In silico analysis of chimeric subunit vaccine containing HER-2-MUC1 against breast cancer

Article *in* Indian Journal of Biotechnology · April 2018

CITATIONS
0

READS
84

4 authors:

Mahdieh Mehrab Mohseni
National Institute of Genetic Engineering and Biotechnology
15 PUBLICATIONS **60** CITATIONS
SEE PROFILE

Jafar Amani
Baqiyatallah University of Medical Sciences, Tehran, Iran
211 PUBLICATIONS **1,258** CITATIONS
SEE PROFILE

Elaheh Gheybi
Ferdowsi University Of Mashhad
9 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Ali hatef Salmanian
National Institute of Genetic Engineering and Biotechnology
154 PUBLICATIONS **1,593** CITATIONS
SEE PROFILE

Indian Journal of Biotechnology
Vol. 17, April 2018, pp 224-233

In silico analysis of chimeric subunit vaccine containing HER-2-MUC1 against breast cancer

Mahdieh Mehrab Mohseni[1], Jafar Amani[2*], Elaheh Gheybi[3] and Ali Hatef Salmanian[1*]

[1]Agricultural Biotechnology, National Institute of Genetic Engineering and Biotechnology (NIGEB), Tehran, Iran

[2]Applied Microbiology Research Center, Systems Biology and Poisonings Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran

[3]Department of Biology, Faculty of Science, University of Guilan, Rasht, Iran

Received 17 June 2015; revised 14 June 2016; accepted 11 July 2016

Breast cancer is a leading cause of cancer-related deaths in women worldwide. Although tumorectomy, radiotherapy, chemotherapy and hormone replacement therapy have been used for the treatment of breast cancer, there is no effective therapy for patients with invasive and metastatic breast cancer. Targeting tumors using cancer vaccine therapeutics has several advantages including the induction of long-term immunity, prime boost strategies for additional treatments and reduced side effects compared to conventional chemotherapeutics. However, one problem in targeting tumor antigens directly is that it can lead to antigen loss or immune editing. We have designed a complex immunogen derived from the extracellular domain of human HER-2/*neu*- (480–620) and seven tandem repeats of MUC1 (VNTR) that represents a three-dimensional epitope. The construct was analyzed by bioinformatics softwares. Linear and discontinuous B-cell epitopes, MHC class I and II binding peptides of chimeric protein were predicted. Results suggest that the construct can be an appropriate vaccine candidate against breast cancer.

Keywords: Breast cancer, chimeric vaccine, cancer antigens, *in silico* analysis

Introduction

Breast cancer is a leading cause of cancer related death in women all over the world. It is estimated that about 1 in 8 women in the United States of America (approximately 12.8 percent) will develop breast cancer during their lifetime. Each year, more than 180,000 new cases of invasive breast cancer are diagnosed, with more than 48,000 deaths attributed to this disease in the USA[1-2]. There are more than 1,000,000 new cases and 370,000 deaths worldwide annually[3]. Clearly, effective new treatments are needed to combat this disease. Although tumorectomy, radiotherapy, chemotherapy and hormone replacement therapy have been used for the treatment of breast cancer, there is no effective therapy for patients with invasive and metastatic breast cancer[4]. Breast cancer is a multifactorial disease caused by complex inherited and environmental factors. Human breast cancers (BC) are highly heterogeneous with respect to their biologic and clinical profiles[5]. The human epidermal growth factor receptor-2 (HER2) proto-oncogene, also known as HER2/neu and c-erbB-2, encodes a growth factor receptor that has been found to play an important role in breast cancer[6]. HER-2/neu is often over expressed in breast cancer, ovary, lung and gastrointestinal tract[7]. The HER2 is usually expressed at low levels on the cell surface of many tissues. But the receptors can increase 40-fold (over expression) in tumors, It can be increased nearly 50,000 to over 2,000,000 HER2 per cell[8]. Despite its high oncogenic potential, HER-2/*neu* oncoprotein has been considered as a promising target for cancer therapeutics as it has been proven to be immunogenic *in vivo* inducing specific cytotoxic T-Lymphocytes (CTL) and immunoglobulin G (IgG) responses in a considerable subset of patients with HER-2/*neu* positive breast cancer[9]. Moreover, various MHC Class I and Class II binding peptides from the HER-2/*neu* sequence were also able to induce tumor reactive CTL and T helper (Th) responses *in vitro*, respectively. However, synthetic peptides corresponding to HER-2/*neu* immunogenic epitopes have greatly failed to function as therapeutic vaccines in phase I/II trials[10]. However, in parallel with the studies aiming at improving the effectiveness of a vaccine, it has been considered essential to identify novel immunogenic CTL epitopes, aiming at

*Author for correspondence:
Tel: +98-21-82482592, +98-21-44580365, Fax: +98-21-88068924, +98-21-44580395
jafar.amani@gmail.com; salman@higeb.ac.in

providing insight into the design of more efficacious peptide-based treatment modalities. Mucin 1 is a well-known tumor marker and cellular tumor antigen[11], It is one of the tumor antigen which is over expressed in 90% of adenocarcinomas including breast cancer[12-13]. MUC1 increase 100-fold in cancer cell in compare to normal cells therefore it is a good target for tumor immunotherapy[14]. MUC1 mucin is a large transmembrane glycoprotein in which an extracellular (MUC1-N) and a membrane spanning subunit (MUC1-C) are held together in a non-covalent association. The larger subunit, MUC1-N, primarily consists of the variable number tandem repeat (VNTR) domain copied up to 125 times per molecule. Because each VNTR contains five potential sites of *O*-linked glycosylation, MUC1 is normally decorated with a network of branched "core 2-based" glycans. As a result, the MUC1 ecto domain assumes a rigid structure and may extend up to 500 nm from the cell surface[15]. Antibodies against MUC1 are used for *in vivo* targeting of breast and ovarian tumors, and there is considerable interest in MUC1 as a possible target antigen for immunotherapy of breast cancer[16]. Subunit vaccines are considerably safer, more specific with less adverse reactions, with the ability to target the site where immunity is required, and large-scale production of recombinant proteins by biotechnological revolution. More complete protective response is generated by multi-component vaccine in compare with a single component[17]. Hence, more efficient immunoprotection could be achieved against cancer by multisubunit vaccine. The identification of vaccine subunits is often a lengthy process and bioinformatics approaches have recently been used to identify candidate protein vaccine antigens[18]. Comparative structural and immunological analysis of antigens could lead to the judicious selection of a combination of immunogens for a multi subunit chimeric vaccine[19].

