

COMPUTATIONAL PROTEIN DESIGN IN GREEN CHEMISTRY

A. A. Parikesit¹ and U. S. F. Tambunan^{2,*}

¹Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, 13210, East Jakarta, Indonesia

²Bioinformatics Research Group, Department of Chemistry, Faculty of Mathematics and Sciences, University of Indonesia, 16424, Depok, Indonesia

*E-mail : usman@ui.ac.id

ABSTRACT

The application of computational design of protein is mainly seen in the mutations of its active sites for increasing the reliability of that particular protein. Two essential examples from our groups are the stability improvement of riboflavin synthase and lipase enzyme. Specific mutations were introduced to their active and binding sites. Thus, by running molecular dynamics simulations, it is confirmed that the mutations improve the protein stability significantly. However, as the future of green chemistry is eventually aligned with nanotechnology, the atomic and even the sub-atomic features of the protein improvement will be put into the discussion. Nanotechnology is nowadays could not separate by the computational procedures. Besides molecular dynamics, finding the binding and active sites, as well as their chemical interactions, are important aspects of the nano-based computational design. In this respect, it could be related that the bioinformatics tools in Green Chemistry should be aligned with nanotechnology to produce fine-grained protein design that environmentally friendly.

Keywords: computational design, protein, bioinformatics, industrial reactor, molecular dynamics, mutation, nanotechnology.

© RASAYAN. All rights reserved

INTRODUCTION

Protein is the modular biomolecules that could act as the independent processing unit.¹ The utilization of protein in our metabolic system is acting as an enzyme, the biocatalyst of a biochemical reaction.^{2,3} In this respect, as biocatalyst, the enzyme has multitudes of feature that could be applied for facilitating organism's metabolism.⁴ This feature is of great importance when discussing the production of biochemical substances for the interest of drug, food, and vaccine products.⁵

Fermentation is the method that is a great use for producing biochemical products.^{6,7} In an industrial context, fermentation is always utilized using large-scale bioreactor with at least 1000 liter capacity.⁸ Hence, there are some constraints in bioreactor parameters. Firstly, bioreactor could only operate at the certain value of pressure and temperature. Secondly, overutilization of bioreactor means wasting energy and water supply. These constraints should be taken into consideration when operating a full-scale bioindustry, as options with the parameters of the working environment are limited⁹. In this end, due to the constraints, certain optimization measures should be taken possibly until the molecular level.

As a measure to leverage the bioreactor versatility, the bacteria and protein themselves should be optimized. Optimizing bacteria could mean optimizing the enzyme. In classical biochemistry, optimizing the enzyme means to find optimum condition within the threshold of the certain value of pH, temperature, and pressure¹⁰. However, in order to gain more flexibility toward optimization pipeline, the more sophisticated molecular approach should devise as the most optimum value is difficult to reach within the boundary of conventional methods.¹¹ Using conventional methods also have their drawbacks, namely energy and water wasting. Then, the molecular mechanism of the particular enzyme should be analyzed in a fine-grained

manner, especially with computer modeling.^{12,13} Optimization of enzyme molecular mechanism would usually be involved mutations on the active and binding site of the particular cavity or crevice.¹⁴ Usually, the mutations were introduced with site-directed mutagenesis PCR or with radioactive bombardment toward the bacteria or fungus.^{15,16} However, without the assistance and guide of computational measures, the experiment could produce random results in some cases and induce unsatisfactory production in the bioreactors.

The most common method for improving the performance of enzyme with computational measures is with improving the protein stability.¹⁷ The pipeline sequence for this methods is started with sequence alignment, structural alignment, search for an active and binding site, inducing mutation with molecular visualization, and lastly, molecular dynamics simulation.¹⁸⁻²⁰ The protein stability computation has eventually widely applied for various enzymatic performance improvement that eventually could save energy and water.²¹ Thus, more details on the impact of computational methods on the bench-side laboratory in protein stability improvement will serve as the objective for this review.

