UGWU PASCHAL
+2348166207095 | ugwupaschal@gmail.com
https://www.linkedin.com/in/paschal-ugwu-52abb6229/
Helix Biogen Institute Immunoinformatics Final Project
Topic: Design of a hypothetical vaccine to target *BRCA1* gene mutations.
September, 2020

Abstract

BRCA1 mutations have been associated with the risk of developing breast and ovarian cancer. The aim of this project was to design a novel vaccine construct against *BRCA1* gene mutation by combination of highly immunogenic epitopes and suitable adjuvants and linkers. I retrieved the tumor proteome, checked the physical and chemical properties of the tumor proteins, predicted B-cells, cytotoxic T-lymphocytes, helper T-lymphocytes, antigenicity, allergenicity, toxicity, and population coverage. I then proceeded to construct the vaccine (which I called *BRCA1* VACC), predicted its structure and docked it with TLR3 and TLR4. I believe that this project will be useful to researchers and pharmaceutical companies to work on and develop a potent vaccine that will be used to arrest breast cancer in the future.

Introduction

Germline mutations in *BRCA1* and *BRCA2* increase the risk for developing breast and ovarian cancer (Sluiter and van Rensburg, 2011). *BRCA1* gene mutations account for about 25-28% of hereditary breast cancer as *BRCA1* is included in the category of high penetrance genes. Except for few common mutations, there is a heterogenous spectrum of *BRCA1* mutations in various ethnic groups. 185AGdel and 5382ins are the most common *BRCA1* alterations (founder mutations) which have been identified in most of the population (Sharma *et al.,* 2018). *BRCA*1 is a tumor suppressor gene which, when mutated, is associated with the development of hereditary breast cancers. In sporadic tumors, although inherent gene mutations are rare, loss of *BRCA1*, resulting from reduced expression or incorrect subcellular localization, is postulated to be important (Rakha *et al.,* 2008). The *BRCA1* gene was identified and cloned in 1994 based its linkage to early onset breast cancer and breast-ovarian cancer syndromes in women. While inherited mutations of *BRCA1* are responsible for about 40–45% of hereditary breast cancers, these mutations account for only 2-3% of all breast cancers, since the *BRCA1* gene is rarely mutated in sporadic breast cancers. However, *BRCA1* expression is frequently reduced or absent in sporadic cancers,

suggesting a much wider role in mammary carcinogenesis (Rosen *et al.,* 2003). Breast cancer is the most frequently diagnosed neoplastic disease in women around menopause often leading to a significant reduction of these women's ability to function normally in everyday life. The increased breast cancer incidence observed in epidemiological studies in a group of women actively participating in social and professional life implicates the necessity of conducting multidirectional studies in order to identify risk factors associated with the occurrence of this type of neoplasm (Kamińska *et al.,* 2015). The most recent global cancer burden figures estimate that there were 2.26 million incident breast cancer cases in 2020 and the disease is the leading cause of cancer mortality in women worldwide. The incidence is strongly correlated with human development, with a large rise in cases anticipated in regions of the world that are currently undergoing economic transformation. Survival, however, is far less favourable in less developed regions (Wilkinson and Gathani, 2022). Over the past few years, substantial advances have been made in the discovery of new drugs for treating Brest Cancer. Improved understanding of the biologic heterogeneity of Breast Cancer has allowed the development of more effective and individualized approach to treatment (Tong *et al.,* 2018). After almost two decades of poor clinical trial results, cancer vaccines (CVs), an active immunotherapy, have come back in the spotlight because of some technological advancements, ultimately boosted by coronavirus disease 2019 pandemic. In particular, neoantigens are emerging as the preferred targets for CVs, with gene-based and viral vector–based platforms in development. Moreover, lipid nanoparticles proved to be immunogenic and efficient delivery vehicles (Corti *et al.,* 2022). Vaccine therapy introduces antigens that act on cancer cells causing prolonged activation of the immune system. In particular, cancer relapse could be avoided due to the presence of a longer period of immunological memory with an effective vaccine that can protect against various tumor antigens. Cancer vaccines are broadly classified as preventive and therapeutic. Preventive vaccines are used to ward off any future infections and therapeutic vaccines are used to treat a person with active disease (Pallerla *et al.,* 2021).

Aim of project

This study aimed to design a novel vaccine construct against *BRCA1* gene mutation by combination of highly immunogenic epitopes with suitable adjuvants and linkers.

Methods:

The retrieval of the tumor proteome.

The entire proteome of *BRCA1* (Homo sapiens) with accession number:AAC37594.1, was retrieved from National Center for Biotechnology Information (NCBI at https://www.ncbi.nlm.nih.gov/protein/AAC37594.1?report=fasta).