In the present study, we designed a novel multi subunit antigen that provides a suitable and safe vaccine candidate against breast cancer. We employed bioinformatics softwares for designing a chimeric protein with high level expression in prokaryotic and eukaryotic systems.

Materials and Methods

Sequence Retrieval

Related sequences of major breast cancer antigens (HER-2 and MUC1) were obtained from UniProtKB/Swiss-Prot.

Segment Selection Based on Prediction of Immunogenic Epitopes

B-cell Epitopes

Linear B-cell epitopes of construct were estimated using BepiPred 1.0 Server (http://www.cbs.dtu.dk/services/BepiPred/) and bcepred (http://www.imtech.res.in/raghava /bcepred)[20,21]. Discotope 1.2 (http://tools.immuneepitope.org/stools/discotope/discotope) and SEPPA (http://lifecenter.sgst.cn/seppa/) were employed to predict discontinuous B-cell epitopes. Additionally, ElliPro server (http://tools.immuneepitope.org/tools/ElliPro/iedb_input) was used to predict both linear and discontinuous B-cell epitopes. This method predicts epitopes based upon solvent accessibility and flexibility[22-24].

T-cell Epitope Prediction

Proteasomal cleavage sites of the chimeric protein were predicted by Netchop 3.1 (http://www.cbs.dtu.dk/services/NetChop/), MAPPP (http://www.mpiib-berlin.mpg.de/MAPPP/binding.html) and PCPS (http://imed.med.ucm.es/Tools/pcps/index.html). Peptides binding affinity to TAP protein was computed by TAPPred (http://www.imtech.res.in/raghava/tappred)[25-27]. ProPred-I (http://www.imtech.res.in/raghava/propred/), nHLAPred (http://www.imtech.res.in/raghava/nhlapred/) and CTLpred (http://www.imtech.res.in/raghava/ctlpred/) were used for prediction of peptide sequences binding to MHC class I[28-30]. HLApred (http://www.imtech.res.in/raghava/hlapred/), MHC2Pred (http://www.imtech.res.in/raghava/mhc2pred/help.html) and Propred (http://www.imtech.res.in/raghava/propred/) were employed to predict peptides binding to MHC class II[28].

Designing the Chimeric Construct (HER-2- MUC1)

The chimeric antigenic construct HER-2-MUC1 (HM) was assembled by fusing two selected segments by a linker into a single molecule. The C-terminal (480-620) of HER-2 (UniProtKB/Swiss-Prot: P04626) and seven tandem repeats (VNTR) of MUC1 (UniProtKB/Swiss-Prot: P15941) were linked together by A $(EAAAK)_5$ A linker. The chimeric construct codons were optimized based on prokaryotic *(Escherichia coli)* and eukaryotic (plant) host by using GenScript service (https://www.genscript.com/ssl-bin/quote), EMBOSS (http://www.ebi.ac.uk/Tools/st/emboss_transeq/) and DyNAVacS software (http://miracle.igib.res.in/dynavac/)[31]. Then the synthetic constructs were deposited into GenBank (KF430636, KF443783).

Bioinformatics Analysis of Chimeric Protein

The messenger RNA (mRNA) secondary structure of the chimeric gene was evaluated by the mfold software (http://www.bioinfo.rpi.edu/applications/mfold). Comparison of folding and thermodynamic features of native and optimized mRNA was performed[32,33]. The prediction of the secondary structure of the recombinant protein was performed using the advanced protein secondary structure prediction server (APSSP2) (http://www.imtech.res.in/raghava/apssp2/) and GOR-IV[34].

Raptor X (http://raptorx.uchicago.edu/predict/) was used for detection of disorder regions in the chimeric protein structure. The iterative threading assembly refinement (I-TASSER) (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) and an integrated platform for automated protein structure and function prediction based on the sequence-to-structure-to-function paradigm were employed for tertiary structure prediction[35]. Solubility and antigenicity of recombinant protein was performed by online programs such as SOLpro and ANTIGENpro, respectively (http://scratch.proteomics.ics.uci.edu/).

Evaluation of Model Stability

Three dimensional (3D) structural stability of the chimeric protein in compare with original their structures were evaluated by energy minimization and RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). C-score were used to estimate the accuracy of 3D model, quantitatively and qualitatively.

Results

Designation and Construction of Chimeric Protein

Two fragments of proteins, 140 amino acids from HER-2 and MUC1 were selected. Linker consisting of EAAAK repeats was selected that is expected to form a monomeric hydrophobic α-helix and separate the different domains. It has been reported that the salt bridge Glu--Lys+ between repeated Ala can stabilize helix formation. Five repeated EAAAK sequences were introduced between two domains for more flexibility and efficient separation. Arrangement of fragment junctions and linker site are shown in Figure 1. The amino acid was back translated based on *E. coli* and plant hosts and synthetic chimeric genes were analyzed for their codon bias and GC content. Codon adaptation index of the optimized chimeric genes was 0.87 and percentage of codon having a frequency distribution of 91-100 were significantly improved to 66% for *E. coli* and 63% for plant. The overall GC content, which is a measure of transcriptional and translational efficiency, was improved to 65.25 and 53.14% after codon optimization. Within the synthetic construct, polyadenylation signal, instability elements, and the *cis*-acting sites that may have a negative influence on the expression rate were removed. Furthermore, the necessary restriction enzyme sites were introduced at the ends of the sequence for cloning purpose.

Messenger RNA Structure Prediction

In order to determine the potential folding of the chimeric genes, genetic algorithm-based RNA secondary structure prediction was combined with comparative sequence analysis. The 5' terminus of the genes was folded in the way typical of bacterial and plant gene structures. The prediction of the minimum free energy for secondary structures formed by RNA molecules was also done. ΔG of the best predicted structure was −177.92 kcal/mol for *E. coli* optimized gene and −114.01 kcal/mol for plant optimized gene. The first nucleotides at 5′ end did not have a long stable hairpin or pseudoknot (Fig. 2)[32].

