ARTICLE

Insilico Anticancer Peptide Prediction from *Curcuma longa*

Sunil Kumar Suryawanshi[1], Usha Chouhan[2] and Jyoti Kant Choudhari[2,*]

[1]Centre for Knowledge Management & e-Governance, Atal Bihari Vajpayee Institute of Good Governance & Policy Analysis, Bhopal, India

[2]Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal, India

*Corresponding Author: Jyoti Kant Choudhari. Email: jtchoudhary27@gmail.com

ABSTRACT

Cancer is the second biggest cause of death globally, and the use of therapeutic peptides to specifically target and destroy cancer cells has gotten much interest. Cancer peptides or vaccinations are utilized to treat cancer nowadays, apart from chemotherapy, which has significant discomfort, side effects and costly. It is time demanding to identify and predict potential anticancer peptides using computational biology approaches. Thus, 3-D molecular modeling is being used to find possible ACP candidates. In this research, *Curcuma longa* has predicted peptide sequences were docked on breast cancer receptors and used a molecular docking technique to assess the anticipated peptides' binding affinities to MHC molecules. A similar approach was utilized to simulate the interactions of the chosen peptide with the TCR. Additionally, the Pep10 LIRQHVASNIGIAKSKIREPIV was examined, and our findings indicated interaction with MHC classes I and II. However, the maximum binding energy was obtained with TCR at 695.61, giving strength through eight hydrogen bonds. Similarly, the Pep20, GAIIGNRKIKLQPHIIIRID, the projected, has the most significant overall binding energy with MHC classes I and II but a lower global E total value with TCR, namely −600.97 kj/Mol, and also four hydrogen bonds. This research could lead to the development of novel anticancer drugs based on the anticancer activity of the *Curcuma longa* medicinal plant.

KEYWORDS

Anticancer; molecular modelling; molecular docking; cancer; peptide

1 Introduction

Cancer therapies have increasingly been associated with surgery, chemotherapy, radiation, and hormone therapy [1,2]. These therapies are critical in treating cancer initially, but the danger of recurrence may continue throughout the dosage used, resulting in drug resistance [3]. In this modern era, novel approaches such as cancer vaccines are promising areas to challenge this condition actively as a clinical therapeutic to recognize the antigens by inducing the immune system. Cancer vaccines are promising because they can help the immune system recognize the antigens [4]. Peptide engineering creates shorter peptide fragments that target specific immune responses, such as peptide vaccines or peptidomimetics [5,6].

The interactions between the compounds and selected targets have been analyzed by *insilico* modeling to verify potential binding for anticancer drug targets. Docking and binding affinity calculations were carried out and compared with known anticancer agents. Regarding other immunogens, selective T and B

cell-mediated positive primary immunological response is evoked by the peptide vaccines with antigenic epitopes ranging from 6–15 amino acids [7]. Antigen presentation processing pathways in the case of T-cell mediated immune response reveal that the peptide antigens are presented on the outer surface of cancer cells specific to T-cell receptors (TCR) recognized by Major Histocompatibility Complex (MHC) of class I and class II molecules [8]. Proteasome machinery is active endogenously in the Antigen-presenting cells (APCs) resulting from converting large protein fragments into short peptides with 6–10 amino acids and presenting this peptide in association with class I MHC molecules recognized by CD8+ cytotoxic T lymphocytes (CTL) [9]. Simultaneously, long peptides ranging from 11–30 amino acids were produced from exogenous antigens processing pathways combined with Class II MHC molecules, accepted by helper (CD4+) T cells [10].

X-ray crystallographic studies revealed the presence of class I MHC/peptide complexes which depict the information about the conformation of a peptide in 3-dimensional space. It is network formation in terms of bonding between peptides, and complementary MHC residues with hydrogen bonds have an affinity to hold this contact tightly [11]. Electrostatic interactions and hydrogen bonding with peptide side chains are common. However, the specific binding is governed by the interaction between charged amino acids and carboxyl-termini of peptides with the binding groove of polar residues at both ends of MHC residues [12]. When the MHC-peptide complex has been formed, the maximum portion of central residues of the peptides disclosed in the MHC complexes is identified by TCR [13].

In the engineering of peptide vaccines, the aim should be to produce efficient induction of either B-cell or cytotoxic T-cell responses. The main aspect of this is capable helper T cell responses, which are necessary for the immune system to control humoral or adaptive immunity [14,15]. So, the challenging opportunities exit remain here to distinguish active epitopes from the peptides that can evoke specific immune responses that encounter T cells to provide defensive immunity [16]. MHC is a widely diversified protein in the human population, so it is vital to deal with the candidate peptide vaccine by the APCs. The tremendous importance of peptide vaccines deals with saving time and expense reduction to recognize reliable identification of peptide vaccines because, in nature, MHC alleles and candidate peptides are found to be diverse [17].

In bioinformatics, molecular docking is a widely utilized methodology for structure-based rational drug designing. It has all the features needed to become a complete tool and solve the binding prediction for MHC and peptide affinity. Most successful attempts have been made with small molecules as ligands with molecular docking. Peptide docking is also a central thirsty area with promising outcomes to produce and subsequently find the best match with MHC classes I and II complexes. This bioinformatics approach was a better, rapid, accurate, and reliable way to predict peptide binding with MHC molecules [18,19].

This study selected only 20 peptide sequences for the three-dimensional modeling and molecular docking analysis against breast cancer. We enforced a molecular docking approach to determine the binding affinities of the predicted peptides with MHC molecules. A similar approach is used to model the interactions between the selected peptide with TCR α and β.

2 Materials and Methods

2.1 Data Sets

In this study, 165 protein sequences were downloaded from NCBI using the keyword "*Curcuma longa.*" Protein sequences have their uniqueness and visual language, retaining a wide variety of means, and from where we find some unique peptide designs [20]. The enormous peptide will be generated using these protein sequences *insilico* digestion, further evaluating the performance of our methodology. The prediction tools available in the CAMP database were utilized. It is based on various machine learning algorithms such as Support Vector Machine (SVM), Random forest (RF), Artificial neural network (ANN) and Discriminant Analysis (DA). The results accuracy for different statistical models ranges from 87% to 93%, and their antimicrobial activity is independent of length. The antibacterial potential of *Curcuma longa*

protein fragments generated following proteolysis was evaluated using four statistical models from the CAMP database. The results were used to manually select peptides where a positive result was reported throughout the four statistical models. Thus, 20 potentially antimicrobial peptides were identified with a prediction value greater than 0.45 for at least three different statistical models. Further, all these peptides have been selected for further processing. Fig. 1 depicts the Insilico analysis process step-by-step.
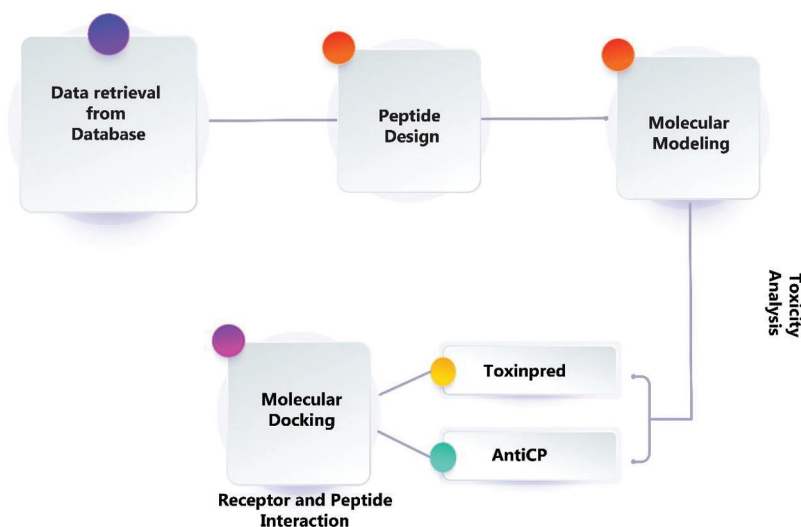

Figure 1: Workflow of *insilico* anticancer peptide design and analysis

2.2 Molecular Modeling

The 3D structure of the peptide sequences is predicted using the PEP-FOLD server. PEP-FOLD is an online tool for predicting the 3D conformation of peptides [21]. It allows the linear and disulfide bonded cyclic peptide with a limitation of 9–36 amino acids for modeling. All models are visualized using the Discovery Studio 2.5 software [22,23].

2.3 Prediction of Anticancer and Toxicity Properties

The AntiCP is a server that predicts the anticancer effects of peptides. It is used to predict the anticancer properties of peptides [23], and The Toxinpred server predicts the peptides' toxicity. It employs a machine learning technique based on the SVM classifier [24]. Algorithms for both the tool are based on machine learning multidimensional techniques. The quantitative matrix belongs to peptides' properties as valuable features related to the anticancer and toxicity. This server is based on the training and testing of specific features related to cancer and toxic protein datasets.