Sequence alignment and determination of the conserved regions

The retrieved protein sequences of *BRCA1* proteins were further aligned to obtain conserved epitopes using multiple sequence alignment (MSA) tools, muscle, embedded in the MEGA 11 software. MSA analysis was performed to analyze 100% conserved epitopes amongst the screened sequences of *BRCA1* protein.

Physical and chemical properties of the tumor proteins and antigenicity.

ProtParam (http://web.expasy.org/protparam/) is a tool allowed the computation of various physical and chemical parameters for a given protein sequence. Each protein was subjected to Protparam server for the physiochemical properties and the computed parameters covered the molecular weight, theoretical pI, amino acid composition, extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Moreover the VaxiJen v2.0 server at (http://www.ddg-pharmfac.net/vaxijen/) which based on autoand cross-covariance transformation of protein sequences into uniform vectors of principal amino acid properties was used to analyze the potent antigenicity of each protein of *BRCA1*.

B-cell epitopes prediction

B-cell epitopes are antigenic determinants recognized by the immune system and represent the specific piece of the antigen to which B lymphocytes bind. These play a vital role in vaccine design. The Immune Epitope Data Base web server (IEDB) at (https://www.iedb.org/) was used for prediction of B cell epitopes. A collection of methods to predict B cell epitopes based on sequence characteristics of the antigen using amino acid scales and hidden Markov Models (HMMs) were used. For instance; Linear Bcell epitopes were predicted using BepiPred linear epitopes prediction tool (Larsen *et al.,* 2006). Emini Surface Accessibility prediction method was used to obtain surface epitopes (Emini *et al.,* 1985). The antigenicity of the predicted epitopes was performed using Kolaskar and Tongaonkar Antigenicity prediction tools (Kolaskar and Tongaonkar, 1990). For prediction of epitopes flexibility and hydrophilicity, the Karplus and Schulz flexibility and Parker hydrophilicity prediction tools were used (Parker *et al.,* 1986).

Cytotoxic T lymphocytes epitopes prediction

The epitopes binding analysis to Major Histocompatibility Complex class I molecules (MHC class I) from *BRCA1* protein was performed using IEDB MHC I tools at (http://tools.iedb.org/mhci/). The MHC I epitope molecules that interacted to T lymphocytes was subjected to multiple steps. This prediction method used an amino acid sequence, or set of sequences and determined each subsequence's ability to bind to a specific MHC class I molecule. The binding of the fragmented peptides to MHC molecules step was predicted by Artificial Neural Network 4.0 (ANN 4.0) method. Prior to the prediction, all lengths of epitope was set as 9mers and all the conserved epitopes that bound to alleles at score of ≤100 half-maximal inhibitory concentration (IC50) were subjected for further analysis (Kim *et al.,* 2012).

Helper T-lymphocytes epitopes prediction

Analysis of peptides binding to MHC II molecules from *BRCA1* protein was assessed by the IEDB MHC II prediction tool at (http://tools.iedb.org/mhcii/result/). For MHC II binding predication, human allele's reference sets (human HLA-DR, HLA-DQ, HLA-DP) were used. MHC II groove has the ability to bind different lengths peptides that makes prediction more difficult and less accurate. Thus Neural Network align (NN-align 2.3; Net MHCII 2.3) was used to identify both the binding affinity and MHCII binding core epitopes. Prior to the prediction, the length of peptide was set as 15mers (15 amino acids) and all the conserved epitopes that bound to alleles at score of score of ≤1000 half-maximal inhibitory concentration (IC50) were subjected for further analysis (Wang *et al.,* 2008).

Antigenicity, allergenicity and toxicity of the predicted epitopes

Analysis of the antigenicity, allergenicity and toxicity of the predicted epitopes from *BRCA1* protein for B and T lymphocytes, was performed using multiple prediction tools. The predicted epitopes were submitted to the VaxiJen v2.0 server for antigenicity prediction. The threshold of VaxiJen v2.0 server was set to the default threshold (0.5). Epitopes that demonstrated antigenicity were further investigated for their allergenicity using AllerTOP server (Dimitrov *et al.,* 2014). Epitopes found to be antigenic and non-allergenic were further assessed for toxicity by ToxinPred server (Gupta *et al.,* 2013).

Population coverage

Epitopes that interacted with MHC I and MHC II from *BRCA1* protein were subjected to population coverage analysis after they were shown to be antigenic, nonallergic and

nontoxic. The population coverage was investigated against the whole world using IEDB population coverage tool at (http://tools.iedb.org/ tools/population/iedb_input).

