAVSAGTSSTCGSYFNPGSRDFPAVPYSGWD

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   WRKTQCMPREVCIDVGKEFGVATNTFFKPPCVSVY---RCGGCCNSEGLQ
sp|P01584|IL1B_HUMAN    EAYVHDAPVRSLNCTLRDSQQKSLVMSGPYELKAL---HLQGQDMEQQVV
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    IGQKVAHALAEGLGVIACIGEKLDEREAGITEKVV---FEQTKVIADN--
sp|P0DTE8|AMY1C_HUMAN   FNDGKCKTGSGDIENYNDATQVRDCRLSGLLDLALGKDYVRSKIAEYMNH

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   CMNTSTSYLSKTLFEITVPLSQGPKPVTISFANHTSCRCMSKLDVYRQVH
sp|P01584|IL1B_HUMAN    FSMS------------FVQGEESNDKIPVALGLKEKNLYLSCVLKDDKPT
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    ------------------VKDWSKVVLAYEPVWAIGTGKTATPQQAQEVH
sp|P0DTE8|AMY1C_HUMAN   LIDIGVAGFRIDASKHMWPGDIKAILDKLHNLNSNWFPEGSKPFIYQEVI

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   SIIRRSLPATLPQCQAANKTCPTNYMWNNHICRCLAQEDFMFSSDAGDDS
sp|P01584|IL1B_HUMAN    LQLESVDPKNYPKKKMEKRFVFNKIEINNKLEFESAQFPNWYISTSQAEN
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    EKLRGWLKSNVSDAVAQSTRIIYGGSVTGATCKELASQPDVDGFLVGGAS
sp|P0DTE8|AMY1C_HUMAN   DLGGEPIKSSDYFGNGRVTEFKYGAKLGTVIRKWNGEKMSYLKNWGEGWG

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   TDGFHDICGPNKELDEETCQCVCRAGLRPASCGPHKELDRNSCQCVCKNK
sp|P01584|IL1B_HUMAN    MPVFLGGTKGGQDITDFTMQFVSS------------------------
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    LKPEFVDIINAKQ-----------------------------------
sp|P0DTE8|AMY1C_HUMAN   FMPSDRALVFVDNHDNQRGHGAGGASILTFWDARLYKMAVGFMLAHPYGF

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   LFPSQCGANREFDENTCQCVCKRTCPRNQPLNPGKCACECTESPQKCLLK
sp|P01584|IL1B_HUMAN    --------------------------------------------------
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    --------------------------------------------------
sp|P0DTE8|AMY1C_HUMAN   TRVMSSYRWPRYFENGKDVNDWVGPPNDNGVTKEVTINPDTTCGNDWVCE

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   GKKFHHQTCSCYRRPCTNRQKACEPGFSYSEEVCRCVPSYWKRPQMS---
sp|P01584|IL1B_HUMAN    --------------------------------------------------
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    --------------------------------------------------
sp|P0DTE8|AMY1C_HUMAN   HRWRQIRNMVNFRNVVDGQPFTNWYDNGSNQVAFGRGNRGFIVFNNDDWT

C3                      --------------------------------------------------
D2                      --------------------------------------------------
sp|P49767|VEGFC_HUMAN   --------------------------------------------------
sp|P01584|IL1B_HUMAN    --------------------------------------------------
sp|P10145|IL8_HUMAN     --------------------------------------------------
sp|P60174|TPIS_HUMAN    --------------------------------------------------
sp|P0DTE8|AMY1C_HUMAN   FSLTLQTGLPAGTYCDVISGDKINGNCTGIKIYVSDDGKAHFSISNSAED