Secondary Structure Prediction

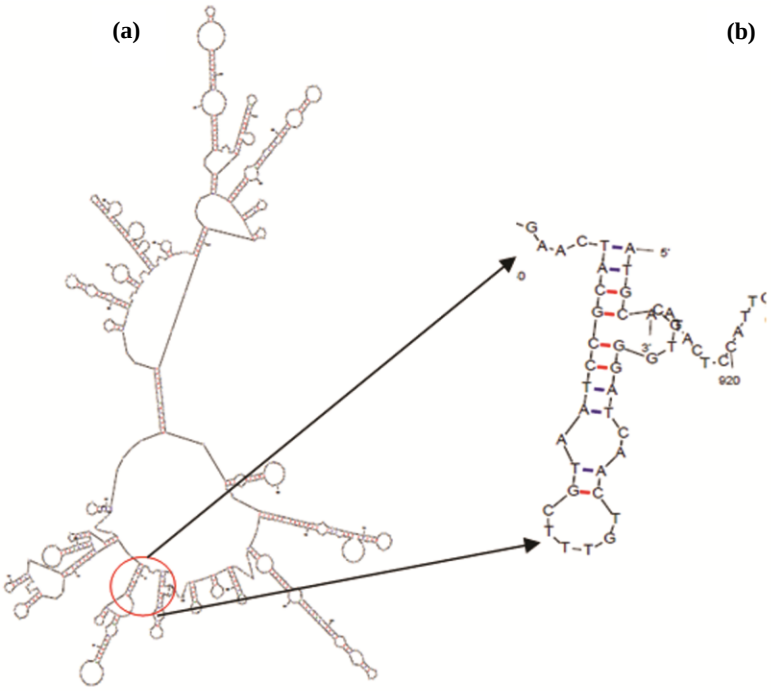
The results showed that number of residues are 308 which folds to 40 strands, 36 helices, and 232 random coils. Figure 3 shows secondary structure of fusion protein construction. The results showed that there is a helix pick between positions 141 to 168 that corresponds to the linker fragment.

Tertiary Structure Prediction of the Fusion Protein

By uploading the sequences to the I-TASSER server, tertiary structure of the fusion protein was predicted[36]. I-TASSER server is a service for protein structure and function predictions. Models are constructed based on multiple-threading alignments by LOMETS and iterative TASSER simulations. It showed a protein with two domains linked together with a linker. In order to check 3D models of protein

Fig. 1 — Schematic representation of chimeric construct consists of *her-2* and *muc1* genes bound together by an appropriate linker for expression in *E. coli* and plant.

(a)



(b)

Structural element	δG	Information
External loop	-1.30	13 ss bases & 1 closing helices.
Stack	-3.30	External closing pair is C4-G914
Helix	-3.30	2 base pairs.
Interior loop	1.10	External closing pair is C5-G913
Stack	-2.10	External closing pair is U7-A911
Stack	-1.50	External closing pair is G8-C910
Stack	-2.10	External closing pair is G9-U909
Helix	-5.70	4 base pairs.
Multi-loop	2.50	External closing pair is G10-C908 6 ss bases & 3 closing helices.
Stack	-2.10	External closing pair is A440-U906
Stack	-3.30	External closing pair is G441-C905
Stack	-2.40	External closing pair is G442-C904
Stack	-2.10	External closing pair is A443-U903
Stack	-3.30	External closing pair is G444-C902

Fig. 2 — Analysis of *her-2-muc1* mRNA stability and start codon position: a. The one predicted folding for the sequence; predicted structure has no long stable hairpin and pseudo knot at 5 ′ site of mRNA and b. Free energy details for mRNA structure. This layout mimics the html output.

structures for potential errors, ProSA was used. The z-score of the input structure was within the range of scores typically found for native proteins of similar size (Fig. 4). Tertiary prediction results of the fusion protein construction showed a protein with two domains linked together by a linker. Furthermore, protein structure prediction by the SCRATCH server showed the formation of two distinct domains including domain I (1 - 163) and domain II (164 – 321). Predicted solubility upon over expression by SOLpro algorithm was 99.31% and predicted probability of antigenicity by ANTIGENpro algorithm was 89.21%.

Evaluation of model stability

In order to calculate the profile of energy minimization, spdbv (Swiss-PdbViewer) was used. A total of -2870.531 kcal/mol indicating that the synthetic chimeric protein has acceptable stability compared to that of original structure of each domain. Additionally, the structural stability of the protein was confirmed by the data generated by a Ramachandran plot (Fig. 5). The Ramachandran plot is the two dimensional plot of the φ-ψ torsion angles of the protein backbone. It provides a simple view of the conformation of a protein. The φ-ψ angles cluster into distinct regions in the Ramachandran plot where each

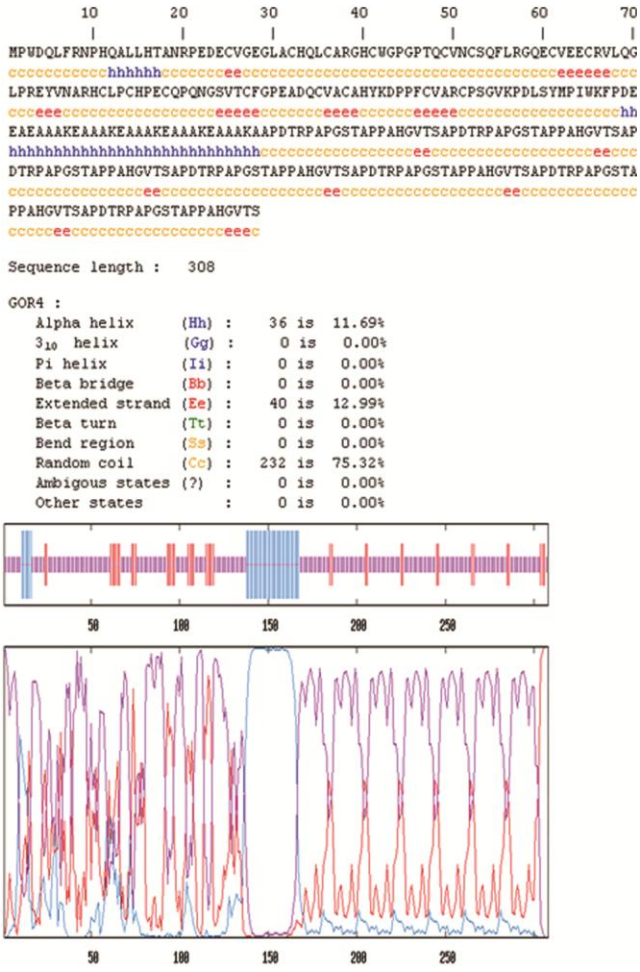


Fig. 3 — Secondary structure analysis of the HER-2-MUC1 protein Helix: Blue, Sheet: Red, Coil: Violet.