2.4 Molecular Docking

The protein-peptide docking of 20 peptides with MHC-I (1i4f, 1a07), MHC-II (1bx2, 1fv1, 1h15, 1a6a, 1j8h, 1jk8, 31qz), and TCR (3mff) was performed to determine their overall binding affinities using Hex 8.0 software. Figs. 2A–2K depict the 3D structures of MHC-I, MHC-II, and TCR αβ. It is a well-known tool for calculating and presenting possible docking pairs of protein and peptide molecules in an interactive molecular graphics computer. In this docking procedure, the ligand molecules assume to be rigid and can superimpose pairs of molecules using only their 3D shapes. Hex is one of the few docking programs based on the spherical polar fourier (SPF) correlations to accelerate the calculations and has built-in graphics to view the results. Furthermore, it is the first protein docking program to use modern graphic processor units (GPUs) to accelerate the calculation.

(A)

(B)

(C)

(D)

(E)

(F)
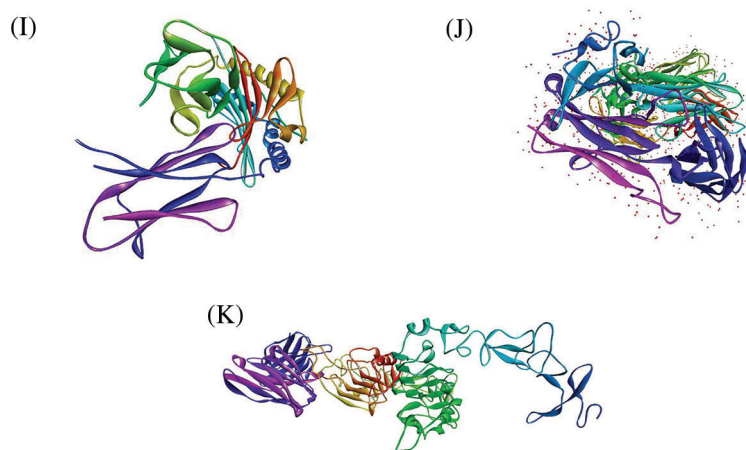
(G)

(H)
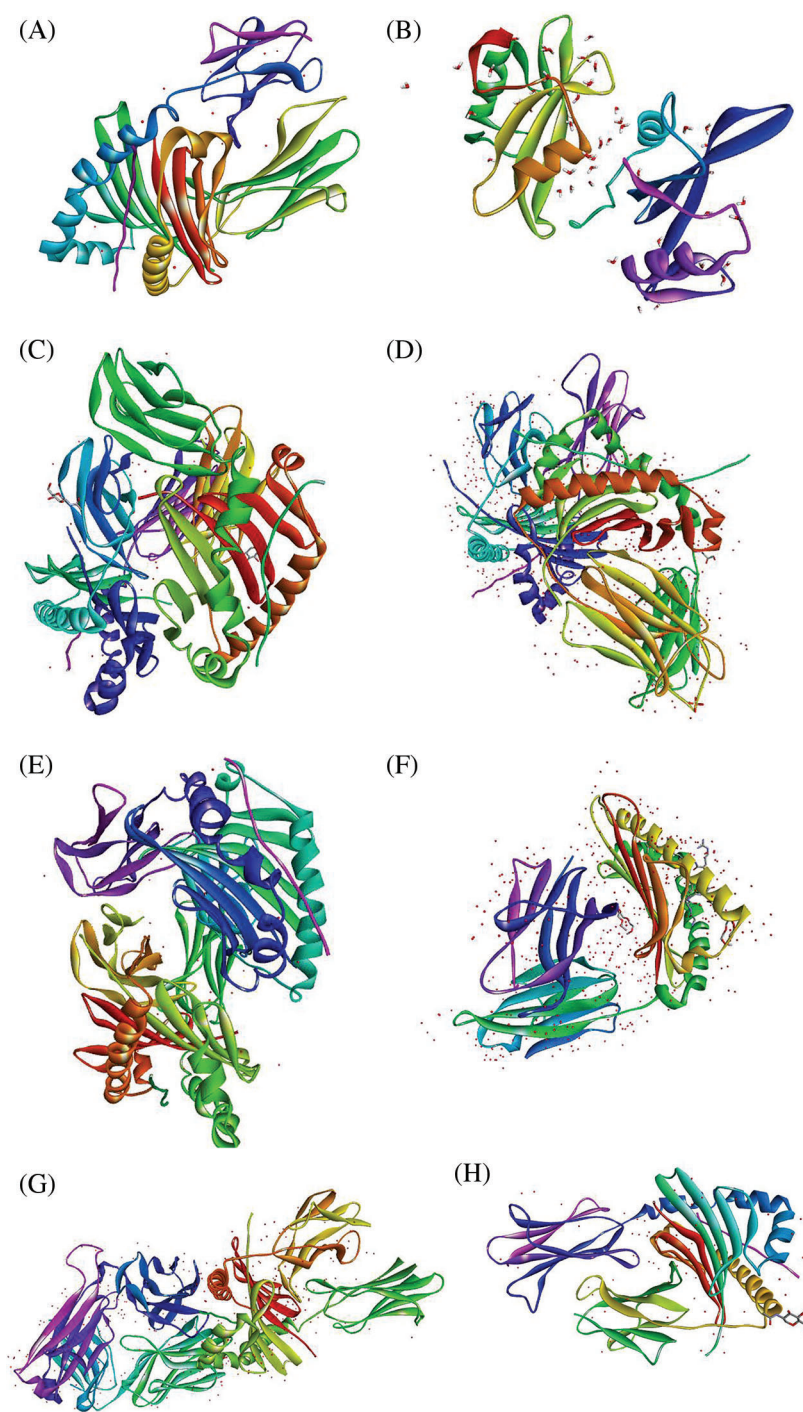
Figure 2: (Continued)

Figure 2: 3D structures of (A) MHC-I (PDB Id: 1a6a) (B) MHC-I (PDB Id: 1a07) (C) MHC-II (PDB Id: 1bx2) (D) MHC-II (PDB Id: 1fv1) (E) MHC-II (PDB Id: 1h15) (F) MHC-II (PDB Id: 1i4f) (G) MHC-II (PDB Id: 1j8h) (H) MHC-II (PDB Id: 1jk8) (I) MHC-II (PDB Id: 3lqz) (J) TCR αβ (PDB Id: 3mff) (K) HER2 and ECD (PDB Id: 6j71)

2.5 Docking Complex Visualization

Accelyra's computational tools for protein-protein docking and protein domain development are used to develop a modified structural model [25]. The Discovery Studio 2.5 is used to analyze the docked complex [26].

3 Results and Discussion

In this study, the toxicity and anticancer properties of the peptide are investigated using the web server based on machine learning approaches. The AntiCp server determined that only Pep11 was toxic, as shown in Table 1. The negative SVM score predicted the possibility of a non-toxic peptide sequence in nature. All other peptides may be regarded as possible bioactive peptides against cancer for additional docking study, except Pep11. Consequently, we have predicted 20 antimicrobial peptides based on physicochemical properties (Protein length, peptide mass, hydrophobicity, steric hindrance, side bulk, hydropathicity, amphipathicity, hydrophilicity, net hydrogen, charge, pI, boman index, hydropathy index, aliphatic index, instability index, GRAVY (grand average of hydropathicity) from the proteins of medicinal plants, i.e., *Curcuma longa*, which can also act as an anticancer peptide. So, we consider these peptides under investigation for anticancer peptides.

Table 1: Prediction of anticancer and toxicity of experimental peptides

S. No.	Peptide sequences	Anti-cancer prediction SVM score prediction	Toxicity prediction SVM score prediction
1	VWVKPWLHFK	0.63 Anticp	−0.93 Non-Toxin
2	AFGWTFYFLNQLPAIISNNIGK	0.72 Anticp	−1.56 Non-Toxin
3	VSAAINAGLVIDPLLQCLNEWNGAPIPIC	0.80 Anticp	−0.73 Non-Toxin
4	LGIQAFQPILVEGSAICLHPLVCK	0.70 Anticp	−1.14 Non-Toxin
5	DIGTLYLIFGAIAGVMGTCFSVLIR	0.69 Anticp	−1.08 Non-Toxin
6	VLGFPSKPIGLFIK	0.71 Anticp	−1.25 Non-Toxin
7	GIVCSAVNSETGEQVAIKKIGNA	0.37 Anticp	−1.06 Non-Toxin

(Continued)

Table 1 (continued)

S. No.	Peptide sequences	Anti-cancer prediction SVM score prediction	Toxicity prediction SVM score prediction
8	RRICERAAIRKRHLYL	1.11 Anticp	−0.84 Non-Toxin
9	RRICEKAAIRKRHLYL	1.09 Anticp	−0.61 Non-Toxin
10	LIRQHVASNIGIAKSKIREKEPIV	0.71 Anticp	−1.00 Non-Toxin
11	LVICQRSNKNTCMHQKPQVSRGKCIKKGQI	0.76 Anticp	0.30 Toxin
12	VNERARRYIRQLPRHIRQSF	0.67 Anticp	−1.09 Non-Toxin
13	FSPTSGGCRGIRCSANIVGECPAQ	0.77 Anticp	−0.21 Non-Toxin
14	GGIRCPLTVVQSRGIGTIISSP	0.54 Anticp	−0.52 Non-Toxin
15	GATIGESRIKLQPHIIQKI	0.65 Anticp	−1.16 Non-Toxin
16	KFPPPKVIQVSKSAW	0.85 Anticp	−0.45 Non-Toxin
17	LGPPWGNAPGAVANRVAL	0.66 Anticp	−0.60 Non-Toxin
18	CTISGHPISKSIW	0.93 Anticp	−0.86 Non-Toxin
19	GAIIGDSKIRLQPHIIKRIIS	0.51 Anticp	−0.69 Non-Toxin
20	GAIIGNRKIKLQPHIIIRIID	0.57 Anticp	−0.61 Non-Toxin