C3                      -----------
D2                      -----------
sp|P49767|VEGFC_HUMAN   -----------
sp|P01584|IL1B_HUMAN    -----------
sp|P10145|IL8_HUMAN     -----------
sp|P60174|TPIS_HUMAN    -----------
sp|P0DTE8|AMY1C_HUMAN   PFIAIHAESKL
```

Figure 1. Sequence alignment with other protein saliva.

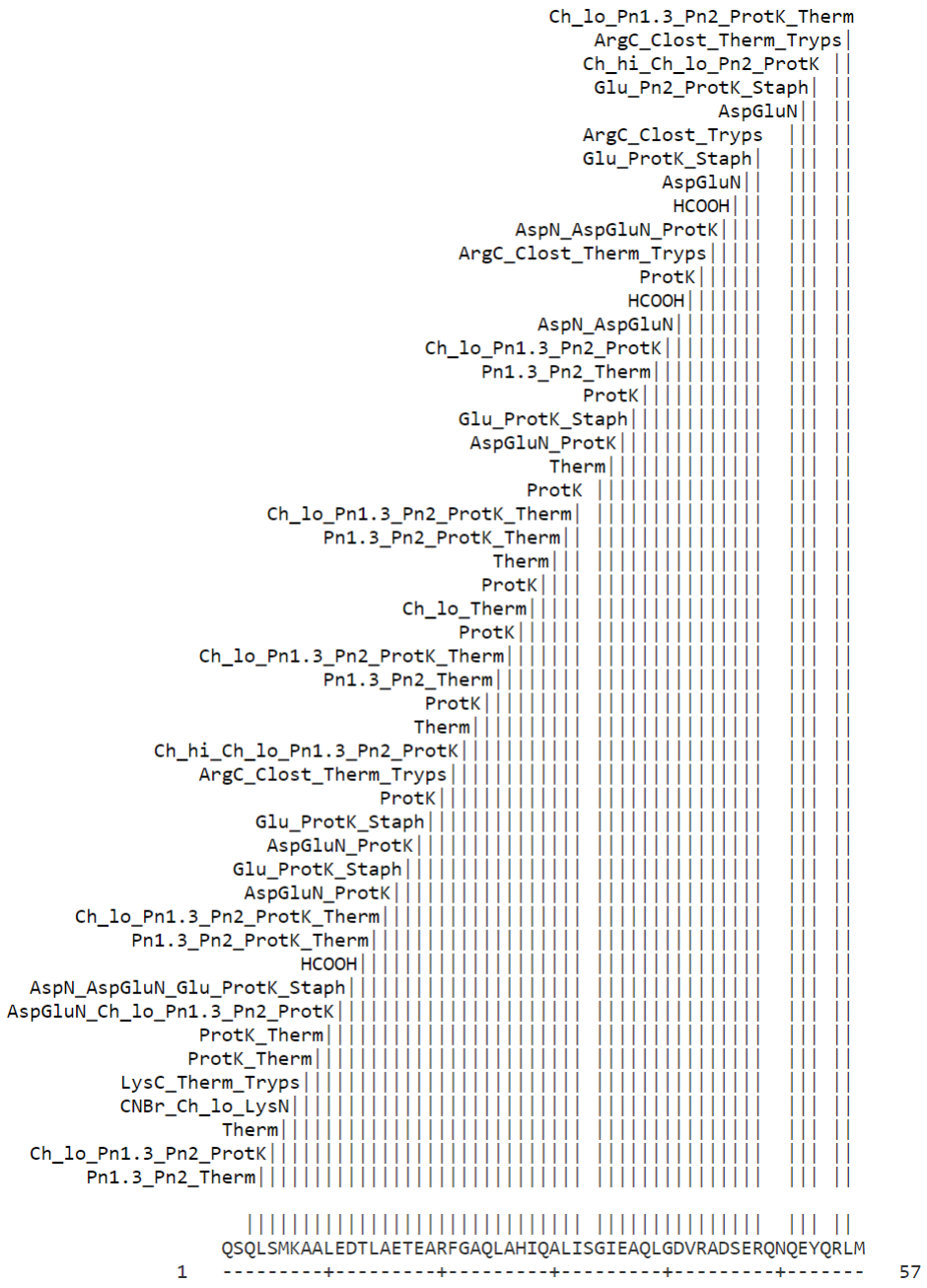

Figure 2. Proteolytic cleavages in CYFRA 21-1.