Vaccine construction (multiepitopes vaccine)

Epitopes that passed the criteria of B cell epitopes prediction, epitopes with high allelic interaction and best population coverage scores against cytotoxic and helper T lymphocytes were used to generate the vaccine construct. Epitopes that overlapped in both MHC I and MHC II were used once in the vaccine construct as MHC I or MHC II epitopes. The vaccine construct was generated as previously described (Hasan *et al.,* 2019) with minor modifications. The GPGPG linker was used to fuse the B cell and T helper predicted epitopes. While AAY linker was used to link the epitopes of T cytotoxic lymphocytes. EAAAK linker was used to link the epitopes with the Human TLR2 protein (accession number AAH33756.1) that was used as an adjuvant on the amino and carboxyl terminals to ameliorate the immunogenicity of the vaccine construct (Tani *et al.,* 2000). Linkers were shown to assist in enhancing expression, stability and folding of the protein by separating the functional domains (Shamriz *et al.,* 2016).

Physical and chemical properties of the vaccine construct

The vaccine construct from predicted epitopes was analyzed for the physical and chemical properties using Protparam analysis tool. The computed parameters covered the molecular weight, theoretical pI, amino acid composition, extinction coefficient, estimated half-life, instability index, aliphatic index and Grand average of hydropathicity (GRAVY).

Secondary structure prediction

Self-optimized prediction method (SOPMA) at (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) was used to predict alpha helix, coiled structures and beta sheets in the secondary structure of the vaccine construct (Combet *et al.,* 2000).

Molecular docking of the vaccine construct with TLR3 and TLR4 (protein-ligand docking)

Protein–Ligand interactions play a fundamental role in a variety of biological processes. Determining the complex structures of these interactions is valuable, in which molecular docking has played an important role. PatchDock web server was used for protein-ligand docking, that is to dock the vaccine construct with human Toll-Like Receptor4 (TLR4) and also human Toll-Like Receptor3 (TLR3). The

vaccine construct PDB file was submitted to the server with TR3 and TLR4 as receptors for the docking process. Visualization to determine the interactions and bonding that exist between human Toll-Like Receptors was carried out using UCSF Chimera software.

Codon adaptation and in silico cloning.

Codon adaptation and in silico cloning were performed in order to express the final vaccine construct in the E. coli (strain K12) host since codon usage optimization demonstrated differences between human and E. coli strain. The purpose of this cloning was to guarantee the expression of the vaccine construct in the selected host. Java Codon Adaptation Tool (JCAT) server (http://www.prodoric.de/JCat) was firstly used for the reverse translation of the protein sequence of the vaccine construct into DNA sequence. The rho independent transcription termination, prokaryote ribosome binding site and cleavage site of restriction enzyme were avoided (Shintouo *et al.,* 2020). In the JACT, codon adaptation index (CAI) score is 1.0 but > 0.8 is considered a good score (Morla *et al.,* 2016). The favourable GC content of a sequence ranged between 30 and 70%. Secondly, XbaI and XhIO restriction enzymes cutting sites sequences were introduced to the DNA sequence obtained by (JCat) server at the N-terminal and C-terminal vicinities, respectively. The SnapGene restriction cloning Module (Shintouo *et al.,* 2020) was used to insert the DNA sequence into pET28a (+) vector between the XbaI and XhIO.

Conclusion

I have been able to hypothetically design a vaccine (which I called *BRCA1* VACC) that will target *BRCA1* gene mutations. This vaccine has been tested using different parameters, and has proven to be save for the body. Molecular docking that was also performed proved excellent interactions between the body and the vaccine. I believe that this project will be useful to researchers and pharmaceutical companies to work on and develop a potent vaccine that will be used to arrest breast cancer in the future.

References

Combet, C., Blanchet, C., Geourjon, C., and Deléage, G. (2000). NPS@: network protein sequence analysis. *Trends in biochemical sciences*, *25*(3), 147–150. https://doi.org/10.1016/s0968-0004(99)01540-6

Corti, C., Giachetti, P. P., Eggermont, A. M., Delaloge, S., and Curigliano, G. (2022). Therapeutic vaccines for breast cancer: Has the time finally come?. *European Journal of Cancer*, *160*, 150-174.