Physicochemical properties of 20 predicted Antimicrobial Peptides (AMPs) from *Curcuma longa* have been computed and compared with the AMP patent database shown in Table 2. The comparison included the frequency and composition (%) of the amino acids in each of the 20 anticipated AMPs. The most frequent amino acids found in AMP are 10 which cover more than 66% in the proteolytic cleavage of *Curcuma longa* and 77% of the overall amino acid composition, i.e., GAVLIKNSTE in the patent database. On the other hand, 44% & 23% of amino acid composition cover M P W Q Y C F D R H, respectively. The content of amino acids in the proteolytic peptides was calculated and compared with the patent peptide database. The frequency of specific amino acids is more common in the AMPs and can influence their biological activity. For docking and interaction analyses, only 20 peptides were considered. The main goal of this study was to figure out which amino acids are involved in MHC classes I, II, and TCR binding to Pep1 at Pep20. Finally, compare the binding affinity of the energy state of the peptide-receptor complex. Subdomain III of Human Epidermal Growth Factor Receptor (HER2) and sub-domain III of the extracellular domain (ECD), HER 2 (Fig. 2K 6j71) is overexpressing SK-BR-3 breast cancer cells, which is one of the best crucial targets in the cancer research for the development of immunotherapeutic strategy because the cell surface proteins are overexpressed under the clinical condition like breast, gastric and ovarian cancer.

Table 2: Comparison of amino acid frequency and its composition between AMPs in *Curcuma longa* and patent database of CAMP

One letter code of amino acid	AMPs in *Curcuma longa* proteins		AMPs patent database at CAMP	
	Amino acid frequency	Amino acid composition (%)	Amino acid frequency	Amino acid composition (%)
G	37	8.02603	25096	7.20841
A	32	6.94143	24649	7.08001
V	25	5.42299	24067	6.91284

(Continued)

Table 2 (continued)

One letter code of amino acid	AMPs in *Curcuma longa* proteins		AMPs patent database at CAMP	
	Amino acid frequency	Amino acid composition (%)	Amino acid frequency	Amino acid composition (%)
L	33	7.15835	41118	11.8105
I	70	15.1844	26000	7.46807
M	4	0.86768	7961	2.28666
P	29	6.29067	9544	2.74135
F	15	3.2538	15310	4.39754
W	7	1.51844	2302	0.66121
N	17	3.68764	17578	5.04899
Q	20	4.33839	13487	3.87392
S	28	6.07375	25356	7.28309
T	10	2.1692	17908	5.14377
Y	8	1.73536	10752	3.08833
C	16	3.47072	4884	1.40285
D	4	0.86768	15706	4.51129
E	11	2.38612	19571	5.62144
K	43	9.32755	25050	7.1952
R	38	8.24295	16227	4.66094
H	14	3.03688	5583	1.60362
Total	**461**	**100**	**348149**	**100**

Table 3 shows that most antimicrobial peptides are cationic amphipathic peptides, also containing one or more hydrophobic residues. Hydrophobic amino acids such as I G L account for more than 30 percent of analyzed AMPs from *Curcuma longa* and play an important role in secondary structure formation and interactions with bacterial membrane. The cationic and hydrophobic characters peptides affect the antimicrobial activity. The analysis of amino acid content of antimicrobial peptides revealed many fragments with the predominance of one or several amino acids. In comparison with all AMPs listed in the patent database, amino acids such as V & S are less frequently encountered from *Curcuma longa*. On average, the evaluated peptides in both cases have a higher content of I, K & G. In both cases, the result shows the significant frequency of serine to be antimicrobial.

The selected peptides in our work are potentially antimicrobial AMPs with 6 to 30 amino acids, corresponding to a peptide mass of 117.175 to 3427.126 Da. The general classification of AMPs includes cationic peptides divided into three subclasses: linear peptides forming helical structures, cysteine-rich open-ended peptides containing single or several disulfide bridges and molecules rich in specific amino acids, such as proline, glycine, histidine and tryptophan. The majority of antimicrobial peptides with positive charges are known as cationic antimicrobial peptides (CAPs). In our study, 85.71%, 3.57% & 7.14% of peptides are found to have a cationic, anionic and neutral charge.

MHC involved in breast cancer interacts with predicted peptides from *Curcuma longa*. The highest binding energy molecules further selected and interacted with TCR to determine its affinity is shown in Table 4.

Table 3: List of predicted AMPs from *Curcuma longa* and its physicochemical properties

| S. No.	Peptide	Protein	Peptide length	Peptide mass	Net hydrogen	Charge	pI	Hydrophobicity	Hydrophilicity	Steric hindrance	Hydropathicity	Amphipathicity	Boman index	Hydropathy	Aliphatic index	Instability index
1	VVVKPWLHFK	Cysteine protease inhibitor	10	1339.649	0.7	2.5	10.02	0.03	−0.86	0.53	0.06	0.88	−43.68	−1.015	97	−25.85
2	AFGWTFYFLNQLPAIISNNIGK	chalcone synthase-like protein	22	2514.909	0.64	1	8.94	0.09	−0.89	0.63	0.41	0.22	−293.8	−0.606	97.73	58.95
3	VSAAINAGLVIDPLLQCLNEWNGAPIPIC	phenylalanine ammonia-lyase, partial	29	3005.54	0.41	−2	3.67	0.12	−0.61	0.6	0.88	0.09	−12.02	−0.556	141.38	21.55
4	LGIQAFQPILVEGSAICLHPLVCK	RNApolymeraseC, partial (plastid)	24	2550.116	0.38	0.5	7.03	0.13	−0.62	0.58	1.09	0.37	−35.16	−1.223	146.25	28.92
5	DIGTLYLIFGAIAGVMGTCFSVLIR	cytochrome oxidase subunit 1, partial (mitochondrion)	25	2631.187	0.36	0	6.15	0.21	−0.83	0.64	1.53	0.1	−75.57	−0.791	140.4	−0.59
6	VLGFPSKPIGLFIK	unnamed proteinproduct	14	1515.902	0.36	2	10.02	0.14	−0.53	0.61	0.99	0.52	−152.1	−0.366	132.14	25.35
7	GIVCSAVNSETGEQVAIKKIGNA	mitogenactivated	23	2288.601	0.6	0	6.46	0	0	0.6	0.2	0.4	−27	0	101.74	−2.3
8	RRICERAAIRKRHLYL	diketide- CoA synthase 2	16	2054.493	1.56	5.5	11.44	−0.49	0.56	0.6	−0.73	1.17	1.52	1.5	110	77.59
9	RRICEKAAIRKRHLYL	diketide- CoA synthase 1	16	2026.48	1.44	5.5	10.95	−0.45	0.56	0.6	−0.69	1.24	−30.96	−0.279	110	77.59
10	LIRQHVASNIGIAKSKIREKEPIV	RNA polymerase C, partial (plastid)	24	2700.223	0.92	3.5	10.29	−0.18	0.27	0.59	−0.13	0.85	−248.7	−1.016	130	33.33
11	LVICQRSNKNTCMHQKPQVSRGKCIKKGQI	RNA polymerase B, partial (plastid)	30	3427.126	1.1	7.5	10.23	−0.31	0.24	0.61	−0.72	0.96	−87.07	−1.048	71.33	53.32
12	VNERARRYIRQLPRHIRQSF	Mitogen-activated protein kinase 1	20	2595.999	1.7	5.5	12.01	−0.53	0.46	0.62	−1.31	1	−55.94	−1.032	78	113.06
13	FSPTSGGCGRGIRCSANIVGECPAQ	Pathogenesis-related protein 5, partial	24	2410.723	0.68	1	8.27	−0.11	−0.07	0.59	−0.01	0.3	3.53	0.86	52.92	68.41
14	GGIRCPLTVVQSRGIGTIISSP	TURM_CU RLORecName; Full=Turmerin	22	2211.608	0.65	2	10.38	−0.01	−0.29	0.59	0.5	0.27	−172	−0.234	115	65.93
15	GATIGESRIKLQPHIIQKI	polyprotein, partial	19	2102.508	0.8	2.5	10.01	−0.11	0.02	0.57	−0.09	0.75	−43.62	−0.429	128.42	58.33
16	KFPPPKVIQVSKSAW	A Chain A, Lipoxygena se InComplex with Protocatechuic Acid	15	1712.068	0.69	3	10.31	−0.1	−0.09	0.54	−0.31	0.77	−204	−0.308	71.33	37.17
17	LGPPWGGNAPGAVANRVAL	ribulose- 1,5-bisphosphate carboxylase/ oxygenase large subunit, partial	18	1760.029	0.47	1	10.11	0.04	−0.45	0.55	0.25	0.13	−203.8	−0.819	97.78	14.05
18	CTISGHPISKSIW	maturase K, partial (chloroplast)	13	1428.669	0.57	1.5	8.57	0.02	−0.49	0.5	0.21	0.37	−5.95	0.683	90	32.09
19	GAIIGDSKIRLQPHIIKRIIS	polyprotein, partial	21	2328.83	0.82	3.5	11.01	−0.09	0.02	0.59	0.28	0.68	−964.1	−0.622	153.33	67.71
20	GAIIGNRKIKLQPHIIRIID	polyprotein, partial	21	2381.937	0.86	3.5	11.01	−0.06	−0.09	0.63	0.41	0.71	−521.7	−1.159	171.9	34.93