Application of Bioinformatics in Green Chemistry

The science for computational measures of protein design is Bioinformatics.²² As an independent research group and study program, Bioinformatics has provided guidelines to leverage the bench-side research to be smarter and more efficient.^{23,24} Thus, bioinformatics also the frontier research and studies for overcoming constrain in bioreactor capacity and environmental hazards.^{25,26} In this respect, the technical specification of the pipelines should be supported with proper methods in accordance to its respective order, such as CLUSTAL for sequence alignment, VAST for structural alignment, PROSITE for the search for active site, visualization and mutation with Pymol or Deepview, and GROMACS/NAMD for molecular dynamics.²⁷⁻³² Herewith, the applications of bioinformatics into the realm of green chemistry will be shown here.

Food Supplements Production

Riboflavin or vitamin B-2 is an important co-factor for supporting the metabolic system of the organism, especially human.³³⁻³⁵ However, producing riboflavin as a food supplement in large scale remain a big issue, as the default condition of the bioreactor remains too harsh for the biosynthesis of that particular compound.^{36,37} In this end, some groups have already tried different approaches to remedy this situation, mainly by introducing mutation to the proteins that responsible for the synthesis of riboflavin.³⁸ These approaches were introduced to many different micro-organisms that commonly utilized in the bioreactor.^{36,39,40} However, an important breakthrough was achieved when the riboflavin synthase of *Eremothecium gossypii* successfully optimized with bioinformatics methods to improve the thermal stability of the enzyme. In this respect, the molecular dynamics simulation at 300K, 315K, 325K, 335K, and 350K can determine the substrate-interacted residues of Ser152, Thr154, Ser167, and Thr172 at the C-terminal domain and Glu22, Cys54, Thr56, Ala70, Thr73, and His108 at the N-terminal. His108 was elucidated as the important active site from N-terminal that could be optimized with a directed mutagenesis method.⁴¹

Moreover, the utilized computational parameters for protein mutation was mainly centered on root mean square fluctuation (RMSF), average solvent accessible surface area (SASA) and Rg parameter.^{42,43} RMSF is the value of the fluctuation of proteins α carbon that is comparable to the scattering data of X-ray B-factor.^{44,45} The RMSF value should be elucidated optimally as it is important for protein not to be too rigid, as this would deter the occurrence of a biochemical reaction, and not be too flexible, as this would, in the end, denatured the protein. SASA is the total, hydrophobic, and hydrophilic accessible surfaces of protein.⁴⁶ Lastly, Rg parameter is a parameter that describes the equilibrium conformation of the total system of protein structure compactness.⁴⁷ These three parameters are equally important in the protein stability computation, especially for the enzymes that commonly utilized in the industrial applications. *Candida Antartica* Lipase B (CALB) is one most common example in that regards.

CALB is hydrolyzing enzyme that works on the ester bonds of carboxyl esters as a means to degrade lipid polymers.^{48,49} It is already applied in a wide range of fields, such as agriculture, food, environment, and others. The main problem that revealed in CALB optimization is the difficulties to reach an optimum

condition in the harsh temperature of the bioreactor. In our fine-grained examples, the utilization of SASA and Rg parameters to obtain the most optimum protein conformation of CALB will be exposed accordingly. In Fig.-1, within the frame of the molecular dynamics method for CALB enzyme, it is seen that both SASA and Rg parameters would rise up when the temperature is increased. This is unsurprisingly true as the kinetics of the atoms will be more dynamics as the temperature increases. Both SASA and Rg parameters are directly affected by the dynamics of the temperature.

