Atom Indonesia

Molecular and *in silico* Study of TP53 Codon 72 Polymorphism (rs1042522) in a Population Exposed to High Background Radiation in Mamuju-West Sulawesi

D. Tetriana[1], S. Purnami[1], T. Rahardjo[1], W. Mailana[1], S. Nurhayati[1], E. Pudjadi[1], S. Tri Widyaningtyas[3], T. Ishida[4], and D. Ramadhani[2,1*]

[1]*Center for Safety Technology and Radiation Metrology, National Nuclear Energy Agency (BATAN),*
Jl. Lebak Bulus Raya No. 49, Pasar Jum'at, Jakarta Selatan 12070, Indonesia.
[2]*Doctoral Program for Biomedical Sciences, Faculty of Medicine, Universitas Indonesia (UI),*
Jl. Salemba Raya 6, Jakarta 10430, Indonesia.
[3]*Virology and Cancer Pathobiology Research Center, Faculty of Medicine, Universitas Indonesia (UI),*
Jl. Salemba Raya 4, Jakarta 10430, Indonesia.
[4]*Unit of Human Biology and Genetics, Department of Biological Sciences, Graduate School of Science,*
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

ARTICLE INFO

ABSTRACT

The evaluation of the tumor protein p53 (TP53) codon 72 polymorphism (rs1042522) status in a population exposed to high background radiation was performed in this study. Real-time polymerase chain reaction (Q-PCR) was used to genotype the rs1042522 polymorphism in 100 subjects from Takandeang, Salleto, and Ahu villages in Mamuju district, West Sulawesi. An *in silico* study was then conducted to identify the potential effects of the proline substitution associated with this polymorphism on protein stability. The *in silico* analysis was performed using three different computational tools, namely I-Mutant Suite, iStable, and Protein Variation Effect Analyzer (PROVEAN). Secondary and three-dimensional (3D) structural models for wild-type (WT) and variant TP53 were generated to predict potential structural changes in the protein. Electrostatic surface charge calculations were also performed to identify changes in the electrostatic charge of codon 72. The Pro72 and Arg72 frequencies among the inhabitants in the Takandeang, Salleto, and Ahu villages, who are Mandarese Austronesian-speaking ethnic group members, were 0.56 and 0.44, respectively. The *in silico* analysis revealed no negative effects on protein stability due to the proline substitution at codon 72, although molecular modeling showed several differences in the secondary and 3D structures of the TP53 variant compared with the structure of WT TP53. To determine the impact of the proline-coding allele on individual sensitivity to radiation exposure, we compared the micronucleus (MN) frequencies of each genotype. The results showed that the MN frequencies in individuals harboring the proline-coding allele were not significantly higher than those expressing the arginine-coding allele. In conclusion, this study revealed that the proline-coding allele for codon 72 of the TP53 gene represented the predominant genotype among Takandeang, Salleto, and Ahu villages inhabitants. The present study also demonstrated that individuals who carried the proline-coding allele were not more sensitive to radiation compared with those expressing the arginine-coding allele for codon 72 of the TP53 gene.

INTRODUCTION

Regions with high levels of environmental radioactivity exist in several locations around the world, such as Ramsar (Iran), Guarapari (Brazil), Kerala (India), and Yangjiang (China) [1,2]. Mamuju in West Sulawesi, Indonesia, is also known for high levels of environmental radioactivity, due to the natural uranium contents of approximately eight times as high as the average for the Earth's crust, which is approximately 3 ppm of equivalent uranium

(eU). Six villages (Ahu, Takandeang, Botteng, Pengasaan, Tande-Tande, and Mamunyu) in Mamuju are considered to be anomalous areas due to background radiation dose rates higher than 400 nSv/h [3].

