

ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)



Bio Technology



ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Conserved Region Analysis of Oncogenic Human *Papillomavirus* Genome

Usman Sumo Friend Tambunan, Herbert Wybert Butar-Butar,
Radya Umbas and Zulfa Hidayah
Department of Chemistry, Faculty of Science, University of Indonesia, Indonesia

Abstract: This study was carried out to determine the conserved regions of late genes from sequenced HPV types. HPV genome sequences were collected from the Los Alamos National Laboratory *papillomavirus* database. There are 74 types of HPV in the database which have completely documented genome sequences as well as their translation product. Specific types of HPV which may cause cervical cancer are grouped into high risk or low risk, according to their risk potential. This classification may differ from one research methodology to another. In order to access a representative classification, three sets of classification were studied for this research. HPV type 16 and 18 are consistently grouped as high risk, while other types of HPV varied randomly. Sequence alignment was taken and the result shows 62 conserved regions as a primer template for L1 and L2 genes. These conserved regions were then subjected to BLASTn operation in order to search the conserved region with least similarity to low risk HPV and human genome. Finally, 7 selected conserved regions were examined for secondary structures using NetPrimer program. From this operation, only region 52 (5'-ACAGGCTATGGTGCTATGGA-3') met the criteria to be used as an oligonucleotide primer.

Key words: Human *papillomavirus*, cervical cancer, hybrid capture, conserved regions

INTRODUCTION

Cervical cancer is a leading cause of death in women worldwide. In Indonesia, it is the cancer with highest malignancy occurring in females, with an estimated incidence of 25-40 per 100,000 women per year (De Boer *et al.*, 2004). The main etiologic factor for cervical cancer is from the infection of high-risk Human *papillomavirus* (HPV) (Park *et al.*, 2003).

HPV has double-stranded DNA and it belongs to the category of papovavirus family. There are more than 100 known types of HPV that are specifically attacking epithelial cells such as skin, respiratory mucosa and the genital tract. Genital tract HPV is further classified into low-risk and high-risk type according to their relative malignancy potential (Janicek *et al.*, 2001).

High-risk HPV type can induce changes in normal cells, which over some period of time may result in a malignant cell growth. This cancer can occur in genital area (cervix, vulva and anus), skin (non-melanoma skin cancer) and in the oropharynx (area located between head and neck). Among those included as high-risk HPV are type 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68 and 69. High risk HPV are found in almost every case of cervical cancer patient (Castle *et al.*, 2003). Meanwhile low-risk HPV type does not cause any malignancy, but may cause

benign hyperplastic lesion. This lesion can occur in both male and female genitals, for example like in vagina, cervix, vulva, penis and rectum.

Conventional method of detecting cervical cancer is done by carrying out cytological examination, which is more widely known as Pap smear. Due to rapid advancement of molecular biology, molecular based diagnostic and early detection methods on cervical cancer has been highly developed to replace conventional method of detection. Examples include polymerase chain reaction (PCR) and hybrid capture 1, 2 and 3. Hybrid capture 1 is liquid hybridization assaying method designed to detect 14 types of HPV, 9 of which are high risk (type 16, 18, 31, 33, 35, 45, 51, 52 and 56), while the other 5 are low risk (type 6, 11, 42, 43 and 44). Hybrid capture 2 is a development of hybrid capture 1 which uses microtitre plates instead of tubes and is capable of detecting four additional types of viral oncogenic (type 39, 58, 59 and 68) (Clavel *et al.*, 1998). Hybrid capture 3, similar to previous hybrid capture tests, relies on the formation of target HPV DNA-RNA probe heteroduplexes during the hybridization step in specimens containing sufficient HPV DNA. The chemiluminescent detection of these hybrids is by adding an alkaline phosphatase-conjugated monoclonal antibody specific to the DNA-RNA complexes with dioxetane substrate in a 96-well

enzyme-linked immunosorbent assay format (Lorincz and Anthony, 2001). Most of the above mentioned methods are specific, sensitive, reliable and easy to perform. Moreover, its routine application has been very much improved by the use of non-radioactive enzyme immunoassay detection procedure (Clavel *et al.*, 1998). Modification of the Hybrid Capture method, is expected to be achieved through the use of a customized oligonucleotide probe able to detect multiple high-risk HPV infection.