Dimitrov, I., Bangov, I., Flower, D. R., and Doytchinova, I. (2014). AllerTOP v.2--a server for in silico prediction of allergens. *Journal of molecular modeling*, *20*(6), 2278. https://doi.org/10.1007/s00894-014-2278-5

Emini, E. A., Hughes, J. V., Perlow, D. S., and Boger, J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of virology*, *55*(3), 836–839. https://doi.org/10.1128/JVI.55.3.836-839.1985

Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery Consortium, et al. (2013) *In Silico* Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE* 8(9): e73957. https://doi.org/10.1371/journal.pone.0073957

Hasan, M., Ghosh, P. P., Azim, K. F., Mukta, S., Abir, R. A., Nahar, J., and Hasan Khan, M. M. (2019). Reverse vaccinology approach to design a novel multi-epitope subunit vaccine against avian influenza A (H7N9) virus. *Microbial pathogenesis*, *130*, 19–37. https://doi.org/10.1016/j.micpath.2019.02.023

Kamińska, M., Ciszewski, T., Łopacka-Szatan, K., Miotła, P., and Starosławska, E. (2015). Breast cancer risk factors. *Menopause Review/Przegląd Menopauzalny*, *14*(3), 196-202.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., and Peters, B. (2012). Immune epitope database analysis resource. *Nucleic acids research*, *40*(Web Server issue), W525–W530. https://doi.org/10.1093/nar/gks438

Kolaskar, A. S., and Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS letters*, *276*(1-2), 172–174. https://doi.org/10.1016/0014-5793(90)80535-q

Larsen, J. E., Lund, O., and Nielsen, M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome research*, *2*, 2. https://doi.org/10.1186/1745-7580-2-2

Morla, S., Makhija, A., and Kumar, S. (2016). Synonymous codon usage pattern in glycoprotein gene of rabies virus. *Gene*, *584*(1), 1–6. https://doi.org/10.1016/j.gene.2016.02.047

Pallerla, S., Abdul, A. U. R. M., Comeau, J., and Jois, S. (2021). Cancer vaccines, treatment of the future: With emphasis on her2-positive breast cancer. *International journal of molecular sciences*, *22*(2), 779.

Parker, J. M., Guo, D., and Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. *Biochemistry*, *25*(19), 5425–5432. https://doi.org/10.1021/bi00367a013

Rakha, E. A., El-Sheikh, S. E., Kandil, M. A., El-Sayed, M. E., Green, A. R., and Ellis, I. O. (2008). Expression of *BRCA1* protein in breast cancer and its prognostic significance. *Human pathology*, *39*(6), 857-865.

Rosen, E. M., Fan, S., Pestell, R. G., and Goldberg, I. D. (2003). *BRCA1* gene in breast cancer. *Journal of cellular physiology*, *196*(1), 19-41.

Shamriz, S., Ofoghi, H., and Moazami, N. (2016). Effect of linker length and residues on the structure and stability of a fusion protein with malaria vaccine application. *Computers in biology and medicine*, *76*, 24–29. https://doi.org/10.1016/j.compbiomed.2016.06.015

Sharma, B., Kaur, R. P., Raut, S., and Munshi, A. (2018). *BRCA1* mutation spectrum, functions, and therapeutic strategies: the story so far. *Current problems in cancer*, *42*(2), 189-207.

Shintouo, C. M., Shey, R. A., Nebangwa, D. N., K Esoh, K., Nongley, N. F., Nguve, J. E., Giron, P., Mutesa, L., Vanhamme, L., Souopgui, J., Ghogomu, S. M., and Njemini, R. (2020). In Silico Design and Validation of OvMANE1, a Chimeric Antigen for Human Onchocerciasis Diagnosis. *Pathogens (Basel, Switzerland)*, *9*(6), 495. https://doi.org/10.3390/pathogens9060495

Sluiter, M. D., and van Rensburg, E. J. (2011). Large genomic rearrangements of the *BRCA1* and *BRCA2* genes: review of the literature and report of a novel *BRCA1* mutation. *Breast cancer research and treatment*, *125*(2), 325-349.

Tani, K., Murphy, W. J., Chertov, O., Salcedo, R., Koh, C. Y., Utsunomiya, I., Funakoshi, S., Asai, O., Herrmann, S. H., Wang, J. M., Kwak, L. W., and Oppenheim, J. J. (2000). Defensins act as potent adjuvants that promote cellular and humoral immune responses in mice to a lymphoma idiotype and carrier antigens. *International immunology*, *12*(5), 691–700. https://doi.org/10.1093/intimm/12.5.691

Tong, C. W., Wu, M., Cho, W. C., and To, K. K. (2018). Recent advances in the treatment of breast cancer. *Frontiers in oncology*, *8*, 227.

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., and Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS computational biology*, *4*(4), e1000048. https://doi.org/10.1371/journal.pcbi.1000048

Wilkinson, L., and Gathani, T. (2022). Understanding breast cancer as a global health concern. *The British Journal of Radiology*, *95*(1130), 20211033.