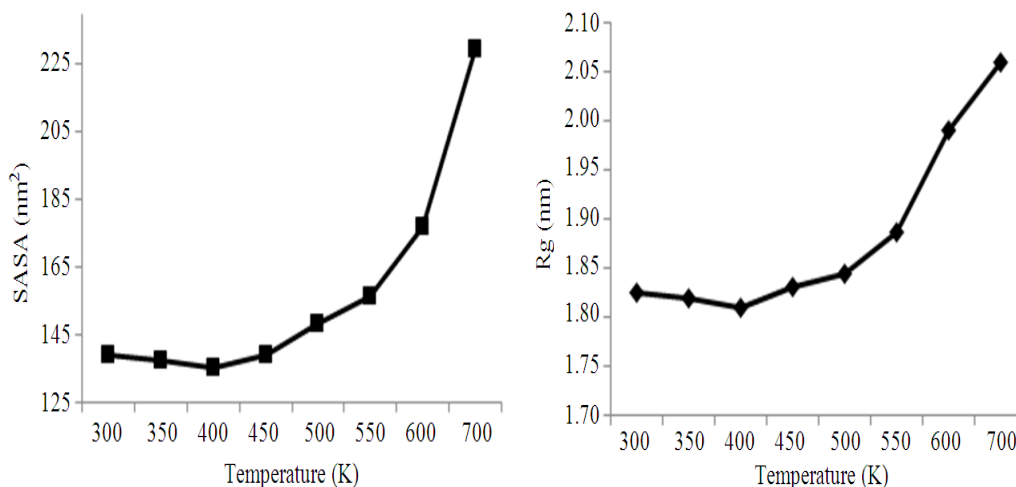


Fig.-1 : Average solvent accessible surface area. (Up) Average Solvent Accessible Surface Area (SASA) of final 2000 ps simulation of wild-type CALB; and (Down) average Radius of gyration (Rg) of final 2000 ps simulation of wild-type CALB (Figure is taken from Tambunan et al 2014⁴⁸)

In Fig.-2, it is seen that there is a significant difference between the SASA value of Mutant-1 and wild type CALB, both in 450 K and 550 K. Temperature difference plays a trivial role in this case. This is surprising as temperature always play role in protein dynamics. The SASA value of the Mutant-1, which is the product of a site-directed mutagenesis experiment, is significantly lower than the wild-type one. This phenomenon also observed in the Rg value of Fig.-3. Direct correlation of SASA with Rg value is predictable as they both depend on the temperature of the system.

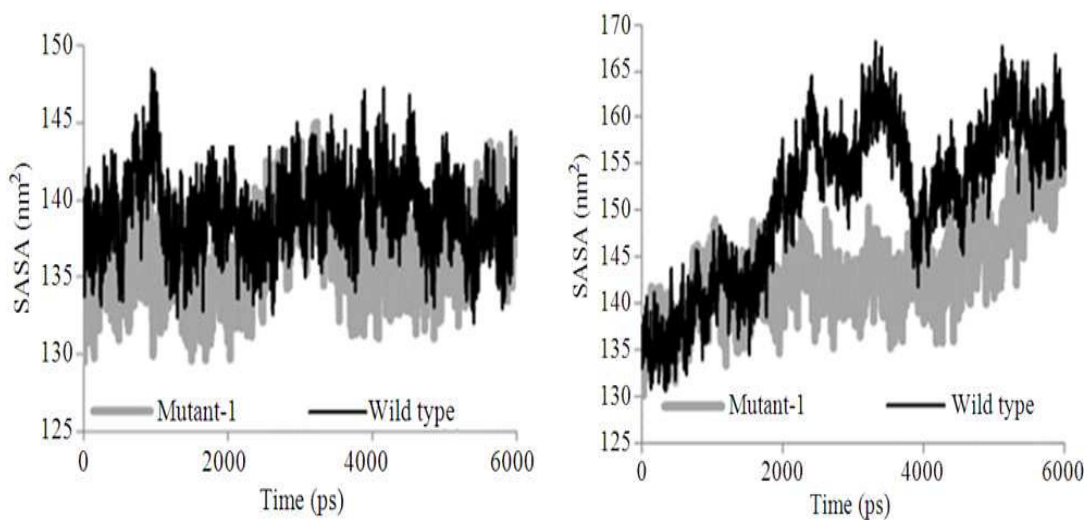


Fig.-2: SASA value of Mutant-1 and wild type CALB 450 K (up) and 550 K (down) simulation (Figure taken from Tambunan et al 2014⁴⁸)