People who live in areas with high environmental radioactivity levels and express the homozygous proline-coding allele (C/C) genotype for codon 72 of the tumor protein p53 gene (*TP53*) should be subjected to public health evaluations. Individuals with the C/C genotype are considered to be more sensitive to radiation exposure because proteins containing proline are less effective as transcription factors normally associated with the cell cycle, the activation of apoptosis, and the expression of genes related to DNA repair, which can lead to a reduced DNA damage repair capacity [4]. Cells with reduced DNA damage repair capacities are at risk of chromosomal instabilities, such as the development of dicentric chromosomes, acentric fragments, or micronucleus (MN) formation, that occur when chromosome fragments or whole chromosomes develop their own nuclear envelopes after failing to merge with daughter nuclei [5,6].

The *TP53* in humans is located on chromosome 17p13.1 and consists of 11 exons [7,8]. The *TP53* is often referred to as the "guardian of the genome" because it plays central roles in the prevention of carcinogenesis, the induction of cell cycle arrest, and apoptosis [9,10]. *TP53* has been extensively studied in humans, particularly its correlation with cancer susceptibility. *TP53* is also highly polymorphic. Furthermore, one of the most widely-investigated single nucleotide polymorphisms (SNPs) in *TP53* is the rs1042522 polymorphism in codon 72 on exon 4, which results in the translation of either a proline or an arginine residue [11]. The TP53 protein is a homotetramer composed of 393 amino acids. It consists of three primary domains, the N-terminal domain (NTD), the DNA-binding domain (DBD), and the C-terminal domain (CTD). The NTD consists of an acidic N-terminal transcription activation region and a proline-rich region (PRR). The CTD consists of a tetramerization domain (TD) and a regulatory domain (RD) [12]. The rs1042522 polymorphism is located in the PRR of TP53 protein, which contains 12 proline residues and the PXXP sequence motif.

In a previous study, the rs1042522 status in twenty samples taken from Takandeang sub-village inhabitants was evaluated [13]. To obtain more precise information regarding the allelic distribution of TP53 codon 72 variants among Mamuju inhabitants, more samples were screened in this study. Single-amino-acid changes caused by SNPs can cause changes in protein structural stability by producing smaller free energy changes (ΔG, or dG) after folding. The differences in folding free energy changes between wild-type (WT) and mutant protein isoforms ($\Delta\Delta G$, or ddG) are often considered to be impact factors associated with changes in protein stability [14]. Based on these explanations, *in silico* predictions and molecular modeling were also performed for TP53 protein to determine the potential effects of the codon 72 polymorphism on the TP53 protein structure. The association between rs1042522 status and the presence of MNs induced by high background radiation exposure was also evaluated in the present study, to determine whether individuals with the C/C genotype were more prone to chromosomal aberrations that result in MN formation due to less efficient DNA repair processes.

EXPERIMENTAL METHODS

Annual effective dose calculation

The detailed process used to perform annual effective dose calculations was described in the previous study [13]. In brief, indoor and outdoor background gamma radiation measurements were performed using a gamma spectrometer (Exploranium GR-135 Plus) to calculate the annual effective doses due to the indoor and outdoor background gamma dose rates in villages in the Takandeang, Salleto, and Ahu regions. The Takandeang, Salleto, and Ahu villages are part of the Mamuju District, located between 1° 38' 110'' - 2° 54' 552'' South and 11° 54' 47'' - 13° 5' 35'' East.

Blood sampling

A total of 100 individuals from Takandeang, Salleto, and Ahu villages were analyzed to assess their rs1042522 status. Blood samples were collected by venipuncture using ethylenediaminetetraacetic acid (EDTA) vacutainer tubes. Informed consent was obtained from all donors. The study was approved by the Ethics Committee of the Indonesian Ministry of Health (number LB.02.01/5.2.KE.051/2015). Detailed descriptions of the inclusion and exclusion criteria used for donors can be found in the previous study [13].

DNA extraction and genotyping

Genomic DNA extraction and TP53 Arg72Pro (rs1042522) polymorphism genotyping were performed as described in [13]. The primer pairs used for rs1042522 were purchased from Applied

Biosystems with assay ID number C_2403545_10. In brief, 2 μL PCR primer, 10 μL Taqman GTXpress Mastermix for genotyping, 4 μL genomic DNA, and 4 μL PCR-grade water were mixed and subjected to the following PCR protocol: 10 min denaturation at 95 °C, 45 cycles at 95 °C for 15 seconds, and 60 °C for 60 seconds.