region corresponds to a particular secondary structure. There are four basic types of Ramachandran plots, depending on the stereo chemistry of the amino acid: generic (which refers to the 18 non-glycine non-proline amino acids), glycine, proline, and pre-proline (which refer to residues preceding a proline). The generic and proline Ramachandran plots are now well understood but the glycine and pre-proline Ramachandran plots are not[37]. These results were verified by the PROCHECK from Swiss model server placed most of the residues (98.4%) in allowed and only 1.6% in the disallowed regions.

Solvent Accessibility Prediction

Major hydrophobic and polarity properties of residual patterns were used to characterize the solvent accessibility distributions[38]. These patterns identified that the mean residue accessible surface area (ASA) gave a high solvent accessibility value, approximately fifty percent (data not shown).

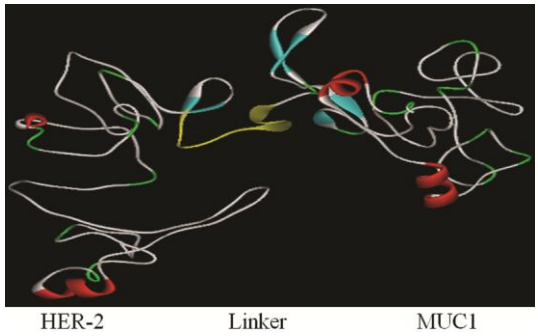


Fig. 4 — The models predicted by I-TASSER server and visualised by Discovery Studio viewer.

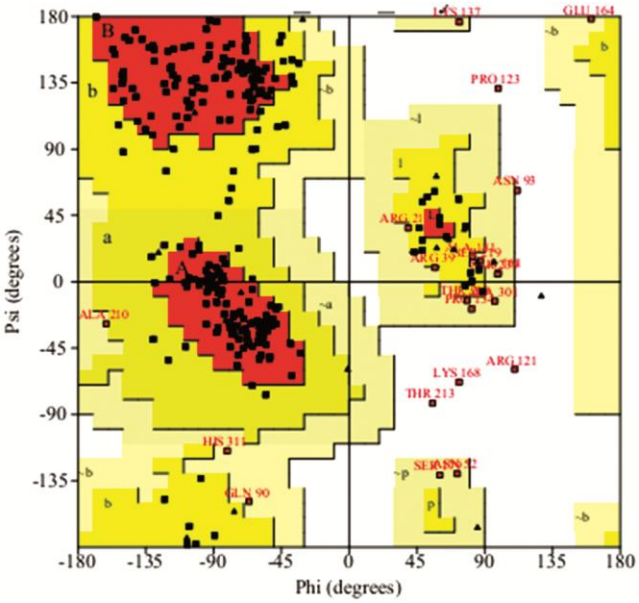


Fig. 5 — Ramachandran plot HER-2-MUC1 model.

Antigenic B-cell and T-cell epitopes

If an antigen is hydrophilic and induce both the B-cell and T-cell mediated immunity, it will be good vaccine candidate. Therefore, BCpreds and AAPpreds were used to predict B-cell epitopes of full length protein. Best epitopes were selected based on the criteria as mentioned in the materials and methods. In general, epitopes with values greater than BCpreds, AAPpreds and VaxiJen cutoff values (> 0.8, > 0.8 and > 0.5, respectively), were selected (Table 1a). The predicted B cell epitopes with ABCpred Prediction Server were ranked according to their score obtained by trained recurrent neural network and higher score of the peptide means the higher probability to be as an epitope. Furthermore, CBTOPE and Discotope servers were used to predict the conformational epitopes for B cells (Table 1b). For screening T-cell epitopes, Propred-I, Propred, and MHCPred were used to identify common T-cell epitopes that share B-

Table 1 — Continues and discontinues epitopes predicted HER2- MUC1 chimeric protein

a. Linear epitopes

No.	Start position	End position	Peptide	Number of residues	Score
1	272	308	TRPAPGSTAPPAHGVTSAPDTRPAPGSTAPPAHGVTS	37	0.816
2	72	107	PREYVNARHCLPCHPECQPQNGSVTCFGPEADQCVA	36	0.749
3	113	136	DPPFCVARCPSGVKPDLSYMPIWK	24	0.743
4	207	217	TSAPDTRPAPG	11	0.677
5	60	69	ECVEECRVLQ	10	0.665
6	152	175	KEAAAKEAAAKEAAAKAAPDTRPA	24	0.624
7	38	52	RGHCWGPGPTQCVNC	15	0.606
8	265	270	GVTSAP	6	0.537

b. Discontinous epitopes

No.	Residues	Number of residues	Score
1	A249, D251, G265, V266, T267, S268, A269, P270, D271, T272, R273, P274, A275, P276, G277, S278, T279, A280, P281, P282, A283, H284, G285, V286, T287, S288, A289, P290, D291, T292, R293, P294, A295, P296, G297, S298, T299, A300, P301, P302, A303, H304, G305, V306, T307, S308	46	0.751
2	C32, A37, R38, G39, H40, C41, W42, G43, P44, G45, P46, T47, Q48, C49, V50, N51, _:C52, _:S53, _:E60, _:C61, V62, E63, E64, C65, R66, V67, L68, Q69, L71, Y75, V76, N77, A78, R79, H80, C81, L82, P83, C84, H85, P86, E87, C88, Q89, P90, Q91, N92, G93, S94, V95, T96, C97, F98, G99, P100, E101, A102, D103, Q104, C105, V106, A107, C108, H110, D113, P114, P115, F116, C117, V118, A119, R120, C121, P122, S123, G124, V125, K126, P127, D128, L129, S130, Y131, M132, P133, I134, W135, E140, E141, A142, E143, A144, A150, A151, K152, A154, A155, A156, _:K157, E158, A159, A160, A161, K162, E163, A164, A165, A166, K167, A168, A169, P170, D171, T172, R173, P174, A175	117	0.671
3	T187, S188, A189, T207, S208, A209, P210, D211, T212, R213, P214, A215, P216, G217		

cell epitope sequence, which can interact with both MHC class I & II with the highest score (Table 2a & b).