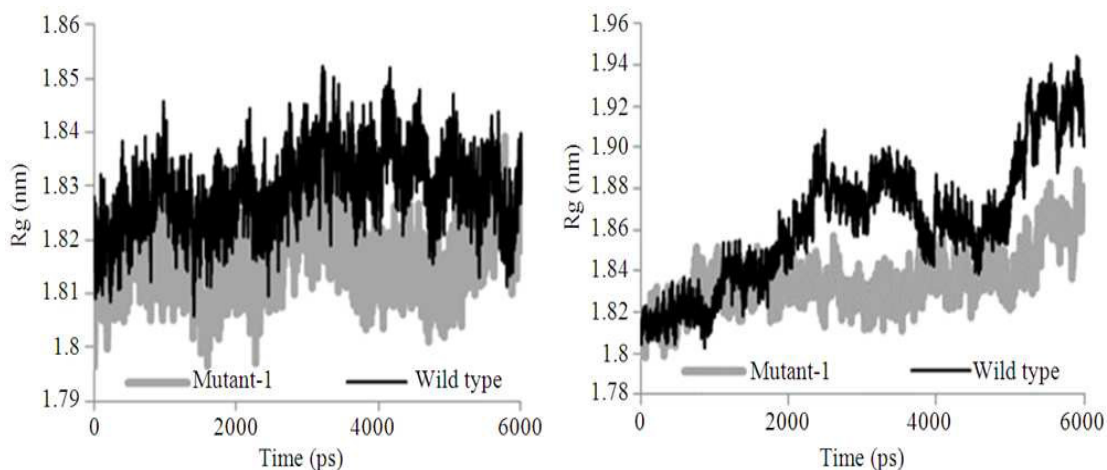


Fig.-3: Rg value of Mutant-1 and wild type CALB 450 K (up) and 550 K (down) simulation (Figure is taken from Tambunan et al 2014⁴⁸)

Cancer Drug Production

Nowadays, cancer drug production was devised by using modern biotechnology approach, namely with cutting-edge bioreactors. Large-scale mammalian cells bioreactors have been utilized for producing vaccines and drugs in bioindustry, and this pipeline is becoming commonplace.⁵⁰ Gendicine, a gene therapy agent for cancer, was produced in a large-scale culture of HEK293 cell with patented bioreactor⁵¹. This enhancement could be achieved because the titers of mammalian cell cultures could do away beyond the 5 g/l threshold, as it is optimized for the non-chromatographic separation process.^{52,8} Thus, plant culture cell already successful to be optimized for producing taxol using cell suspension culture technology⁵³. In this end, bioreactor technology has been successful in producing anti-cancer agents. In this end, producing natural products, gene, and protein-based products are doable and feasible within the frame of current technology. However, current drug and vaccine products for cancer are proved to be volatile^{54,55}. The biochemical substance such as silencing (si)RNA is very prone to the degradation process in the bioreactor. In this respect, the modification for the improvement should be conducted to the ligand as the product, not only the protein. This provides to be difficult as the mechanism to create a stable product is still not determined yet. Some possible scenarios are including sulfide bonds and incorporating pro-drugs into the lead. However, this scenario is not proven yet. The stability of products should be determined based on outputs of both in silico and wet laboratory experiments.⁵⁶⁻⁵⁸

CONCLUSION

Protein optimization in the bioreactor is necessary to obtain optimal condition to produce the necessary products such as food supplements and drugs. This optimization condition could be achieved with the assistance of solid bioinformatics tools. In this respect, it is concluded that the protein stability optimization could help to increase the production of the therapeutic agent. The particular optimization was reached by inducing mutations to the active and binding sites of the proteins.

ACKNOWLEDGMENT

The authors would like to thanks to the Directorate of Research and Community Engagement (DRPM), University of Indonesia, for their heartfelt support for this project.

REFERENCES

1. A. A. Parikesit, L. Steiner, P. F. Stadler, and S. J. Prohaska, *Malaysian J. Fundam. Appl. Sci.*, **10**, 65(2014), DOI: [10.11113/mjfas.v10n2.57](https://doi.org/10.11113/mjfas.v10n2.57)
2. M. K. Basu, L. Carmel, I. B. Rogozin, and E. V. Koonin, *Genome Res.*, **18**, 449(2008), DOI: [10.1101/251113](https://doi.org/10.1101/251113)