In silico analysis of the rs1042522 effect to TP53 protein

The amino acid sequence of the TP53 protein was retrieved from the National Center for Biotechnology Information (NCBI) database, using accession number NP_000537. Because rs1042522 is located in the PPR of TP53, it could interfere with protein stability. Therefore, three different online computational tools were chosen to perform *in silico* analyses to determine the impact of the rs1042522 polymorphism on TP53 protein stability. All three computational tools were capable of predicting the effects of the SNP on protein stability. The online tools used were I-Mutant Suite, iStable, and Protein Variation Effect Analyzer (PROVEAN). I-Mutant Suite is a support vector machine (SVM) that calculates changes in protein stability induced by specific SNP. The inputs included the WT protein sequence, the mutated residue, temperature, and pH, and the output reported whether or not the mutated protein would be stable. The output was the free energy change value (ddG) of the protein, both before and after mutation. Based on the ddG value, the mutated protein was categorized as either neutral or having a large decrease in stability (ddG = 0.5 kcal/mol) [12]. iStable, an integrated predictor, integrates the results from several different element predictors to predict protein stability following a single-amino-acid mutation. The integration aims to combine the results from different predictors based on the logic that the predictive power of multiple methods is larger than that for any single method alone [15]. PROVEAN is an online computational tool that can predict stability not only for single-amino-acid substitutions but also for multiple-amino-acid substitutions, insertions, and deletions, using the same underlying scoring scheme. PROVEAN categorizes the output as 'deleterious' or 'neutral', based on the amino acid sequence information [16].

Molecular modelling and electrostatic surface charge calculations of TP53 protein

Molecular models of the three-dimensional (3D) protein structures of WT TP53 and variant TP53 were generated using RaptorX (http://raptorx.uchicago.edu/StructurePrediction/predict/). The structures of WT and variant TP53 were superimposed using PyMol. The electrostatic surface charge calculations for WT and variant TP53 were performed using the PDB2PQR server (http://nbcr-222.ucsd.edu/pdb2pqr_2.1.1/) and then visualized using PyMol. The secondary structure predictions for WT and variant TP53 were conducted using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/).

Micronucleus (MN) assay

The MN assay was conducted as described in previous publication [13]. Blood samples were cultured in medium consisting of 4.5 mL Roswell Park Memorial Institute (RPMI) 1640 supplemented with 20 % heat-inactivated fetal bovine serum (FBS) 1 % streptomycin/penicillin, and 0.1 mL phytohemagglutinin (PHA); in a 5 % CO_2 atm at 37 ^{o}C for 72 h. Cytochalasin-B was then added to the culture at 44 h at a final concentration of 6 μg/mL. The blood cultures were then centrifuged for 10 min at 1000 rpm and resuspended in 7 mL 0.075 M cold (4 °C) KCl. Subsequently, the tubes were centrifuged for 8 min at 1000 rpm and the cells were resuspended in freshly prepared fixative consisting of methanol:acetic acid (10:1) diluted 1:1 with Ringer's solution. The cells were then washed three times with freshly prepared fixative, consisting of methanol:acetic acid (10:1) without Ringer's solution, until the cell suspension was clear. The cell suspension then was stored at -20 °C for at least one night until slide preparation was performed. Fixed cells were dropped onto clean, wet slides and then dried and stained with 4 % Giemsa solution (pH 6.8) for 12 mins. Three slides were prepared for each sample. For each sample, one reader scored 1000 binucleated cells (BNC) from at least two slides, at a magnification of 400×.

Statistical analysis

The Hardy-Weinberg equilibrium (HWE) was tested on genotypic data using SNPStats (https://www.snpstats.net/start.htm) [17]. The genotype distribution was considered to follow the HWE if the *p*-value was greater than 0.05. The significance of differences in the MN frequencies between each genotype was evaluated using a one-way analysis of variance (ANOVA). A *p*-value of less than 0.05 ($p<0.05$) was considered significant. All tests were conducted using Statistical Package for Social Sciences (SPSS) for Windows, version 22.0.