Other Properties of the Construct

AlgPred tool and SDAP allergen library were used to predict the allergenicity of the sequence. Based on different allergenicity prediction approaches in AlgPred tool and SDAP allergen library, this protein was not detected as a potential allergen (data not shown).

Discussion

Breast cancer is the most frequent cancer in women and the second cause of cancer-related death in women all over the word. Despite advances in common treatment, a significant proportion of patients with breast cancer dies from recurrent disease, especially aggressive tumors. Hence, novel strategies are needed to improve therapies among breast cancer patients, and one of such strategies is vaccine therapy. Tumor-associated antigens (TAAs) are proteins expressed by cancer cells capable of eliciting a specific immune response[39]. Two types of TAAs, which are extensively investigated in breast cancer, are MUC1 and HER2[39,40]. HER2 is the most important TAA in breast cancer, which is normally expressed at low levels on the cell surface of many tissues about 40- fold less than tumor cells[8]. HER2 is a 1255 amino acid trans-membrane glycoprotein with a molecular weight of 185 kDa involved in many epithelial-derived cancers[41]. HER2 over expresses in 20–30% of metastatic breast cancers and its over expression causes the disruption of normal signaling pathways, and the development of resistance to apoptosis[8]. HER2 over expression by cancer cells is quite specific, therefore, therapies against this receptor have been considered as an effective anti-tumor therapy. For instance, DNA vaccine encoding full-length or truncated HER2 has protective and therapeutics potentials against HER2 positive breast tumors. However, plasmid-based DNA vaccines have not been potent for human clinical use[42]. Hence, new approaches are needed to enhance both the quality and quantity of the immune response against HER2-expressing tumors.

MUC1 (mucin 1) is a high molecular weight antigen that compromises of three domains, a 69

Table 2 — Prediction of MHC I & II epitopes

a. MHC I epitopes prediction

Types of MHC I	Sequence	Types of MHC I	Sequence	Types of MHC I	Sequence	Types of MHC I	Sequence
HLA-A1	MPWDQLFRNPHQ	HLA-B14	MPWDQLFRNPHQ	HLA-B*5103	MPWDQLFRNPHQ	HLA-Cw*0301	MPWDQLFRNPHQ
HLA-A2	MPWDQLFRNPHQ	HLA-B*2702	MPWDQLFRNPHQ	HLA-B*5201	MPWDQLFRNPHQ	HLA-Cw*0401	MPWDQLFRNPHQ
HLA-A*0201	MPWDQLFRNPHQ	HLA-B*2705	MPWDQLFRNPHQ	HLA-B*5301	MPWDQLFRNPHQ	HLA-Cw*0602	MPWDQLFRNPHQ
HLA-A*0205	MPWDQLFRNPHQ	HLA-B*3501	MPWDQLFRNPHQ	HLA-B*5401	MPWDQLFRNPHQ	HLA-Cw*0702	MPWDQLFRNPHQ
HLA-A*1101	MPWDQLFRNPHQ	HLA-B*3701	MPWDQLFRNPHQ	HLA-B*51	MPWDQLFRNPHQ	MHC-Db	MPWDQLFRNPHQ
HLA-A24	MPWDQLFRNPHQ	HLA-B*3801	MPWDQLFRNPHQ	HLA-B*5801	MPWDQLFRNPHQ	MHC-Db revised	MPWDQLFRNPHQ
HLA-A3	MPWDQLFRNPHQ	HLA-B*3901	MPWDQLFRNPHQ	HLA-B60	MPWDQLFRNPHQ	MHC-Dd	MPWDQLFRNPHQ
HLA-A*3101	MPWDQLFRNPHQ	HLA-B*3902	MPWDQLFRNPHQ	HLA-B61	MPWDQLFRNPHQ	MHC-Kb	MPWDQLFRNPHQ
HLA-A*3302	MPWDQLFRNPHQ	HLA-B40	MPWDQLFRNPHQ	HLA-B62	MPWDQLFRNPHQ	MHC-Kd	MPWDQLFRNPHQ
HLA-A68.1	MPWDQLFRNPHQ	HLA-B*4403	MPWDQLFRNPHQ	HLA-B7	MPWDQLFRNPHQ	MHC-Kk	MPWDQLFRNPHQ
HLA-A20 Cattle	MPWDQLFRNPHQ	HLA-B*5101	MPWDQLFRNPHQ	HLA-B*0702	MPWDQLFRNPHQ	MHC-Ld	MPWDQLFRNPHQ
HLA-A2.1	MPWDQLFRNPHQ	HLA-B*5102	MPWDQLFRNPHQ	HLA-B8	MPWDQLFRNPHQ		

b. MHC II epitopes prediction

RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.	RANK	POS.	N	SEQUENCE	C	MW (Da)	SCORE	% OPT.
1	104	EAD	QCVACAHYK	DPP	1004.19	21.083	39.67 %	16	32	GEG	LACHQLCAR	GHC	996.22	11.586	21.80 %
2	77	EYV	NARHCLPCH	PEC	1032.21	18.359	34.55 %	17	107	GLA	CHQLCARGH	CWG	1006.17	11.237	21.14 %
3	102	GPE	ADQCVACAH	YKD	899.01	16.816	31.64 %	18	284	QCV	ACAHYKDPP	FCV	983.12	10.637	20.02 %
4	296	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	19	264	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
5	276	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	20	244	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
6	256	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	21	224	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
7	236	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	22	204	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
8	216	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	23	184	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
9	196	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	24	12	PPA	HGVTSAPDT	RPA	865.89	9.248	17.40 %
10	176	RPA	PGSTAPPAH	GVT	815.89	14.798	27.84 %	25	105	NPH	QALLHTANR	PED	1005.14	9.193	17.30 %
11	140	FPD	EEAEAAAKE	AAA	928.97	13.593	25.58 %	26	11	ADQ	CVACAHYKD	PPF	991.15	8.022	15.09 %
12	155	KEA	AAKEAAAKE	AAA	869.98	12.331	23.20 %	27	79	RNP	HQALLHTAN	RPE	986.09	7.803	14.68 %
13	150	KEA	AAKEAAAKE	AAA	869.98	12.331	23.20 %	28	48	VNA	RHCLPCHPE	CQP	1073.27	7.783	14.64 %
14	145	AEA	AAKEAAAKE	AAA	869.98	12.331	23.20 %	29	163	GPT	QCVNCSQFL	RGQ	1023.19	7.399	13.92 %