- 10.1101/gr.6943508
3. E. V Koonin, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **370**, 20140333 (2015), DOI: [10.1098/rstb.2014.0333](https://doi.org/10.1098/rstb.2014.0333)
 4. M. Lee and D. Kim, *BMC Bioinformatics*, **13 Suppl 1**, S6 (2012), DOI: [10.1186/1471-2105-13-s17-s6](https://doi.org/10.1186/1471-2105-13-s17-s6)
 5. T. Cajka, L. A. Garay, I. R. Sitepu, K. L. Boundy-Mills, and O. Fiehn, *J. Nat. Prod.*, **79**, 2580(2016), DOI: [10.1021/acs.jnatprod.6b00497](https://doi.org/10.1021/acs.jnatprod.6b00497)
 6. S. M. Kotay and D. Das, *Bioresour. Technol.*, **98**, 1183(2007), DOI: [10.1016/j.biortech.2006.05.009](https://doi.org/10.1016/j.biortech.2006.05.009)
 7. C. Dinamarca and R. Bakke, *Int. J. energy Environ.*, **2**, 1067(2011) .
 8. A. A. Shukla and J. Thömmes, *Trends Biotechnol.*, **28**, 253(2010), DOI: [10.1016/j.tibtech.2010.02.001](https://doi.org/10.1016/j.tibtech.2010.02.001)
 9. S. Pawiroharsono, Perspective of bioindustry in Indonesia, (2000).
 10. F. S. Hanschen et al., *Sci. Rep.*, **7**, 40807 (2017), DOI: [10.1038/srep40807](https://doi.org/10.1038/srep40807)
 11. S. C. L. Kamerlin and A. Warshel, *Proteins*, **78**, 1339(2010), DOI: [10.1002/prot.22654](https://doi.org/10.1002/prot.22654)
 12. J. Lameira, C. N. Alves, I. Tuñón, S. Martí, and V. Moliner, *J. Phys. Chem. B*, **115**, 6764(2011), DOI: [10.1021/jp202079e](https://doi.org/10.1021/jp202079e)
 13. R. Sattar, S. A. Ali, M. Kamal, A. A. Khan, and A. Abbasi, *Biochem. Biophys. Res. Commun.*, **314**, 755(2004), DOI: [10.1016/j.bbrc.2003.12.177](https://doi.org/10.1016/j.bbrc.2003.12.177)
 14. B. Steipe, B. Schiller, A. Plückthun, and S. Steinbacher, *J. Mol. Biol.*, **240**, 188(1994), DOI: [10.1006/jmbi.1994.1434](https://doi.org/10.1006/jmbi.1994.1434)
 15. K. Yamamoto et al., *Biochim. Biophys. Acta*, **1820**, 1469(2012), DOI: [10.1016/j.bbagen.2012.04.022](https://doi.org/10.1016/j.bbagen.2012.04.022)
 16. R. Bellini et al., *J. Med. Entomol.*, **50**, 94(2013), DOI: [10.1603/ME11058](https://doi.org/10.1603/ME11058)
 17. M.-Y. Jeong, S. Kim, C.-W. Yun, Y.-J. Choi, and S.-G. Cho, *J. Biotechnol.*, **127**, 300(2007), DOI: [10.1016/j.jbiotec.2006.07.005](https://doi.org/10.1016/j.jbiotec.2006.07.005)
 18. H. Ogawa et al., *Comput. Biol. Chem.*, **32**, 102(2008), DOI: [10.1016/j.compbiolchem.2007.11.001](https://doi.org/10.1016/j.compbiolchem.2007.11.001)
 19. T. Huang et al., *PLoS One*, **5**, e10972 (2010). DOI: [10.1371/journal.pone.0010972](https://doi.org/10.1371/journal.pone.0010972)