RESULTS AND DISCUSSION

Annual effective dose

The average annual effective dose for the Takandeang, Salleto, and Ahu villages was 2.95 mSv/y, with a range from 1.99 to 5.33 mSv/y, which is 2.2 - 6.12 times as great as the average effective environmental gamma dose rate due to cosmic rays and terrestrial gamma radiation estimated for the world in a report by the United Nations Scientific Committee on the Effects of Atomic Radiation (0.87 mSv/y). However, as described in a previous publication, the average annual effective dose was calculated without considering ingestion and inhalation exposure due to radon gas [18]. The annual effective dose for the Takandeang, Salleto, and Ahu villages were comparable with each other. In detail, the average annual effective dose for Takandeang village was 2.52 mSv/y. The average annual effective dose for Salleto and Ahu villages was higher than for Takandeang which was 3.38 mSv/y.

Distribution of the *TP53* polymorphism (rs1042522) in Takandeang, Salleto, and Ahu village inhabitants

The genotype distribution among Takandeang, Salleto, and Ahu village inhabitants followed the HWE ($p = 0.84$) (Table 1). The present study revealed that the WT genotype among Takandeang, Salleto, and Ahu village inhabitants was the C/C allele that encoded a proline at codon 72. Based on this finding, the presence of an arginine residue will be considered the variant TP53 protein for this region. The Arg72 allele frequency among all samples of the Mandarese Austronesian-speaking ethnic group members was 0.44. A previous study revealed that the Arg72 allele frequencies among the two largest ethnic groups in South Sulawesi (Torajans and Buginese) from locations near West Sulawesi were 0.45 and 0.55, respectively [19].

Table 1. Distribution of each TP53 genotype among Takandeang, Salleto, and Ahu village inhabitants

Gene	Genotype	Number	*p*-value of HWE
TP53	*CC*	31	
	CG	51	
	GG	18	0.84
Allele	*C*	113	
frequency	*G*	87	

Based on those findings, it can be postulated that the historical backgrounds of the Takandeang,

Salleto, and Ahu village inhabitants were potentially closer to the Torajans than to the Buginese. However, a more extensive analysis should be performed to validate this hypothesis. Moreover, the rs1042522 status among Torajans and Bugines should also be validated to further verify this hypothesis.

In silico analyses of the effect of the rs1042522 polymorphism on TP53 protein structure

The *in silico* analysis results examining the effect of the rs1042522 polymorphism on TP53 protein structure using three different online computational tools, revealed that the substitution from proline to arginine did not have a negative effect on protein stability. Analysis using I-Mutant Suite predicted an increase in the structural stability of the variant protein with a ddG value of -0.29. Similarly, iStable predicted an increase in the protein structural stability due to the substitution from proline to arginine with a score of 0.703838. The analysis using PROVEAN predicted the result of the proline substitution to be neutral with a score of -0.230 (Table 2). Overall, the *in silico* analyses of the effects of the proline to arginine substitution did not show any negative effects on protein stability.

Table 2. *In silico* results using three different computational tools

Computational Tools	Prediction	Value/Score
I-Mutant	Increase	-0.29
iStable	Increase	0.703838
PROVEAN	Neutral	-0.230

Molecular modeling and electrostatic surface charge calculations for TP53 protein

Visualization of the 3D structures of WT and variant TP53 revealed that position 72 in the TP53 protein was located in the loop region. The proline-to-arginine substitution at position 72 represents a change from a nonpolar residue to a polar residue. The superimposed 3D structures of WT and variant TP53 models are shown in Fig. 1, which shows a large difference between the WT and variant TP53 structural models. Proline has a very rigid structure; therefore, the proline to arginine substitution would be expected to result in a conformational change [20]. Arginine also has a smaller size than proline. The proline-to-arginine substitution can also lead to the loss of hydrophobic interactions with other amino acids. In addition, the charges possessed by mutant amino acids can cause repulsion with

neighboring amino acids [21]. The electrostatic surface charge calculations also reveal that the interface of codon 72, which is equally positive and negative for proline, becomes mostly positive for arginine (Figs. 2 and 3).
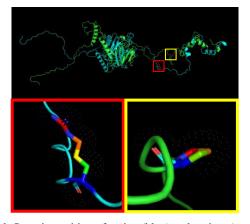