There are 10 gene encoded in HPV genome. These gene may be classified into two groups, namely Gene E (early) which encodes regulatory proteins and Gene L which encodes structural proteins. The region of L1 and L2 of Gene L is responsible for encoding capsid protein to be used as DNA envelope in HPV. The capsid protein will serve as protection system to HPV genetic materials. It is generally assumed that there are many nucleotides sequence in the region of L1 and L2 that is conserved throughout the evolution process of HPV (Dahlgren, 2005).

The aim of this study was to determine the conserved regions of late genes L1 and L2 from 74 sequenced and published HPV (Icenogle, 1995). The result was used to predict candidate templates for oligonucleotide probes that are specific on types of HPV types which cause cervical cancer. Nevertheless, the specific purpose of this study is to design primer that is able to detect on the open reading frames region and also to develop a new assay for the detection of high risk HPV DNA.

MATERIALS AND METHODS

Location and time of research: This online research was conducted early in 2006 at the Laboratory of Bioinformatics, Department of Chemistry, Faculty of Science, University of Indonesia.

HPV genome searching: HPV genome sequences were collected in GenBank Flatfile Format (GBFF) from the Los Alamos National Laboratory *papillomavirus* database (<http://hpv-web.lanl.gov/stdgen/virus/cgi-bin/hpv-organisms.cgi?dbname=hpv>).

Sequence alignment of HPV genome: A series of multiple alignments were carried out to HPV genomes that had already been classified according to its habitat and relevancy. HPV genomes were also classified into high risk and low risk groups (Park *et al.*, 2003). Sequence alignment was carried out for L1 and L2 open reading frame by using Clustal W version 1.8.

Table 1: Grouping of HPV based on high risk type and low risk type

HPV type	High risk	Low risk
Malloy <i>et al.</i> (2000)	16; 18; 31; 45	33; 35; 39; 51; 52; 56; 58; 59; 61; 68
Karlsen <i>et al.</i> (1996)	11; 16; 18; 31; 33; 35	39; 45; 51; 52; 56; 58; 59; 61; 68
Schellekens <i>et al.</i> (2004)	16; 18; 52	31; 33; 35; 45; 51; 58; 59; 61; 68

Conserved region and template searching: Conserved region and template were thoroughly searched using Bio Edit version 7.0.1

Template region and associated human genome searching and evaluation of primer candidates: Selection of the template regions was done by database similarity searching, using BLASTn (Lipman and Pearson, 1985). NetPrimer was used to evaluate the ability of the selected regions to be used as oligonucleotide primers.

RESULTS AND DISCUSSION

HPV genome searching: Based on the database collected from Los Alamos National Laboratory, there are 100 known types of HPV, but only 74 among them which have complete documented genome sequences and translation products.

Grouping of HPV genomes: It is necessary to group all this 74 types according to their risk potential (high risk or low risk) before aligning the sequences. Only specific types of HPV cause cervical cancer and this widely vary for patients from different parts of the world. For this reason, we had grouped HPV by using 3 major literatures as Malloy *et al.* (2000), Karlsen *et al.* (1996) Malloy *et al.* (2000) and Schellekens *et al.* (2004), which were supported by laboratory experiments from cervical cancer patient's specimens. The result of this grouping is shown in Table 1.

According to Table 1, each HPV type shows type 16 and 18 as high risk, while other types of HPV varies randomly. In order for the input to be recognized by bioinformatical programs which will be used, all the sequences were converted from GBFF format into FASTA format. Position of Gene L1 and L2 from the nucleotide sequences were then identified for each type of HPV. The total genome of one HPV sequence consists of 8000 nucleotide base pairs. Therefore it will be very difficult to manually find those position. For this reason, Bio Edit program was used as a tool to cut sequences on the targeted positions.

Sequence alignment, conserved region and template searching of HPV genomes: The aim of sequence alignment is to find highest similarity among inputs.

Sequence alignments in this study were carried out using CLUSTALW version 1.83. Sequence alignment was conducted to HPV that had first been grouped according to their ability to initiate normal cell into cancerous cell (High-Risk and Low-Risk cancerous). The results showed 62 conserved regions as a primer template for L1 and L2 genes.

The parameters used for determining conserve regions from sequence alignments were gap limit until x in conserved region, limit contiguous gap till x, average maximum of entropy and the sum of exceptions taken from the maximum entropy that was chosen.