 20. P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, and E. Lindahl, *J. Chem. Theory Comput.*, **6**, 459(2010). DOI: [10.1021/ct900549r](https://doi.org/10.1021/ct900549r)
 21. T. J. Magliery, *Curr. Opin. Struct. Biol.*, **33**, 161(2015), DOI: [10.1016/j.sbi.2015.09.002](https://doi.org/10.1016/j.sbi.2015.09.002)
 22. W. Iks, *Significan Gice*, **1**, 7(2004), DOI: [10.1111/j.1740-9713.2004.00001.x](https://doi.org/10.1111/j.1740-9713.2004.00001.x)
 23. L. R. Welch, R. Schwartz, and F. Lewitter, P. E. Bourne, Editor. *PLoS Comput. Biol.*, **8**, e1002570 (2012), DOI: [10.1371/journal.pcbi.1002570](https://doi.org/10.1371/journal.pcbi.1002570)
 24. L. Welch et al., *PLoS Comput. Biol.*, **12**, e1004943 (2016), DOI: [10.1371/journal.pcbi.1004943](https://doi.org/10.1371/journal.pcbi.1004943)
 25. D. Sheyn, G. Pelled, D. Netanel, E. Domany, and D. Gazit, *Tissue Eng. Part A*, **16**, 3403(2010), DOI: [10.1089/ten.tea.2009.0834](https://doi.org/10.1089/ten.tea.2009.0834)
 26. R. Almstrand, A. J. Pinto, L. A. Figueroa, and J. O. Sharp, *Genome Announc.*, **4**, e01540-15 (2016), DOI: [10.1128/genomeA.01540-15](https://doi.org/10.1128/genomeA.01540-15)
 27. J. D. Thompson, T. J. Gibson, and D. G. Higgins, *Curr. Protoc. Bioinformatics*, **Chapter 2**, Unit 2.3 (2002), DOI: [10.1002/0471250953.bi0203s00](https://doi.org/10.1002/0471250953.bi0203s00)
 28. C. J. A. Sigrist et al., *Nucleic Acids Res.*, **38**, D161-6 (2010), DOI: [10.1093/nar/gkp885](https://doi.org/10.1093/nar/gkp885)
 29. T. Makarewicz and R. Kaźmierkiewicz, *J. Chem. Inf. Model.*, **53**, 1229(2013), DOI: [10.1021/ci400071x](https://doi.org/10.1021/ci400071x)
 30. T. D. Team, DeepView – The Swiss-PdbViewer User Guide, World Trade (2001).
 31. W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graph.*, **14**, 33, 27(1996), DOI: [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
 32. Y. Wang et al., *Trends Biochem. Sci.*, **25**, 300(2000), DOI: [10.1016/S0968-0004\(00\)01561-9](https://doi.org/10.1016/S0968-0004(00)01561-9)
 33. V. Massey, *Biochem. Soc. Trans.*, **28**, 283(2000), DOI: [10.1042/bst0280283](https://doi.org/10.1042/bst0280283)
 34. H. J. Powers, *Am. J. Clin. Nutr.*, **77**, 1352(2003), DOI: [10.1093/ajcn/77.6.1352](https://doi.org/10.1093/ajcn/77.6.1352)
 35. R. P. Goodrich, *Vox Sang.*, **78 Suppl 2**, 211(2000).
 36. S. H. Lim, J. S. Choi, and E. Y. Park, *Biotechnol. Bioprocess Eng.*, **6**, 75(2001), DOI: [10.1007/BF02931951](https://doi.org/10.1007/BF02931951)
 37. E. Choe, R. Huang, and D. B. Min, *J. Food Sci.*, **70**, R28(2005), DOI: [10.1111/j.1365-2621.2005.tb09055.x](https://doi.org/10.1111/j.1365-2621.2005.tb09055.x)