Table 4: Interaction of MHC and T-cytotoxic receptor with 20 different peptides (1–20) by using Hex 8.0 docking tool

Peptide	MHC-I (kj/Mol)		MHC-II (kj/Mol)							TCR (kj/Mol)
	HLA-A0201		HLA-DR2		DRB50 101	HLA-DR3	HLA-DR4	HLA-DQ8	HLA-DP2	TCR-αβ
	1i4f	1a07	1bx2	1fv1	1h15	1a6a	1j8h	1jk8	31qz	3mff
Pep1	−425.34	−43.75	−556.01	−317.09	−530.47	−458.87	−41.57	−461.24	−441.44	−499.57
Pep2	−426.35	−129.26	−734.46	−333.74	−659.66	−516.32	−93.54	−464.14	−577.94	−625.60
Pep3	−502.20	−101.54	−734.46	−356.31	−622.01	−534.52	−96.73	−566.08	−587.34	−587.72
Pep4	−297.44	−156.80	−711.18	−193.05	−621.48	−494.84	−166.64	−330.78	−564.69	−546.74
Pep5	−324.99	−210.97	−717.08	−235.38	−697.57	−526.40	−145.95	−325.85	−563.98	−558.23
Pep6	−380.88	−87.70	−608.29	−284.86	−615.24	−507.39	−66.33	−374.70	−536.06	−509.08
Pep7	−404.71	−161.36	−688.98	−245.67	−672.33	−487.59	−108.07	−382.04	−523.26	−560.76
Pep8	−476.00	−126.82	−696.78	−344.27	−679.03	−549.43	−99.45	−468.22	−510.97	−562.79
Pep9	−431.19	−122.1	−697.63	−286.93	−695.46	−538.61	−76.07	−371.04	−566.23	−582.03
Pep10	−402.40	−192.73	−733.60	−328.82	−666.53	−551.25	−129.96	−399.71	−575.76	**−695.61**
Pep11	−509.29	−109.09	−676.82	−352.06	−666.86	−588.69	−128.58	−441.59	−577.13	−647.20
Pep12	−407.00	−183.00	−694.56	−277.49	−689.17	**−621.01**	−134.90	−381.26	−543.31	−580.70
Pep13	**−521.58**	−68.73	−675.37	−345.52	−662.13	−529.89	−77.80	−482.36	−529.25	−550.00
Pep14	−435.13	−114.38	−687.94	−300.80	−653.34	−522.23	−89.09	−409.07	−564.50	−545.59
Pep15	−406.22	−111.05	−644.25	−286.49	−630.63	−579.12	−57.24	−387.08	−523.27	−592.09
Pep16	−429.93	−60.14	−542.93	−353.55	−546.86	−453.89	−56.18	−431.47	−445.11	−510.73
Pep17	−466.07	−69.89	−710.98	**−378.87**	−660.04	−477.68	−59.56	−446.68	−559.88	−509.83
Pep18	−429.93	−60.14	−542.93	−353.55	−546.86	−453.89	−56.18	−431.47	−445.11	−483.05
Pep19	−473.68	−98.50	−648.60	−354.31	−657.44	−535.19	−71.67	−447.10	−524.21	−629.49
Pep20	−441.87	−80.54	**−747.85**	−203.43	**−728.14**	−555.01	−116.62	−395.20	−550.00	−600.97

In the breast cancer, 1i4f, 1a07 MHC I, MHC II 1bx2, 1fv1, 1h15, 1a6a, 1j8h, 1jk8, 31qz and TCR 3mff. Pep5, DIGTLYLIFGAIAGVMGTCFSVLIR with MHC class I with binding global energy −210.97. Pep13, FSPTSGGCRGIRCSANIVGECPAQ with MHC class I with binding global energy −521.58. Pep3, VSAAINAGLVIDPLLQCLNEWNGAPIPIC with MHC class II binding global energy 1jk8 with −566.98 and 3Iq2 with −587.34. Pep20, GAIIGNRKIKLQPHIIIRIID with MHC class II with ng global energy of 1bx2 and 1hI5 is −747.85 and −728.14 kj/Mol, respectively. Finally, Pep10, LIRQHVASNIGIAKSKIREKEPIV with 3mff TCR has the highest binding energy −695.61 compared with Pep20 GAIIGNRKIKLQPHIIIRIID, which is −600.97 kj/Mol, is the second-highest. Binding interaction site for each receptors represents in terms of hydrogen binding HLA-A0201, HLA-DR2, DRB50101, HLA-DR3, HLA-DR4, HLA-DQ8 and HLA-DP2 MHC receptor. Pep10 also have moderate binding affinity with MHC classes I and II but further it achieves the highest binding affinity with TCR, i.e., −695.61 which is highest among all other peptides. Fig. 3A shows the interaction sites between peptide 10 with TCR are B: GLN 139: HE 21 – LEU 11: O, B: THR 138: O – GLY 1: H 1, C: PRO 90: O – GLY 5: H and C: GLN 45: O – ARG 7: HH 21. The binding interaction site for each receptor represents hydrogen binding HLA-A0201, HLA-DR2, DRB50101, HLA-DR3, HLA-DR4, HLA-DQ8, and HLA-DP2 MHC receptor. Pep10 also has a moderate binding affinity with MHC Classes I and II, but it achieves the highest binding affinity with TCR, i.e., −695.61, the highest among all other peptides. Overall, all the receptors have the highest binding energy with the peptide 20 and have the highest binding energy of −747.85 with HLA-DR

2 and −728.14 kj/Mol with DRB50101. Further, Pep20 docked to TCR and achieved a −600.97 kj/Mol global energy level, the second-highest energy among all other predicted peptides shown in Fig. 3B. The interaction sites between peptide 10 with TCR are GLN 123 : HE Z1 – PRO 22: O, LYS 140 : HZ 1 – GLU 21 : OE 1, LYS 140 : HZ 3 – GLU 21 : OE 1, LYS 140 : HZ 3 – GLU 21 : OE 2, PRO 174 : O – ARG 3 : HH 12, HIS 167 : NE2 – ASN 9 : HD 22, GLN 139: OE 1 – SER 15: HG and GLN 123 : OE 1 – ARG 18 : HH 11, ARG 18 (see Appendix Tables 1A and 1B in the supplemental material).
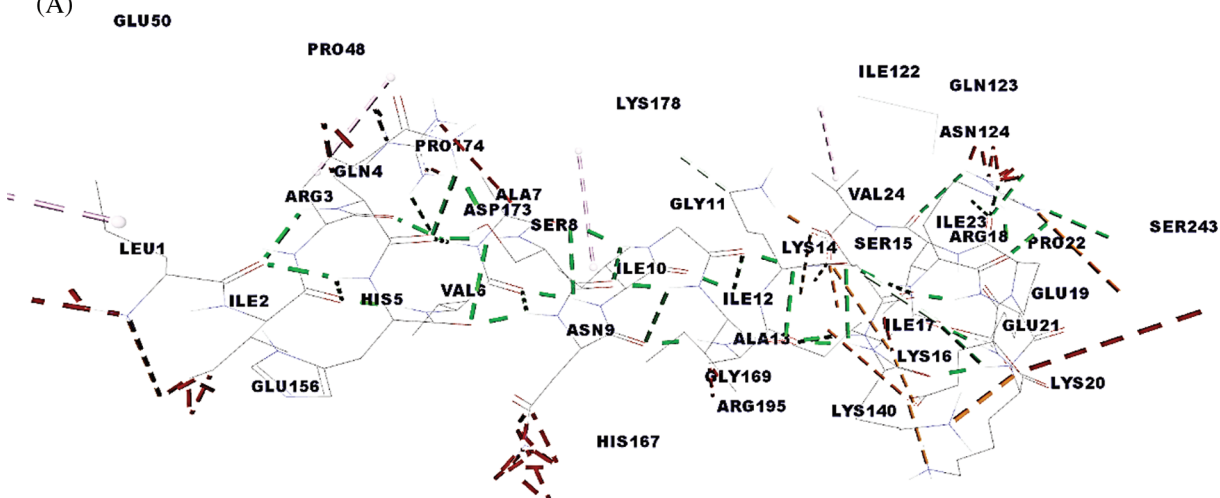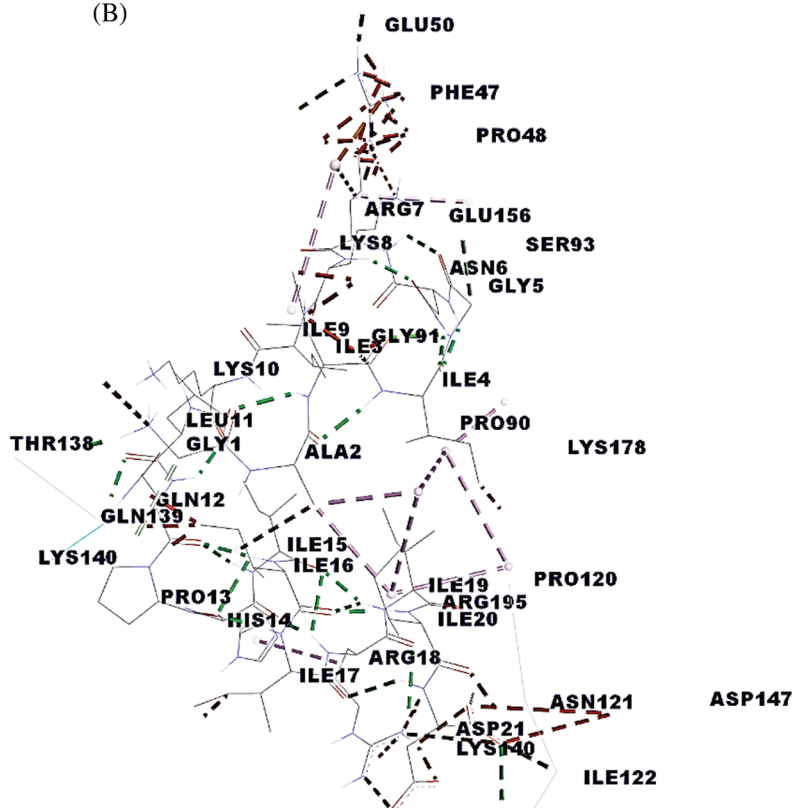