The results were 14 conserved regions for alignment L1 genes for HPV type 16, 18, 31 and 45. Alignment for L1 genes for HPV type 16, 18, 31, 33, 35, 39, 45, 51, 52, 58, 59, 61 and 68 resulted 4 conserved regions. Alignment L1 genes for HPV type 11, 16, 18, 31 and 35 resulted 8 conserved regions. Alignment L1 genes for HPV type 16, 18 and 52 resulted 4 conserved regions. Meanwhile, alignment for L2 genes HPV type 16, 18, 31 and 45 resulted 4 conserved regions. Alignment for L2 genes position for HPV type 16, 18, 31, 33, 35, 39, 45, 51, 52, 58, 59, 61 and 68 resulted 1 conserved region. Alignment L2 genes for HPV type 11, 16, 18, 31, 35 resulted 1 conserved region. Alignment L2 genes for HPV type 16, 18 and 52 resulted 6 conserved regions. The result in overall alignments were 62 conserved regions (Table 2). From these regions, sequences that uniquely occur in oncogenic HPV were selected by using BLASTn. From the BLAST results, regions with high similarity to oncogenic HPV types and low similarity to non-cancerous HPV types were then compared to the patented universal primer used to detect oncogenic HPV 5'-TTTGTTACTGTGGTAGATAC-3' (Anthony *et al.*, 2004).

Seven regions were selected for use as templates, region 1 from the alignment of L1 genes from HPV type 16, 18, 31 and 45, regions 21, 31, 43, 45 and 46 from the alignment of L1 genes from HPV type 11, 16, 18, 31, 35 and 68 and region 52 from the alignment of L2 genes from HPV type 16, 18 and 52 (Table 2).

Comparison of the selected templates to the human genome was also carried out in order to assess their occurrence in the human genome. From the BLAST results for the 7 regions, there were 4 regions that have no similarity with the human genome; namely region 31, 43, 45 and 52. Region 45 gave hit on human DNA cloning from chromosome 5 and 10, while region 31 and 43 gave hit on chromosome 3, 5, 9, 13 and 22. Region 52 gave hit on chromosome 9, 13 and 22 (Table 2).

Evaluation of primer candidates: Oligonucleotide primers are considered based upon certain properties, namely:

Table 2: Conserved regions selected as templates

Nucleotide Sequence	Alignment	Region	BLASTn result and Position on genome according to HPV type
5'-CATATAT	Type	1	Type 16 (5734-5751)
TATCATGCA	16; 18; 31; 45		Type 31 (5647-5668)
GGCAGTTC-3'			Type 33 (1920-1929)
			Type 35 (5703-5719)
			Type 45 (5703-5726)
			Type 51 (5616-5639)
			Type 52 (5742-5761)
			Type 53 (7424-7434)
			Type 56 (5693-5715)
			Type 59 (5714-5724)
			Type 66 (5743-5759)
5'-GATGGGG	Type	21	Type 11 (6350-6372)
ATATGGTTG	11; 16; 18; 31;		Type 16 (6227-6246)
ATACAGGAT-3'	35; 68		Type 31 (6143-6165)
			Type 68 (2205-2229)
5'-ACATATGAT	Type	31	Type 16 (4987-5003)
AACCGTGC-3'	11; 16; 18; 31;		Type 33 (4960-4976)
	35; 68		Type 35 (4958-4974)
			Type 58 (4994-5004)
			Type 59 (6143-6165)
5'-ACCGTTTA	Type	43	Type 16 (5933-5945)
GTTTGGCCCT	11; 16; 18;		Type 31(5846-5861)
TG-3'	31; 35; 68		Type 35 (5889-5906)
			Type 39 (7001-7014)
			Type 45 (2951-2961)
			Type 52 (7096-7083)
			Type 68 (4997-4987)
5'-CACAACTTT	Type	45	Type 06(6063-6081)
AGTTTGGCC-3'	11; 16; 18; 31;		Type 18 (5896-5915)
	35; 68		Type 31 (5840-5857)
			Type 39 (5926-5944)
			Type 45 (2951-2961)
			Type 52 (7069-7084)
5'-GTGTGGCAT	Type	46	Type 16 (5979-5989)
TAGTGGCCAT-3	11; 16; 18;		Type 31 (2383-2393)
	31; 35;68		Type 42 (6176-6193)
5'-ACAGGCTATG	Type	52	Type 16 (6248-6264)
GTGCTATGGA-3'	16; 18; 52		Type 61 (6373-6383)
			Type 68 (2223-2242)
			Type 70 (6190-6209)

they must not have potential secondary structures such as hairpins or dimmers; have a GC content of 45-60%; have a Tm between 52-58°C; their 5' ends stability has to be greater than the stability of their 3' ends; be 17-25 nucleotides in length (Dieffenbach *et al.*, 1995). From NetPrimer analysis results of the seven regions, only region 52 meet with the above mentioned criteria, with a NetPrimer rating of 100 (maximum).