amino acid cytoplasmatic domain involved in signal transduction, a hydrophobic trans-membrane domain of 31 amino acids and a very large extracellular domain, which consists a repetitive (20 to 125 times) 20 amino acid sequence (VTSAPDTR-PAPGSTAPPAHG), which is termed variable number of tandem repeats (VNTR)[40,43]. This core peptide (VNTR) encodes B and T cell epitopes[12]. The APDTRPA sequence within the repeats has been found to be the most immunogenic sequence of MUC1 reacting with almost all anti-MUC1 monoclonal antibodies (mAb) as well as CTLs[44]. This unique structure makes it possible to design a vaccine that exclusively targets MUC1 antigens on tumor cells while avoiding normal cells. Vaccination of breast cancer patients in remission with MUC1-keyhole limpet hemocyanin (KLH) conjugate induced high titers of specific anti-MUC1 IgM and IgG without any significant T cell activation[12]. It seems that there are some relationships between HER2 and MUC1 proteins. Of the 42,000 HER2 positive breast cancer patients in USA, < 35% are responsive to treatment with herceptin. HER2-positive cancer cells that are originally resistant to herceptin acquire sensitivity to herceptin when treated with MUC1 antagonists and herceptin. Additionally tumor cells that had acquired herceptin resistance had also acquired resistance to standard chemotherapy agents. Acquired resistance to these standard chemotherapy drugs was also reversed by combined treatment with the original drug plus a MUC1 inhibitor[45]. Further more patients with HER2-positive breast cancer often exhibit intrinsic or acquired resistance to trastuzumab treatment. The silencing MUC1, C-terminal subunit (MUC1-C) in HER2-overexpressing SKBR3 and BT474 breast cancer cells results in the down regulation of constitutive HER2 activation. Moreover, treatment with the MUC1-C inhibitor, GO-203, was associated

with disruption of MUC1-C/HER2 complexes and decreases in tyrosine phosphorylated HER2 (p-HER2) levels. Also silencing MUC1-C reverses resistance to trastuzumab and that the combination of GO-203 and trastuzumab is highly synergistic. These findings indicate that MUC1-C contributes to constitutive activation of the HER2 pathway and that targeting MUC1-C represents a potential approach to abrogate trastuzumab resistance[46].

Based on the knowledge the chimeric construct containing MUC1- HER2 peptide for expression in *E. coli* and plant was designed (Fig. 1). For vaccine designing, it should be determined which epitopes of ECD of HER2 and MUC1 are the most immunogenic epitopes that can be predicted by bioinformatics tools. Then the peptide sequence have been back translated based on *E. coli* and plant codon optimization tables to optimize the codons of these epitopes with the purpose of deletion of rare codons, instability sequences, unwanted splice sites and unwanted polyadenylation signals for expressing in *E. coli* and plant. This was achieved by improving codon adaptation index (CAI) and codon frequency distribution, and the overall GC content of the gene as well as removing negative elements that may incorporate unfavorable secondary structures on mRNA. The codon adaptation index was improved to 0.87, indicating that the optimized gene sequence could be expressed well. An appreciate gene possesses a CAI of 1.0. Also mRNA stability, its ΔG and Tm were considered[47]. Mfold is the software used for prediction of RNA secondary structure. mfold was operated by a theoretically tractable DP algorithm which can find the minimum ΔG structure within its thermodynamic model and have high ability to predict true positive base pairs. The data from mRNA structure prediction showed that the mRNA was stable enough for efficient translation in *E. coli* and plant (Fig. 2). Both *ab initio* and comparative methods were used for predicting three-dimensional structure of chimeric protein. Synthetic chimeric protein has templates in the PDB library, which based on them the 3D structures of the protein, was generated by I-TASSER server. Our result showed that *ab initio* I-TASSER software could predict the folds of synthetic chimeric protein. Estimation of the reliability and accuracy of experimental and theoretical models of protein structures is necessary. For the evaluation of the predicted models, both RMSD and TM score were used. The best RMSD value was the result of our model on template which consisted of 308 amino acids; 100% of total protein residues. Expected TM score of 0.37 ± 0.12 validates the accuracy of the model. A model of correct topology was indicated by A TM score > 0.5. Its confidence was achieved by other scores including Z-score. The Z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. Z-scores outside a range characteristic for native proteins indicate erroneous structures (Fig. 3, 4). The ProSA-web results showed that synthetic chimeric protein has features like the characteristics of native proteins. Our chimeric structure showed desirable protein stability based on Ramachandran plot predictions. A negligible 1.6% of the residues were found in Ramachandran plot analysis to be in outliner regions that could probably be due to the presence of chimeric junctions (Fig. 6).

An efficient vaccine candidate should be able to induce strong B and T cell responses. The general opinion of antigen presentation to T cells is that endogenous proteins are presented on MHC I molecules and exogenous proteins on MHC II molecules. However, many reports demonstrate that this division is not true absolutely and a significant level of cross priming can occur, because peptides

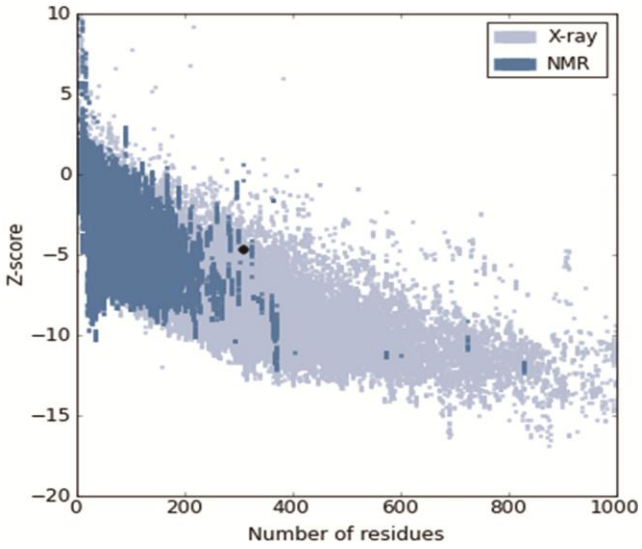


Fig. 6 — ProSA -web Z-score chimeric protein plot. The Z-score indicates overall model quality. ProSA-web Z-scores of protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The HER-2-MUC1 Z-score is highlighted as a black dot, Therefore HM is located in the range of native conformations.