Figure 3: Docking complex with interaction (A) Pep10 with TCR (B) Pep20 with TCR

These findings indicated that Pep10 and 20 could strongly bind to MHC and TCR. This interaction leads to only four hydrogen bonds, which have effective binding. Thus, it could generate an effective immune response against HER 2 receptors to prevent the growth of human breast cancer cells.

4 Conclusion

Our knowledge of the significance of peptides may be used directly as a cytotoxic agent or as a transporter of cytotoxic chemicals and radioisotopes to cancer cells. Peptide-based hormone treatment has been thoroughly studied and treated with breast and prostate cancers. In this research work, two peptides, Pep10 and Pep20, have been identified as having a significant impact on stopping cancer progression. The Pep10 has moderate binding energy with eight hydrogen bonds, and Pep20 achieves the highest binding energy with four hydrogen bonds on the MHC classes I and II. Both peptides are harmless, according to computational methods based on experiments. It could ultimately serve as a novel essential therapeutic target as a cancer peptide vaccine in the future. This method might aid researchers in achieving better disease regulation and predicting treatment targets of disease.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Garrett, D., Yoder, L. H. (2007). An overview of stem cell transplant as a treatment for cancer. *MedSurg Nursing, 16(3),* 183.
2. Garrett, J. T. (2007). *Peptide-based B-cell epitope vaccines targeting HER-2/neu (Ph.D. Thesis)*. The Ohio State University.
3. Calabrich, A., Fernandes, G. D. S., Katz, A. (2008). Trastuzumab: Mechanisms of resistance and therapeutic opportunities. *Oncology, 22(11),* 1250–1258.
4. Xiang, B., Snook, A. E., Magee, M. S., Waldman, S. A. (2013). Colorectal cancer immunotherapy. *Discovery Medicine, 15(84),* 301–308.
5. Li, W., Joshi, M. D., Singhania, S., Ramsey, K. H., Murthy, A. K. (2014). Peptide vaccine: Progress and challenges. *Vaccines, 2(3),* 515–536.
6. Reche, P. A., Fernandez-Caldas, E., Flower, D. R., Fridkis-Hareli, M., Hoshino, Y. (2014). Peptide-based immunotherapeutics and vaccines. *Journal of Immunology Research, 2014,* 256784. DOI 10.1155/2014/256784.
7. Slingluff Jr., C. L. (2011). The present and future of peptide vaccines for cancer: Single or multiple, long or short, alone or in combination? *Cancer Journal, 17(5),* 343–350.
8. Zhang, X. W. (2013). A combination of epitope prediction and molecular docking allows for good identification of MHC class I restricted T-cell epitopes. *Computational Biology and Chemistry, 45(23–24),* 30–35. DOI 10.1016/j.compbiolchem.2013.03.003.
9. Bian, H., Reidhaar-Olson, J. F., Hammer, J. (2003). The use of bioinformatics for identifying class II-restricted T-cell epitopes. *Methods, 29(3),* 299–309. DOI 10.1016/S1046-2023(02)00352-3.
10. Martin, W., Sbai, H., de Groot, A. S. (2003). Bioinformatics tools for identifying class I-restricted epitopes. *Methods, 29(3),* 289–298. DOI 10.1016/S1046-2023(02)00351-1.
11. Garboczi, D. N., Ghosh, P., Utz, U., Fan, Q. R., Biddison, W. E. et al. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature, 384(6605),* 134–141. DOI 10.1038/384134a0.
12. Jardetzky, T. S., Brown, J. H., Gorga, J. C., Stern, L. J., Urban, R. G. et al. (1996). Crystallographic analysis of endogenous peptides associated with HLA-DR1 suggests a common, polyproline II-like conformation for bound peptides. *Proceedings of the National Academy of Sciences, 93(2),* 734–738. DOI 10.1073/pnas.93.2.734.

13. Tong, J. C., Tan, T. W., Ranganathan, S. (2004). Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Science: A Publication of the Protein Society, 13(9),* 2523–2532. DOI 10.1110/(ISSN)1469-896X.

14. Dakappagari, N. K., Douglas, D. B., Triozzi, P. L., Stevens, V. C., Kaumaya, P. T. (2000). Prevention of mammary tumors with a chimeric HER-2 B-cell epitope peptide vaccine. *Cancer Research, 60(14),* 3782–3789.

15. Manijeh, M., Mehrnaz, K., Violaine, M., Hassan, M., Abbas, J. et al. (2013). In silico design of discontinuous peptides representative of B and T-cell epitopes from HER2-ECD as potential novel cancer peptide vaccines. *Asian Pacific Journal of Cancer Prevention, 14(10),* 5973–5981. DOI 10.7314/APJCP.2013.14.10.5973.

16. Wang, B., Kaumaya, P. T., Cohn, D. E. (2010). Immunization with synthetic VEGF peptides in ovarian cancer. *Gynecologic Oncology, 119(3),* 564–570. DOI 10.1016/j.ygyno.2010.07.037.

17. Flower, D. R. (2003). Databases and data mining for computational vaccinology. *Current Opinion in Drug Discovery & Development, 6(3),* 396–400.

18. Patronov, A., Dimitrov, I., Flower, D. R., Doytchinova, I. (2011). Peptide binding prediction for the human class II MHC allele HLA-DP2: A molecular docking approach. *BMC Structural Biology, 11(1),* 1–10. DOI 10.1186/1472-6807-11-32.

19. Young, D. C. (2009). *Computational drug design: A guide for computational and medicinal chemists*. USA: John Wiley & Sons. DOI 10.1002/9780470451854.

20. D'Incalci, M., Steward, W. P., Gescher, A. J. (2005). Use of cancer chemopreventive phytochemicals as antineoplastic agents. *The Lancet Oncology, 6(11),* 899–904. DOI 10.1016/S1470-2045(05)70425-3.

21. Maupetit, J., Derreumaux, P., Tuffery, P. (2009). PEP-FOLD: An online resource for *de novo* peptide structure prediction. *Nucleic Acids Research, 37(suppl_2),* W498–W503. DOI 10.1093/nar/gkp323.

22. Guex, N., Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis, 18(15),* 2714–2723. DOI 10.1002/(ISSN)1522-2683.

23. Scott, W. R., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R. et al. (1999). The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A, 103(19),* 3596–3607. DOI 10.1021/jp984217f.

24. Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R. et al. (2013). *In silico* approach for predicting toxicity of peptides and proteins. *PLoS One, 8(9),* e73957. DOI 10.1371/journal.pone.0073957.

25. Huang, H. J., Lee, K. J., Yu, H. W., Chen, H. Y., Tsai, F. J. et al. (2010). A novel strategy for designing the selective PPAR agonist by the "Sum of Activity" model. *Journal of Biomolecular Structure and Dynamics, 28(2),* 187–200. DOI 10.1080/07391102.2010.10507352.