Fig. 1. Superimposition of native (blue) and variant (green) TP53 modeled cartoon structures. The amino acids arginine and proline are presented as stick shapes (red and yellow squares) at position 72
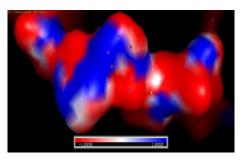

Fig. 2. Electrostatic surface charge model of WT (Pro72) TP53. Dots denote the proline at position 72
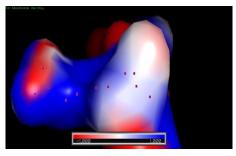

Fig. 3. Electrostatic surface charge model of the Arg72 TP53 variant. Dots denote the arginine at position 72

In this study, also evaluated were the secondary structures of WT and variant TP53. The secondary structure prediction revealed a structural rearrangement due to the variant sequence compared with the native sequence. The variant TP53 protein showed a longer beta-strand in region 110 to 115, and shorter beta-strands for the regions 124-126, 193-198, 251-258, and 263-276. Overall,

both 3D and secondary structure modeling revealed significant differences between WT and variant TP53 structures. Although large differences were observed between WT and variant TP53, the *in silico* analyses showed no adverse effects on protein stability caused by the proline to arginine substitution at codon 72. Thus, to determine whether individuals with arginine at codon 72 of the TP53 protein have reduced levels of chromosomal instability as a result of increased DNA damage repair capacities, we evaluated the association between the Pro72 polymorphism and MN formation.

Association between the Pro72 TP53 polymorphism and MN formation induced by high level background radiation exposure

In a previous study, no association was found between Pro72 TP53 polymorphism and individual radiosensitivity among Mamuju inhabitants assessed by a G_2 MN assay [13]. Interestingly, the present study also found no association between the Pro72 TP53 polymorphism and baseline MN frequencies among Mamuju inhabitants. As shown in Fig. 4, the MN frequencies of the two genotypes were not significantly different ($p>0.05$). To the best of our knowledge, no study has reported an association between the Pro72 TP53 polymorphism and MN frequency in populations living with high-level background radiation exposure. The present study is considered to be the first study to evaluate the correlation between the Pro72 TP53 polymorphism and the baseline MN frequency in lymphocytes from people exposed to high-level background radiation.
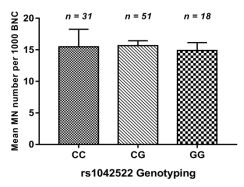

Fig. 4. Mean MN numbers (per 1,000 BNC) for each genotype (*n*: number of samples)

Bayram *et al*. in 2016 evaluated the association between the Pro72 TP53 polymorphism and MN frequency in a population exposed to a different environmental genotoxicant, aflatoxin B1 [22]. Unfortunately, because the environmental

genotoxicant evaluated in this study was radiation exposure, the results of this study are difficult to compare to those reported by Bayram *et al*. However, they stated that the Arg72 and Pro72 alleles have different functions, which could affect the DNA repair capacity of TP53 and influence the MN frequencies. Interestingly, even though the present study revealed different 3D and secondary structures for the WT and variant TP53 proteins, the *in silico* analysis showed no changes in protein stability. The comparison of MN frequencies observed for each genotype also supported the *in silico* analysis results. A possible explanation for the finding of the current study is that the Pro72Arg substitution is probably not suitable for reflecting differences among individual sensitivities to radiation exposure. Other SNPs, particularly those located in the DBD region of *TP53*, would likely better reflect variations among individual sensitivities to radiation exposure. *TP53* is a well-known polymorphic gene [7], with more than 200 polymorphic variants reported in the exon, promoter, and intron regions [21]. The hotspot mutations that affect *TP53* are usually located in the DBD (exons 5-8), which plays key roles in the transcriptional activation of TP53 target genes [23]. Other genes that have strong associations with MN frequencies are those involved in DNA damage repair, such as the X-ray repair cross-complementing group 1 (*XRCC1*) or X-ray repair cross-complementing group 3 (*XRCC3*) genes, which are involved in single-strand break (SSB) repairs, through the base excision repair (BER) pathway, and double-strand break (DSB) repairs, through the homolog recombination (HR) pathway [24,25].