CONCLUSION

Based on sequence similarity, 62 conserved regions were found. Out of the 62, 7 regions were then used as templates for primers used in detection of high and low risk HPV. From the 7 template candidates, only one met the criteria to be used as an oligonucleotide primer, namely region 52. From the study, region 52 is predicted to be selective to be used in the detection of oncogenic Human papillomavirus.

ACKNOWLEDGMENTS

This research was supported by Ministry of Education (Grant Hibah Pasca No. 011/SP3/PP/DP2M/II/2006). The authors are grateful to Dr. Jarnuzi Gunlajuardi, Chairman of Department of Chemistry, Faculty of Science, University of Indonesia for his support and critical comments on the manuscript.

REFERENCES

- Anthony, J., A. Lorincz, I. Williams, J. Troy and Y. Tang, 2004. Detection of Nucleic Acid by Type-Specific Hybrid Capture Method. United States Patent No. 20040214302.
- Castle P.E. *et al.*, 2003. Comparison between prototype hibrid capture 3 and hybrid capture 2 human papillomavirus DNA assay for detection of high-grade cervical intraepithelial neoplasia and cancer. *J. Clin. Microbiol.*, 41: 4022-4030.
- Clavel, C., M. Masure and I. Putaut, 1998. Hybrid capture II, a new sensitive test for human *papillomavirus* detection. Comparison with hybrid capture I and PCR results in cervical lesions. *J. Clin. Pathol.*, 51: 737-740.
- Dahlgren, L., 2005. Studies on the Presence and Influence of human *papillomavirus* (HPV) in Head and Neck Tumors. Stockholms
- De Boer, M.A., L.A. Peters, M.F. Aziz, B. Siregar, S. Cornain, M.A. Vrede, F.S. Jordanova, S. Kolkman-Uljee and G.J. Fleuren, 2004. Human *papillomavirus* Type 16 E6 E7 and li variants in cervical cancer in Indonesia, Suriname and The Netherlands. *Gynecol. Oncol.*, 94: 488-494.
- Dieffenbach, C.W., T.M.J. Lowe and G.S. Dveksler, 1995. PCR Primer, A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.
- Icenogle, J., 1995. Analysis of the Sequences of the L1 and L2 Capsid Proteins of *Papillomaviruses*. Centers for Disease Control, Atlanta, Georgia.
- Janicek, M.E. and H.E. Averet, 2001. Cervical cancer: Prevention, diagnosis and therapeutics. *CA Cancer J. Clin.*, 51: 92-114.
- Karlsen F., M. Kalantari, A. Jenkins, E. Petersen, G. Kristensen, R. Holm, B. Johansson and B. Hagmar, 1996. Use of multiple pcr primer sets for optimal detection of human *papillomavirus*. *J. Clin. Microbiol.*, 34: 2095-2108.
- Lipman, D.J. and W.R. Pearson, 1985. Rapid and sensitive protein similarity searches. *Science*, 227: 1435-1441.
- Lorincz, A. and J. Anthony, 2001. Hybrid capture method of detection of human *Papillomavirus* DNA in clinical specimens. *Papillomavirus Rep.*, 12: 145-154.
- Malloy, C., J. Sherris and C. Herdman, 2000. HPV DNA Testing: Technical and Programmatic Issues for Cervical Cancer in Low- Resource Setting. Alliance for Cervical Cancer Prevention.
- Park, Seong Bae, Sohyun Hwang and Byoung-Tak Zhang, 2003. Classification of human *papillomavirus* (HPH) risk type via text mining. *Genomics and Informatics*, 1: 80-86.
- Schellekens, M.C., A. Dijkman, M.F. Aziz, B. Siregar, S. Cornain, M.A. Vrede, F.S. Jordanova, S. Kolkman-Uljee, L.A. Peters and G.J. Fleuren, 2004. Prevalence of single and multiple HPV Types in cervical carcinomas in Jakarta, Indonesia. *Gynecol. Oncol.*, 93: 49-53.