26. Accelrys Software Inc. (2009). *Discovery studio 2.5*. San Diego: Accelrys Software Inc.

Appendix Table 1: Docking complex with interaction (A) Pep10 with TCR bonding of amino acid (B) Pep20 with TCR bonding of amino acid

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
(A)							
:ARG7:HH12 - C: GLU50:OE1	2.71155	Hydrogen Bond; Electrostatic	Salt bridge; Attractive charge	:ARG7: HH12	H-Donor; Positive	C:GLU50: OE1	H-Acceptor; Negative
:ARG18:HH11 - : ASP21:O	3.03632	Hydrogen Bond; Electrostatic	Salt bridge; Attractive charge	:ARG18: HH11	H-Donor; Positive	:ASP21:O	H-Acceptor; Negative
:ARG18:HH11 - : ASP21:OD2	2.10191	Hydrogen Bond; Electrostatic	Salt bridge; Attractive charge	:ARG18: HH11	H-Donor; Positive	:ASP21: OD2	H-Acceptor; Negative
D:ARG195:NH2 - : ASP21:O	4.84956	Electrostatic	Attractive charge	D:ARG195: NH2	Positive	:ASP21:O	Negative
:LYS8:NZ - D:GLU156: OE2	4.27326	Electrostatic	Attractive charge	:LYS8:NZ	Positive	D:GLU156: OE2	Negative
:ARG18:NH1 - :ASP21: OXT	5.35163	Electrostatic	Attractive charge	:ARG18: NH1	Positive	:ASP21: OXT	Negative
:ARG18:NH2 - :ASP21: OD1	3.08868	Electrostatic	Attractive charge	:ARG18: NH2	Positive	:ASP21: OD1	Negative
B:GLN139:HE21 - : LEU11:O	2.01025	Hydrogen Bond	Conventional hydrogen bond	B:GLN139: HE21	H-Donor	:LEU11:O	H-Acceptor
C:GLN123:HN - : ASP21:OXT	2.6577	Hydrogen Bond	Conventional hydrogen bond	C:GLN123: HN	H-Donor	:ASP21: OXT	H-Acceptor
:GLY1:HT1 - B: THR138:O	1.94776	Hydrogen Bond	Conventional hydrogen bond	:GLY1:HT1	H-Donor	B:THR138: O	H-Acceptor
:ILE3:HN - :GLY1:O	2.2139	Hydrogen Bond	Conventional hydrogen bond	:ILE3:HN	H-Donor	:GLY1:O	H-Acceptor
:ILE4:HN - :ALA2:O	2.12503	Hydrogen Bond	Conventional hydrogen bond	:ILE4:HN	H-Donor	:ALA2:O	H-Acceptor
:GLY5:HN - C: PRO90:O	2.2299	Hydrogen Bond	Conventional hydrogen bond	:GLY5:HN	H-Donor	C:PRO90: O	H-Acceptor
:GLY5:HN - :ILE3:O	2.12522	Hydrogen Bond	Conventional hydrogen bond	:GLY5:HN	H-Donor	:ILE3:O	H-Acceptor
:ASN6:HN - :ILE4:O	2.0869	Hydrogen Bond	Conventional hydrogen bond	:ASN6:HN	H-Donor	:ILE4:O	H-Acceptor
:ARG7:HN - :GLY5:O	2.73665	Hydrogen Bond	Conventional hydrogen bond	:ARG7:HN	H-Donor	:GLY5:O	H-Acceptor
:ARG7:HH11 - C: GLN45:O	2.53695	Hydrogen Bond	Conventional hydrogen bond	:ARG7: HH11	H-Donor	C:GLN45: O	H-Acceptor
:LYS8:HN - :ASN6:OD1	2.15072	Hydrogen Bond	Conventional hydrogen bond	:LYS8:HN	H-Donor	:ASN6: OD1	H-Acceptor
:GLN12:HN - :LYS10:O	2.08151	Hydrogen Bond	Conventional hydrogen bond	:GLN12:HN	H-Donor	:LYS10:O	H-Acceptor
:ILE15:HN - :GLN12:O	2.40051	Hydrogen Bond	Conventional hydrogen bond	:ILE15:HN	H-Donor	:GLN12:O	H-Acceptor
:ILE16:HN - :GLN12:O	2.25961	Hydrogen Bond	Conventional hydrogen bond	:ILE16:HN	H-Donor	:GLN12:O	H-Acceptor
:ILE16:HN - :PRO13:O	2.69335	Hydrogen Bond	Conventional hydrogen bond	:ILE16:HN	H-Donor	:PRO13:O	H-Acceptor
:ILE17:HN - :PRO13:O	1.92206	Hydrogen Bond	Conventional hydrogen bond	:ILE17:HN	H-Donor	:PRO13:O	H-Acceptor

(Continued)

Appendix Table 1 (continued)

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
:ARG18:HN - :HIS14:O	1.97778	Hydrogen Bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:HIS14:O	H-Acceptor
:ARG18:HN - :ILE15:O	2.67954	Hydrogen Bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:ILE15:O	H-Acceptor
:ARG18:HH12 - :ARG18:O	2.81755	Hydrogen Bond	Conventional hydrogen bond	:ARG18:HH12	H-Donor	:ARG18:O	H-Acceptor
:ILE19:HN - :ILE15:O	2.08178	Hydrogen Bond	Conventional hydrogen bond	:ILE19:HN	H-Donor	:ILE15:O	H-Acceptor
:ILE19:HN - :ILE16:O	2.49149	Hydrogen Bond	Conventional hydrogen bond	:ILE19:HN	H-Donor	:ILE16:O	H-Acceptor
:ILE20:HN - :ILE16:O	1.93209	Hydrogen Bond	Conventional hydrogen bond	:ILE20:HN	H-Donor	:ILE16:O	H-Acceptor
:ASP21:HN - :ILE17:O	2.9193	Hydrogen Bond	Conventional hydrogen bond	:ASP21:HN	H-Donor	:ILE17:O	H-Acceptor
C:ILE122:CA - :ASP21:OXT	3.46316	Hydrogen Bond	Carbon hydrogen bond	C:ILE122:CA	H-Donor	:ASP21:OXT	H-Acceptor
:GLY1:CA - B:GLN139:O	3.0816	Hydrogen Bond	Carbon hydrogen bond	:GLY1:CA	H-Donor	B:GLN139:O	H-Acceptor
:GLY5:CA - C:SER93:O	3.63125	Hydrogen Bond	Carbon hydrogen bond	:GLY5:CA	H-Donor	C:SER93:O	H-Acceptor
:ARG7:NH2 - C:PHE47	2.77047	Electrostatic	Pi-cation	:ARG7:NH2	Positive	C:PHE47	Pi-Orbitals
C:PRO90 - :ILE4	4.79367	Hydrophobic	Alkyl	C:PRO90	Alkyl	:ILE4	Alkyl
C:PRO90 - :ILE20	4.64613	Hydrophobic	Alkyl	C:PRO90	Alkyl	:ILE20	Alkyl
C:PRO120 - :ILE4	4.87024	Hydrophobic	Alkyl	C:PRO120	Alkyl	:ILE4	Alkyl
C:PRO120 - :ILE20	4.20855	Hydrophobic	Alkyl	C:PRO120	Alkyl	:ILE20	Alkyl
C:LYS178 - :ILE4	3.99255	Hydrophobic	Alkyl	C:LYS178	Alkyl	:ILE4	Alkyl
:ALA2 - C:PRO90	3.73315	Hydrophobic	Alkyl	:ALA2	Alkyl	C:PRO90	Alkyl
:ALA2 - :ILE16	4.74294	Hydrophobic	Alkyl	:ALA2	Alkyl	:ILE16	Alkyl
:ALA2 - :ILE20	4.39053	Hydrophobic	Alkyl	:ALA2	Alkyl	:ILE20	Alkyl
:ARG7 - C:PRO48	4.37033	Hydrophobic	Alkyl	:ARG7	Alkyl	C:PRO48	Alkyl
C:PHE47 - :ILE3	5.14388	Hydrophobic	Pi-Alkyl	C:PHE47	Pi-orbitals	:ILE3	Alkyl
C:PHE47 - :ARG7	4.49233	Hydrophobic	Pi-Alkyl	C:PHE47	Pi-orbitals	:ARG7	Alkyl
:HIS14 - :ARG18	4.64901	Hydrophobic	Pi-Alkyl	:HIS14	Pi-orbitals	:ARG18	Alkyl
(B)							
D:ARG195:HH11 - :VAL24:O	2.52106	Hydrogen bond; Electrostatic	Salt bridge; Attractive charge	D:ARG195:HH11	H-Donor; Positive	:VAL24:O	H-Acceptor; Negative
D:ARG195:HH21 - :VAL24:O	2.64985	Hydrogen bond; Electrostatic	Salt bridge; Attractive charge	D:ARG195:HH21	H-Donor; Positive	:VAL24:O	H-Acceptor; Negative
:LYS14:HZ3 - :VAL24:OXT	1.99951	Hydrogen bond; Electrostatic	Salt bridge; Attractive charge	:LYS14:HZ3	H-Donor; Positive	:VAL24:OXT	H-Acceptor; Negative
D:LYS140:NZ - :GLU21:OE1	2.50493	Electrostatic	Attractive charge	D:LYS140:NZ	Positive	:GLU21:OE1	Negative
D:LYS140:NZ - :VAL24:O	5.3876	Electrostatic	Attractive charge	D:LYS140:NZ	Positive	:VAL24:O	Negative
D:ARG195:NH1 - :GLU21:OE2	5.23935	Electrostatic	Attractive charge	D:ARG195:NH1	Positive	:GLU21:OE2	Negative
D:ARG195:NH2 - :VAL24:OXT	5.1305	Electrostatic	Attractive charge	D:ARG195:NH2	Positive	:VAL24:OXT	Negative