from exogenous proteins may appear on MHC I molecules[48,49]. MHC I and II restricts to CD8 + cytotoxic T cells (CTL) and CD4 + T helper (TH) cells, respectively[49]. Therefore, T cell and B cell epitopes mapping is important for designing an optimal vaccine. In order to predict linear B-cell epitopes BCpred was used which works based on a novel method of subsequence kernel. The methods employed for prediction of linear epitopes led to almost identical results with some minor differences (Table 1a). Prediction of conformational epitopes in antibody-antigen interaction is a crucial step for a good design of novel drugs and vaccines. Conformational epitopes were predicted with structure-based and sequence information methods. The results (Tables 1b) showed that the structural epitopes derived from CBTOPE and DiscoTope were almost similar. In order to predict T-cell epitopes and binding affinity of MHC molecules, MHCPred, Propred and Propred-1 were used (Table 2 a, b). In this research, the potential T cell epitopes derived from antigenic B cell epitopes of chimeric protein were identified. Selected T cell epitopes were antigenic with potential to interact with human HLA alleles. Therefore, this synthetic chimeric protein has potential to induce both the B and T-cell mediated immune responses. The solubility chance of protein (20%) showed that it could be purified under denaturation condition when expressed in *E. coli*. Finally, the prediction of allergenicity of chimeric protein showed it was no allergen.

Conclusion

Our data suggest this synthetic chimeric protein as a vaccine candidate. Much work should be performed to establish this notion, which is the theme of our future research.

Acknowledgements

This investigation was supported by Applied Biotechnology Research Center, Baqiyatallah Medical Science University and National Institute of Genetic Engineering and Biotechnology (NIGEB) Tehran, Iran.

References

1 Apantaku L M, Breast cancer diagnosis and screening, American Family Physician, 62 (2000).

2 Taillefer R, Robidoux A, Lambert R, Turpin S & Laperrière J, Technetium-99m-sestamibi prone scintimammography to detect primary breast cancer and axillary lymph node involvement, *Journal of Nuclear Medicine*, 36 (1995) 1758-1765.

3 Piperkova E, Raphael B, Altinyay M E, Castellon I, Libes R *et al*, Impact of PET/CT in comparison with same day contrast enhanced CT in breast cancer management, *Clinical Nuclear Medicine*, 32 (2007) 429-434.

4 Zhou J & Zhong Y, Breast cancer immunotherapy, *Cell Mol Immunol*, 1 (2004) 247-255.

5 Hamai A, Duperrier-Amouriaux K, Pignon P, Raimbaud I, Memeo L *et al*, Antibody responses to NY-ESO-1 in primary breast cancer identify a subtype target for immunotherapy, *PLoS One*, 6 (2011) e21129.

6 Ludovini V, Gori S, Colozza M, Pistola L, Rulli E *et al*, Evaluation of serum HER2 extracellular domain in early breast cancer patients: correlation with clinicopathological parameters and survival, *Ann Oncol*, 19 (2008) 883-890.

7 Mohanty K, Saha A, Pal S, Mallick P, Chatterjee S K *et al*, Anti-tumour immunity induced by an anti-idiotype antibody mimicking human Her-2/neu, *Breast Cancer Res Treat*, 104 (2007) 1-11.

8 Daniele L & Sapino A, Anti-HER2 treatment and breast cancer: state of the art, recent patents, and new strategies, *Recent Pat Anticancer Drug Discov*, 4 (2009) 9-18.

9 Chentoufi A A, Nesburn A B & BenMohamed L, Recent advances in multivalent self adjuvanting glycolipopeptide vaccine strategies against breast cancer, *Archivum Immunologiae Et Therapiae Experimentalis*, 57 (2009) 409-423.

10 Lekka E, Gritzapis A D, Perez S A, Tsavaris N, Missitzis I *et al*, Identification and characterization of a HER-2/neu epitope as a potential target for cancer immunotherapy, *Cancer Immunol Immunother*, 59 (2010) 715-727.

11 Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag P M *et al*, Identification of early molecular markers for breast cancer, *Mol Cancer*, 10 (2011) 15.

12 Mushenkova N, Moiseeva E, Chaadaeva A, Den Otter W, Svirshchevskaya E, Antitumor effect of double immunization of mice with mucin 1 and its coding DNA, *Anticancer Res*, 25 (2005) 3893-3898.

13 Rakha E A, Boyce R W, El-Rehim D A, Kurien T, Green A R *et al*, Expression of mucins (MUC1, MUC2, MUC3, MUC4, MUC5AC and MUC6) and their prognostic significance in human breast cancer, *Modern Pathology*, 18 (2005) 1295-1304.

14 Apostolopoulos V, Pietersz G A, Tsibanis A, Tsikkinis A, Drakaki H *et al*, Pilot phase III immunotherapy study in early-stage breast cancer patients using oxidized mannan - MUC1 [ISRCTN71711835], *Breast Cancer Res*, 8 (2006) R27.

15 Wilkie S, Picco G, Foster J, Davies D M, Julien S *et al*., Retargeting of human T cells to tumour-associated MUC1: the evolution of a chimeric antigen receptor, *J Immunol*, 180 (2008) 4901-4909.

16 Rahbarizadeh F, Rasaee M J, Forouzandeh M, Allameh A A, Over expression of anti-MUC1 single-domain antibody fragments in the yeast *Pichia pastoris*, *Mol Immunol*, 43 (2006) 426-435.

17 Nazarian S, Mousavi Gargari S L, Rasooli I, Amani J *et al*, An *in silico* chimeric multi subunit vaccine targeting virulence factors of enterotoxigenic *Escherichia coli* (ETEC) with its bacterial inbuilt adjuvant, *J Microbiol Methods*, 90 (2012) 36-45.