CONCLUSION

The results show that the WT TP53 genotype among Takandeang, Salleto, and Ahu village inhabitants was the C/C allele, which encodes a proline at codon 72. *In silico* analyses showed that the proline-to-arginine substitution did not affect protein stability, although molecular modeling and electrostatic surface charge calculations showed large differences between the WT and variant TP53 proteins. In addition, no association between the Pro72 TP53 polymorphism and MN frequencies was found in the present study. Further investigations of other SNPs in *TP53*, particularly those located in the DBD region, or of different genes involved in the DNA damage repair process should be performed in the future.

REFERENCES

1. C.V. Karuppasamy, E.N. Ramachandran, V.A. Kumar *et al*., Mutat. Res. - Genet. Toxicol. Environ. Mutagen. **800-801** (2016) 40.

2. V. Jain, D. Saini, P.R.V. Kumar *et al*., Mutat. Res. Mol. Mech. Mutagen. **806** (2017) 39.

3. H. Syaeful, I.G. Sukadana and A. Sumaryanto, Atom Indonesia **40** (2014) 33.

4. I. Brasil-Costa, D.O. Alencar, M. Raiol-Moraes *et al*., Biomed Res. Int. **2013** (2013) 1.

5. M. Soto, I. García-Santisteban, L. Krenning *et al*., J. Cell Sci. **131** (2018) jcs214742.

6. J. Pajic, B. Rakic, B. Rovcanin *et al*., Radiat. Environ. Biophys. **54** (2015) 317.

7. J. Huszno and E. Grzybowska, Oncol. Lett. **16** (2018) 34.

8. Y. Zhang, E3S Web Conf. **78** (2019) 1.

9. A. Coelho, A. Nogueira, S. Soares *et al*., Oncol. Lett. **15** (2018) 8165.

10. E. Toufektchan and F. Toledo, Cancers (Basel). **10** (2018) 1.

11. N. Bilous, I. Abramenko, A. Chumak *et al*., Exp. Oncol. **38** (2016) 252.

12. R.M. Elfaki, M.S. Abdelaziz, H.N. Altayb, *et al*., F1000Research **7** (2018) 1.

13. D. Ramadhani, S. Purnami, S. Nurhayati *et al*., Atom Indonesia **45** (2019) 27.

14. C.W. Chen, J. Lin and Y.W. Chu, BMC Bioinformatics **14** (2013) S5.

15. S.P. Sujitha, D.T. Kumar, C.G.P. Doss *et al*., PLoS One **11** (2016) 1.

16. Y. Choi and A.P. Chan, Bioinformatics **31** (2015) 2745.

17. X. Solé, E. Guinó, J. Valls *et al*., Bioinformatics **22** (2006) 1928.

18. D. Ramadhani, S. Nurhayati, T. Rahardjo *et al*., Indones. Biomed. J. **10** (2018).

19. T. Kashima, K. Makino, A. Soemantri *et al*., J. Hum. Genet. **52** (2007) 694.

20. C.G.P. Doss, B. Rajith, N. Garwasis *et al*., Appl. Transl. Genomics **1** (2012) 37.

21. J.S.M. Sabir, A. El, N.A. Shaik *et al*., Saudi J. Biol. Sci. **XXX** (2019) In Press.

22. S. Bayram, E. Rencüzoğulları, A.M. Almas *et al*., Drug Chem. Toxicol. **39** (2016) 331.

23. S. Antoun, D. Atallah, R. Tahtouh *et al*., Cancer Cell Int. **18** (2018) 1.

24. M. Shakeri, F. Zakeri, V. Changizi *et al*., Radiat. Environ. Biophys. **58** (2019) 247.

25. A.M. Ali, H. Abdulkareem, M. Al Anazi *et al*., Biomed Res. Int. **2016** (2016) 1.