(Continued)

Appendix Table 1 (continued)

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
:LEU1:N - D:GLU156:OE2	4.97065	Electrostatic	Attractive charge	:LEU1:N	Positive	D:GLU156:OE2	Negative
:LYS16:NZ - :GLU19:OE2	2.96463	Electrostatic	Attractive charge	:LYS16:NZ	Positive	:GLU19:OE2	Negative
:ARG18:NH2 - B:SER243:O	5.08644	Electrostatic	Attractive charge	:ARG18:NH2	Positive	B:SER243:O	Negative
:LYS20:NZ - :GLU21:OE2	3.11508	Electrostatic	Attractive charge	:LYS20:NZ	Positive	:GLU21:OE2	Negative
C:GLN123:HE21 - :PRO22:O	2.47629	Hydrogen bond	Conventional hydrogen bond	C:GLN123:HE21	H-Donor	:PRO22:O	H-Acceptor
C:ASN124:HD21 - :PRO22:O	3.00104	Hydrogen bond	Conventional hydrogen bond	C:ASN124:HD21	H-Donor	:PRO22:O	H-Acceptor
:ARG3:HH12 - D:ASP173:O	2.83228	Hydrogen bond	Conventional hydrogen bond	:ARG3:HH12	H-Donor	D:ASP173:O	H-Acceptor
:GLN4:HN - :LEU1:O	2.16474	Hydrogen bond	Conventional hydrogen bond	:GLN4:HN	H-Donor	:LEU1:O	H-Acceptor
:GLN4:HE21 - :GLN4:O	2.96028	Hydrogen bond	Conventional hydrogen bond	:GLN4:HE21	H-Donor	:GLN4:O	H-Acceptor
:GLN4:HE21 - :SER8:OG	1.84217	Hydrogen bond	Conventional hydrogen bond	:GLN4:HE21	H-Donor	:SER8:OG	H-Acceptor
:HIS5:HN - :LEU1:O	2.16847	Hydrogen bond	Conventional hydrogen bond	:HIS5:HN	H-Donor	:LEU1:O	H-Acceptor
:HIS5:HN - :ILE2:O	2.83972	Hydrogen bond	Conventional hydrogen bond	:HIS5:HN	H-Donor	:ILE2:O	H-Acceptor
:VAL6:HN - :ILE2:O	1.89431	Hydrogen bond	Conventional hydrogen bond	:VAL6:HN	H-Donor	:ILE2:O	H-Acceptor
:ALA7:HN - :ARG3:O	2.07957	Hydrogen bond	Conventional hydrogen bond	:ALA7:HN	H-Donor	:ARG3:O	H-Acceptor
:ALA7:HN - :GLN4:O	2.77744	Hydrogen bond	Conventional hydrogen bond	:ALA7:HN	H-Donor	:GLN4:O	H-Acceptor
:SER8:HN - :GLN4:O	1.72702	Hydrogen bond	Conventional hydrogen bond	:SER8:HN	H-Donor	:GLN4:O	H-Acceptor
:SER8:HN - :HIS5:O	2.89215	Hydrogen bond	Conventional hydrogen bond	:SER8:HN	H-Donor	:HIS5:O	H-Acceptor
:ASN9:HN - :HIS5:O	1.87443	Hydrogen bond	Conventional hydrogen bond	:ASN9:HN	H-Donor	:HIS5:O	H-Acceptor
:ASN9:HN - :VAL6:O	2.88727	Hydrogen bond	Conventional hydrogen bond	:ASN9:HN	H-Donor	:VAL6:O	H-Acceptor
:ILE10:HN - :VAL6:O	1.91642	Hydrogen bond	Conventional hydrogen bond	:ILE10:HN	H-Donor	:VAL6:O	H-Acceptor
:ILE10:HN - :ALA7:O	2.90224	Hydrogen bond	Conventional hydrogen bond	:ILE10:HN	H-Donor	:ALA7:O	H-Acceptor
:GLY11:HN - :ALA7:O	1.93857	Hydrogen bond	Conventional hydrogen bond	:GLY11:HN	H-Donor	:ALA7:O	H-Acceptor
:GLY11:HN - :SER8:O	2.72278	Hydrogen bond	Conventional hydrogen bond	:GLY11:HN	H-Donor	:SER8:O	H-Acceptor
:ILE12:HN - :SER8:O	1.85618	Hydrogen bond	Conventional hydrogen bond	:ILE12:HN	H-Donor	:SER8:O	H-Acceptor
:ILE12:HN - :ASN9:O	2.98726	Hydrogen bond	Conventional hydrogen bond	:ILE12:HN	H-Donor	:ASN9:O	H-Acceptor

(Continued)

Appendix Table 1 (continued)

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
:ALA13:HN - :ASN9:O	1.86548	Hydrogen bond	Conventional hydrogen bond	:ALA13:HN	H-Donor	:ASN9:O	H-Acceptor
:LYS14:HN - :ILE10:O	1.84719	Hydrogen bond	Conventional hydrogen bond	:LYS14:HN	H-Donor	:ILE10:O	H-Acceptor
:LYS14:HN - :GLY11:O	2.69889	Hydrogen bond	Conventional hydrogen bond	:LYS14:HN	H-Donor	:GLY11:O	H-Acceptor
:SER15:HN - :GLY11:O	1.90957	Hydrogen bond	Conventional hydrogen bond	:SER15:HN	H-Donor	:GLY11:O	H-Acceptor
:SER15:HN - :ILE12:O	2.78525	Hydrogen bond	Conventional hydrogen bond	:SER15:HN	H-Donor	:ILE12:O	H-Acceptor
:LYS16:HN - :ILE12:O	1.94709	Hydrogen bond	Conventional hydrogen bond	:LYS16:HN	H-Donor	:ILE12:O	H-Acceptor
:LYS16:HN - :ALA13:O	2.80139	Hydrogen bond	Conventional hydrogen bond	:LYS16:HN	H-Donor	:ALA13:O	H-Acceptor
:ILE17:HN - :ALA13:O	1.94714	Hydrogen bond	Conventional hydrogen bond	:ILE17:HN	H-Donor	:ALA13:O	H-Acceptor
:ILE17:HN - :LYS14:O	3.00526	Hydrogen bond	Conventional hydrogen bond	:ILE17:HN	H-Donor	:LYS14:O	H-Acceptor
:ARG18:HN - :LYS14:O	1.90732	Hydrogen bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:LYS14:O	H-Acceptor
:ARG18:HN - :SER15:O	2.68651	Hydrogen bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:SER15:O	H-Acceptor
:ARG18:HE - :ARG18:O	1.93021	Hydrogen bond	Conventional hydrogen bond	:ARG18:HE	H-Donor	:ARG18:O	H-Acceptor
:ARG18:HH11 - :PRO22:O	2.31411	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH11	H-Donor	:PRO22:O	H-Acceptor
:ARG18:HH12 - :ILE23:O	2.66519	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH12	H-Donor	:ILE23:O	H-Acceptor
:ARG18:HH21 - B:SER243:OG	2.79253	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH21	H-Donor	B:SER243:OG	H-Acceptor
:ARG18:HH22 - :ARG18:O	2.87732	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH22	H-Donor	:ARG18:O	H-Acceptor
:GLU19:HN - :SER15:O	1.81301	Hydrogen bond	Conventional hydrogen bond	:GLU19:HN	H-Donor	:SER15:O	H-Acceptor
:LYS20:HN - :LYS16:O	1.99565	Hydrogen bond	Conventional hydrogen bond	:LYS20:HN	H-Donor	:LYS16:O	H-Acceptor
:LYS20:HN - :ILE17:O	3.06892	Hydrogen bond	Conventional hydrogen bond	:LYS20:HN	H-Donor	:ILE17:O	H-Acceptor
:GLU21:HN - :ILE17:O	1.79476	Hydrogen bond	Conventional hydrogen bond	:GLU21:HN	H-Donor	:ILE17:O	H-Acceptor
:LYS14:CE - C:LYS178:O	2.87082	Hydrogen bond	Carbon hydrogen bond	:LYS14:CE	H-Donor	C:LYS178:O	H-Acceptor
:SER15:CB - :GLU19:OE1	3.61548	Hydrogen bond	Carbon hydrogen bond	:SER15:CB	H-Donor	:GLU19:OE1	H-Acceptor
C:LYS178 - :ILE10	5.09183	Hydrophobic	Alkyl	C:LYS178	Alkyl	:ILE10	Alkyl
D:PRO107 - :LEU1	4.92164	Hydrophobic	Alkyl	D:PRO107	Alkyl	:LEU1	Alkyl
:ARG3 - C:PRO48	4.287	Hydrophobic	Alkyl	:ARG3	Alkyl	C:PRO48	Alkyl
:VAL24 - C:ILE122	4.75696	Hydrophobic	Alkyl	:VAL24	Alkyl	C:ILE122	Alkyl