18 Akhoon B A, Slathia P S, Sharma P, Gupta S K & Verma V, *In silico* identification of novel protective VSG antigens expressed by *Trypanosoma brucei* and an effort for designing a highly immunogenic DNA vaccine using IL-12 as adjuvant, *Microbial Pathogenesis*, 51 (2011) 77-87.

19 Jacob L & Vert J P, Efficient peptide-MHC-I binding prediction for alleles with few known binders, *Bioinformatics*, 24 (2008) 358-366.

20 Larsen J E P, Lund O & Nielsen M, Improved method for predicting linear B-cell epitopes, *Immunome Research*, 2 (2006) 2.

21 Saha S & Raghava G, Prediction of continuous B cell epitopes in an antigen using recurrent neural network, *Proteins: Structure, function, and bioinformatics*, 65 (2006) 40-48.

22 Haste Andersen P, Nielsen M & Lund O, Prediction of residues in discontinuous B-cell epitopes using protein 3D structures, *Protein Sci*, 15 (2006) 2558-2567.

23 Sun J, Wu D, Xu T, Wang X, Xu X *et al*, SEPPA: a computational server for spatial epitope prediction of protein antigens, *Nucleic Acids Research*, 37 (2009) W612-W616.

24 Ponomarenko J, Bui H H, Li W, Fusseder N, Bourne P E *et al*, ElliPro: a new structure-based tool for the prediction of antibody epitopes, *BMC Bioinformatics*, 9 (2008) 514.

25 Nielsen M, Lundegaard C, Lund O & Keaymir C, The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage, *Immunogenetics*, 57 (2005) 33-41.

26 Diez-Rivero C M, Lafuente E M & Reche P A, Computational analysis and modeling of cleavage by the immunoproteasome and the constitutive proteasome, *BMC Bioinformatics*, 11 (2010) 479.

27 Bhasin M & Raghava G P S, Analysis and prediction of affinity of TAP binding peptides using cascade SVM, *Protein Science*, 13 (2004) 596-607.

28 Singh H & Raghava G, ProPred: prediction of HLA-DR binding sites, *Bioinformatics*, 17 (2001) 1236-1237.

29 Bhasin M & Raghava G P S, A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes, *J Biosciences*, 32 (2007) 31-42.

30 Bhasin M & Raghava G P S, Prediction of CTL epitopes using QM, SVM and ANN techniques, *Vaccine*, 22 (2004) 3195-3204.

31 Harish N, Gupta R, Agarwal P, Scaria V & Pillai B, DyNAVacS: an integrative tool for optimized DNA vaccine design, *Nucleic Acids Research*, 34 (2006) W264-W266.

32 Zuker M, Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Research*, 31 (2003) 3406-3415.

33 Sato K, Hamada M, Asai K & Mituyama T, CENTROIDFOLD: a web server for RNA secondary structure prediction, *Nucleic Acids Research*, 37 (2009) W277-W280.

34 Garnier J, Gibrat J-F, Robson B & Russell F D, GOR method for predicting protein secondary structure from amino acid sequence, *Methods in Enzymology*: Academic Press; 1996. 540-553.

35 Roy A, Kucukural A & Zhang Y, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat Protoc*, 5 (2010) 725-738.

36 Zhang Y, I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics*, 9 (2008) 40.

37 Ho B K & Brasseur R, The Ramachandran plots of glycine and pre-proline, *BMC Structural Biology*, 5 (2005) 1.

38 Rost B & Sander C, Conservation and prediction of solvent accessibility in protein families, *Proteins*, 20 (1994) 216-226.

39 Peoples G E, Holmes J P, Hueman M T, Mittendorf E A, Amin A *et al*, Combined clinical trial results of a HER2/neu (E75) vaccine for the prevention of recurrence in high-risk breast cancer patients: U.S. Military Cancer Institute Clinical Trials Group Study I-01 and I-02, *Clin Cancer Res*, 14 (2008) 797-803.

40 Chentoufi A A, Nesburn A B & BenMohamed L, Recent advances in multivalent self adjuvanting glycolipopeptide vaccine strategies against breast cancer, *Arch Immunol Ther Exp*, 57 (2009) 409-423.

41 Peoples G E, Gurney J M, Hueman M T, Woll M M, Ryan G B *et al*, Clinical trial results of a HER2/neu (E75) vaccine to prevent recurrence in high-risk breast cancer patients, *J Clin Oncol*, 23 (2005) 7536-7545.

42 Baloria U, Akhoon B A, Gupta S K, Sharma S, Verma V, *In silico* proteomic characterization of human epidermal growth factor receptor 2 (HER-2) for the mapping of high affinity antigenic determinants against breast cancer, *Amino Acids*, 42 (2012) 1349-1360.

43 Croce M V, Isla-Larrain M T, Capafons A, Price M R & Segal-Eiras A, Humoral immune response induced by the protein core of MUC1 mucin in pregnant and healthy women, *Breast Cancer Res Treat*, 69 (2001) 1-11.

44 Singh R, Samant U, Hyland S, Chaudhari P R & Wels W S, Target-specific cytotoxic activity of recombinant immunotoxin scFv (MUC1)-ETA on breast carcinoma cells and primary breast tumors, *Mol Cancer Ther*, 6 (2007) 562-569.

45 Fessler S P, Wotkowicz M T, Mahanta S K & Bamdad C, MUC1 is a determinant of trastuzumab (herceptin) resistance in breast cancer cells, *Breast Cancer Res & Treat*, 118 (2009) 113-124.

46 Raina D, Uchida Y, Kharbanda A, Rajabi H, Panchamoorthy G *et al*, Targeting the MUC1-C oncoprotein downregulates HER2 activation and abrogates trastuzumab resistance in breast cancer cells, *Oncogene*, 33 (2014) 3422-3431.

47 Amani J, Mousavi S L, Rafati S & Salmanian A H, *In silico* analysis of chimeric espA, eae and tir fragments of *Escherichia coli* O157:H7 for oral immunogenic applications, *Theor Biol Med Model*, 6 (2009) 28.

48 Norbury C C, Hewlett L J, Prescott A R, Shastri N & Watts C, Class I MHC presentation of exogenous soluble antigen via macropinocytosis in bone marrow macrophages, *Immunity*, 3 (1995) 783-791.

49 Zinkernagel R M, On cross-priming of MHC class I-specific CTL: rule or exception?, *Eur Jr Immunol*, 32 (2002) 2385-2392.