38. W. Sybesma, C. Burgess, M. Starrenburg, D. van Sinderen, and J. Hugenholtz, *Metab. Eng.*, **6**, 109 (2004), DOI: [10.1016/j.ymben.2003.11.002](https://doi.org/10.1016/j.ymben.2003.11.002)
39. T. Kato and E. Y. Park, *Biotechnol. Lett.*, **34**, 611(2012), DOI: [10.1007/s10529-011-0833-z](https://doi.org/10.1007/s10529-011-0833-z)
40. A. Bacher, S. Eberhardt, M. Fischer, K. Kis, and G. Richter, *Annu. Rev. Nutr.*, **20**, 153(2000), DOI: [10.1146/annurev.nutr.20.1.153](https://doi.org/10.1146/annurev.nutr.20.1.153)
41. S. Idrus and U. S. F. Tambunan, *Online J. Bioinforma.*, **13**, 41(2012).
42. E. Durham, B. Dorr, N. Woetzel, R. Staritzbichler, and J. Meiler, *J. Mol. Model.*, **15**, 1093(2009), DOI: [10.1007/s00894-009-0454-9](https://doi.org/10.1007/s00894-009-0454-9)
43. P. Ferrara, J. Apostolakis, and A. Caflisch, *Proteins Struct. Funct. Genet.*, **46**, 24(2002), DOI: [10.1002/prot.10001](https://doi.org/10.1002/prot.10001)
44. D. Reichert, T. Zinkevich, K. Saalwächter, and A. Krushelnitsky, *J. Biomol. Struct. Dyn.*, **30**, 617 (2012), DOI: [10.1080/07391102.2012.689695](https://doi.org/10.1080/07391102.2012.689695)
45. E. Fuglebakk, J. Echave, and N. Reuter, *Bioinformatics*, **28**, 2431(2012), DOI: [10.1093/bioinformatics/bts445](https://doi.org/10.1093/bioinformatics/bts445)
46. L. Lins, A. Thomas, and R. Bresseur, *Protein Sci.*, **12**, 1406(2003), DOI: [10.1110/ps.0304803](https://doi.org/10.1110/ps.0304803)
47. M. Y. Lobanov, N. S. Bogatyreva, and O. V. Galzitskaya, *Mol. Biol.*, **42**, 623(2008), DOI: [10.1134/S0026893308040195](https://doi.org/10.1134/S0026893308040195)
48. U. S. F. Tambunan, A. Randy, and A. A. Parikesit, *Online J. Biol. Sci.*, **14**, 108(2014), DOI: [10.3844/ojbsci.2014.108.118](https://doi.org/10.3844/ojbsci.2014.108.118)
49. M. R. Ganjalikhany, B. Ranjbar, A. H. Taghavi, and T. Tohidi Moghadam, *PLoS One*, **7** (2012), DOI: [10.1371/journal.pone.0040327](https://doi.org/10.1371/journal.pone.0040327)
50. L. Chu and D. K. Robinson, *Curr. Opin. Biotechnol.*, **12**, 180(2001), DOI: [10.1016/S0958-1669\(00\)00197-X](https://doi.org/10.1016/S0958-1669(00)00197-X)
51. Z. Peng, *Hum. Gene Ther.*, **16**, 1016(2005). DOI: [10.1089/hum.2005.16.1016](https://doi.org/10.1089/hum.2005.16.1016)
52. A. Grover, A. Shandilya, V. S. Bisaria, and D. Sundar, *BMC Genomics*, **11 Suppl 4**, S15 (2010), DOI: [10.1186/1471-2164-11-S4-S15](https://doi.org/10.1186/1471-2164-11-S4-S15)
53. S. Malik et al., *Process Biochem.*, **46**, 23(2011), DOI: [10.1016/j.procbio.2010.09.004](https://doi.org/10.1016/j.procbio.2010.09.004)
54. A. A. Parikesit and D. Anurogo, *Chim. Nat. Acta*, **4**, 27 (2016), DOI: [10.24198/cna.v4.n1.10445](https://doi.org/10.24198/cna.v4.n1.10445)
55. A. A. Parikesit and R. Nurdiansyah, *IOP Conf. Ser. Mater. Sci. Eng.*, **299**, 12059 (2018), DOI: [10.1088/1757-899X/299/1/012059](https://doi.org/10.1088/1757-899X/299/1/012059)
56. B.V. Udugade, and S. P. Gawade, *Rasayan J. Chem*, **11(1)**, 118(2018), DOI: [10.7324/RJC.2018.1111860](https://doi.org/10.7324/RJC.2018.1111860)
57. S. Marupati, R. K. Gaddam, B. Prasanna, and A. Thirupathaiah, *Rasayan J. Chem*, **11(2)**, 661(2018), DOI: [10.31788/RJC.2018.1122062](https://doi.org/10.31788/RJC.2018.1122062)
58. S. Ganguly and S. Murugesan, *Rasayan J. Chem*, **1(2)**, 251(2008).

[RJC-3038/2018]