(Continued)

Appendix Table 1 (continued)

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
D:ARG195:HH11 - :VAL24:O	2.52106	Hydrogen bond; Electrostatic	Salt bridge	D:ARG195:HH11	H-Donor	:VAL24:O	H-Acceptor
D:ARG195:HH21 - :VAL24:O	2.64985	Hydrogen bond; Electrostatic	Salt bridge	D:ARG195:HH21	H-Donor	:VAL24:O	H-Acceptor
:LYS14:HZ3 - :VAL24:OXT	1.99951	Hydrogen bond; Electrostatic	Salt bridge	:LYS14:HZ3	H-Donor	:VAL24:OXT	H-Acceptor
C:GLN123:HE21 - :PRO22:O	2.47629	Hydrogen bond	Conventional hydrogen bond	C:GLN123:HE21	H-Donor	:PRO22:O	H-Acceptor
C:ASN124:HD21 - :PRO22:O	3.00104	Hydrogen bond	Conventional hydrogen bond	C:ASN124:HD21	H-Donor	:PRO22:O	H-Acceptor
D:ARG195:HN - D:GLY169:O	1.88274	Hydrogen bond	Conventional hydrogen bond	D:ARG195:HN	H-Donor	D:GLY169:O	H-Acceptor
:ARG3:HH12 - D:ASP173:O	2.83228	Hydrogen bond	Conventional hydrogen bond	:ARG3:HH12	H-Donor	D:ASP173:O	H-Acceptor
:GLN4:HN - :LEU1:O	2.16474	Hydrogen bond	Conventional hydrogen bond	:GLN4:HN	H-Donor	:LEU1:O	H-Acceptor
:GLN4:HE21 - :GLN4:O	2.96028	Hydrogen bond	Conventional hydrogen bond	:GLN4:HE21	H-Donor	:GLN4:O	H-Acceptor
:GLN4:HE21 - :SER8:OG	1.84217	Hydrogen bond	Conventional hydrogen bond	:GLN4:HE21	H-Donor	:SER8:OG	H-Acceptor
:HIS5:HN - :LEU1:O	2.16847	Hydrogen bond	Conventional hydrogen bond	:HIS5:HN	H-Donor	:LEU1:O	H-Acceptor
:HIS5:HN - :ILE2:O	2.83972	Hydrogen bond	Conventional hydrogen bond	:HIS5:HN	H-Donor	:ILE2:O	H-Acceptor
:VAL6:HN - :ILE2:O	1.89431	Hydrogen bond	Conventional hydrogen bond	:VAL6:HN	H-Donor	:ILE2:O	H-Acceptor
:ALA7:HN - :ARG3:O	2.07957	Hydrogen bond	Conventional hydrogen bond	:ALA7:HN	H-Donor	:ARG3:O	H-Acceptor
:ALA7:HN - :GLN4:O	2.77744	Hydrogen bond	Conventional hydrogen bond	:ALA7:HN	H-Donor	:GLN4:O	H-Acceptor
:SER8:HN - :GLN4:O	1.72702	Hydrogen bond	Conventional hydrogen bond	:SER8:HN	H-Donor	:GLN4:O	H-Acceptor
:SER8:HN - :HIS5:O	2.89215	Hydrogen bond	Conventional hydrogen bond	:SER8:HN	H-Donor	:HIS5:O	H-Acceptor
:ASN9:HN - :HIS5:O	1.87443	Hydrogen bond	Conventional hydrogen bond	:ASN9:HN	H-Donor	:HIS5:O	H-Acceptor
:ASN9:HN - :VAL6:O	2.88727	Hydrogen bond	Conventional hydrogen bond	:ASN9:HN	H-Donor	:VAL6:O	H-Acceptor
:ILE10:HN - :VAL6:O	1.91642	Hydrogen bond	Conventional hydrogen bond	:ILE10:HN	H-Donor	:VAL6:O	H-Acceptor
:ILE10:HN - :ALA7:O	2.90224	Hydrogen bond	Conventional hydrogen bond	:ILE10:HN	H-Donor	:ALA7:O	H-Acceptor
:GLY11:HN - :ALA7:O	1.93857	Hydrogen bond	Conventional hydrogen bond	:GLY11:HN	H-Donor	:ALA7:O	H-Acceptor
:GLY11:HN - :SER8:O	2.72278	Hydrogen bond	Conventional hydrogen bond	:GLY11:HN	H-Donor	:SER8:O	H-Acceptor
:ILE12:HN - :SER8:O	1.85618	Hydrogen bond	Conventional hydrogen bond	:ILE12:HN	H-Donor	:SER8:O	H-Acceptor

(Continued)

Appendix Table 1 (continued)

Name	Distance	Category	Types	From	From chemistry	To	To chemistry
:ILE12:HN - :ASN9:O	2.98726	Hydrogen bond	Conventional hydrogen bond	:ILE12:HN	H-Donor	:ASN9:O	H-Acceptor
:ALA13:HN - :ASN9:O	1.86548	Hydrogen bond	Conventional hydrogen bond	:ALA13:HN	H-Donor	:ASN9:O	H-Acceptor
:LYS14:HN - :ILE10:O	1.84719	Hydrogen bond	Conventional hydrogen bond	:LYS14:HN	H-Donor	:ILE10:O	H-Acceptor
:LYS14:HN - :GLY11:O	2.69889	Hydrogen bond	Conventional hydrogen bond	:LYS14:HN	H-Donor	:GLY11:O	H-Acceptor
:SER15:HN - :GLY11:O	1.90957	Hydrogen bond	Conventional hydrogen bond	:SER15:HN	H-Donor	:GLY11:O	H-Acceptor
:SER15:HN - :ILE12:O	2.78525	Hydrogen bond	Conventional hydrogen bond	:SER15:HN	H-Donor	:ILE12:O	H-Acceptor
:LYS16:HN - :ILE12:O	1.94709	Hydrogen bond	Conventional hydrogen bond	:LYS16:HN	H-Donor	:ILE12:O	H-Acceptor
:LYS16:HN - :ALA13:O	2.80139	Hydrogen bond	Conventional hydrogen bond	:LYS16:HN	H-Donor	:ALA13:O	H-Acceptor
:ILE17:HN - :ALA13:O	1.94714	Hydrogen bond	Conventional hydrogen bond	:ILE17:HN	H-Donor	:ALA13:O	H-Acceptor
:ILE17:HN - :LYS14:O	3.00526	Hydrogen bond	Conventional hydrogen bond	:ILE17:HN	H-Donor	:LYS14:O	H-Acceptor
:ARG18:HN - :LYS14:O	1.90732	Hydrogen bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:LYS14:O	H-Acceptor
:ARG18:HN - :SER15:O	2.68651	Hydrogen bond	Conventional hydrogen bond	:ARG18:HN	H-Donor	:SER15:O	H-Acceptor
:ARG18:HE - :ARG18:O	1.93021	Hydrogen bond	Conventional hydrogen bond	:ARG18:HE	H-Donor	:ARG18:O	H-Acceptor
:ARG18:HH11 - :PRO22:O	2.31411	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH11	H-Donor	:PRO22:O	H-Acceptor
:ARG18:HH12 - C:GLN123:OE1	1.57376	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH12	H-Donor	C:GLN123:OE1	H-Acceptor
:ARG18:HH12 - :ILE23:O	2.66519	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH12	H-Donor	:ILE23:O	H-Acceptor
:ARG18:HH21 - B:SER243:OG	2.79253	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH21	H-Donor	B:SER243:OG	H-Acceptor
:ARG18:HH22 - :ARG18:O	2.87732	Hydrogen bond	Conventional hydrogen bond	:ARG18:HH22	H-Donor	:ARG18:O	H-Acceptor
:GLU19:HN - :SER15:O	1.81301	Hydrogen bond	Conventional hydrogen bond	:GLU19:HN	H-Donor	:SER15:O	H-Acceptor
:LYS20:HN - :LYS16:O	1.99565	Hydrogen bond	Conventional hydrogen bond	:LYS20:HN	H-Donor	:LYS16:O	H-Acceptor
:LYS20:HN - :ILE17:O	3.06892	Hydrogen bond	Conventional hydrogen bond	:LYS20:HN	H-Donor	:ILE17:O	H-Acceptor
:GLU21:HN - :ILE17:O	1.79476	Hydrogen bond	Conventional hydrogen bond	:GLU21:HN	H-Donor	:ILE17:O	H-Acceptor
:LYS14:CE - C:LYS178:O	2.87082	Hydrogen bond	Carbon hydrogen bond	:LYS14:CE	H-Donor	C:LYS178:O	H-Acceptor
:SER15:CB - :GLU19:OE1	3.61548	Hydrogen bond	Carbon hydrogen bond	:SER15:CB	H-Donor	:GLU19:OE1	H-Acceptor