Computers in Biology and Medicine 160 (2023) 106929

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

TNFepitope: A webserver for the prediction of TNF-α inducing epitopes

Anjali Dhall [1], Sumeet Patiyal [1], Shubham Choudhury, Shipra Jain, Kashish Narang, Gajendra P.S. Raghava [*]

Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Phase 3, New Delhi, 110020, India

ARTICLE INFO

Keywords:
TNF-α inducing epitopes
Prediction
Designing
Hybrid method
Subunit vaccines

ABSTRACT

Tumor Necrosis Factor alpha (TNF-α) is a pleiotropic pro-inflammatory cytokine that is crucial in controlling the signaling pathways within the immune cells. Recent studies reported that higher expression levels of TNF-α are associated with the progression of several diseases, including cancers, cytokine release syndrome in COVID-19, and autoimmune disorders. Thus, it is the need of the hour to develop immunotherapies or subunit vaccines to manage TNF-α progression in various disease conditions. In the pilot study, we proposed a host-specific in-silico tool for predicting, designing, and scanning TNF-α inducing epitopes. The prediction models were trained and validated on the experimentally validated TNF-α inducing/non-inducing epitopes from human and mouse hosts. Firstly, we developed alignment-free (machine learning based models using composition-based features of peptides) methods for predicting TNF-α inducing peptides and achieved maximum AUROC of 0.79 and 0.74 for human and mouse hosts, respectively. Secondly, an alignment-based (using BLAST) method has been used for predicting TNF-α inducing epitopes. Finally, a hybrid method (combination of alignment-free and alignment-based method) has been developed for predicting epitopes. Hybrid approach achieved maximum AUROC of 0.83 and 0.77 on an independent dataset for human and mouse hosts, respectively. We have also identified potential TNF-α inducing peptides in different proteins of HIV-1, HIV-2, SARS-CoV-2, and human insulin. The best models developed in this study has been incorporated in the webserver TNFepitope (https://webs.iiitd.edu.in/raghava/tnfepitope/), standalone package and GitLab (https://gitlab.com/raghavalab/tnfepitope).

1. Introduction

Tumor Necrosis Factor alpha (TNF-α) is a classical, pleiotropic pro-inflammatory cytokine that functions by promoting the cellular signal activation and trafficking of leukocytes to the inflammatory sites [1]. During acute inflammation, TNF-α cytokine is released by macrophages/monocytes or via other cell types (e.g., B cells, T cells, mast cells, fibroblasts), which further regulates hematopoiesis, immune responses, tumor regression and various infections [2–6]. TNF-α is the first "adipokine" reported in the literature to be produced from the adipose tissue [7–9]. It plays a significant role in various biological processes, including immunomodulation, fever, inflammatory response, inhibition of tumor formation, and inhibition of virus replication [10]. TNF-α is involved in various physiological processes, for instance, the induction of pro-inflammatory interleukins (IL-1 and IL-6) [11–13]. It also interacts with various cytokines/chemokines and regulates signaling pathways in different disease states [14]. Studies have demonstrated that peptide-based vaccines are used for the treatment of various diseases, including cancer [15–20]. For instance, Probst et al., conducted a study in which peptide vaccination strategies and tumor-homing TNF fusion proteins are used for cancer treatment [21]. Sluis et al., revealed that the vaccine induced TNF-α cytokine significantly causes tumoricidal effects and promotes cisplatin-mediated death of tumor cells [22]. Moreover, TNF-receptor superfamily agonists are used as adjuvants for cancer vaccines [23].

Recent studies also showed that, the higher expression of TNF-α cytokine leads to the pathogenesis of numerous diseases including ischemia-reperfusion injury, sepsis, chronic heart failure, viral

* Corresponding author. Department of Computational Biology Indraprastha Institute of Information Technology, Delhi Okhla Industrial Estate, Phase III, (Near Govind Puri Metro Station), Office: A-302 (R&D Block), New Delhi, 110020, India.

E-mail addresses: anjalid@iiitd.ac.in (A. Dhall), sumeetp@iiitd.ac.in (S. Patiyal), shubhamc@iiitd.ac.in (S. Choudhury), shipra@iiitd.ac.in (S. Jain), kashish19026@iiitd.ac.in (K. Narang), raghava@iiitd.ac.in (G.P.S. Raghava).

URL: http://webs.iiitd.edu.in/raghava/ (G.P.S. Raghava).

[1] Equal Authors.

https://doi.org/10.1016/j.compbiomed.2023.106929
Received 9 November 2022; Received in revised form 30 March 2023; Accepted 13 April 2023
Available online 20 April 2023

myocarditis, and cardiac allograft rejection [24–26]. For example, Guo et al., reported that, the cytokine release syndrome in COVID-19 patients is associated with increased levels of TNF-α, IL-6, IL-2, IL-7, and IL-10 cytokines [27]. Moreover, there is a direct relationship between TNF-α and IL-6 cytokines in the severity and survival of COVID-19 patients [28–30]. Therefore, several anti-TNF inhibitors or drugs (including etanercept, infliximab, adalimumab, certolizumab pegol, and golimumab) are approved by FDA to treat a number of diseases. These inhibitors are used to block the overproduction of TNF-α in different disease conditions like ankylosing spondylitis, Crohn's disease, hidradenitis suppurativa, juvenile idiopathic arthritis, plaque psoriasis, polyarticular juvenile idiopathic arthritis, psoriatic arthritis, rheumatoid arthritis, ulcerative colitis, and uveitis [31–34]. Anti-TNF-α therapy has reported beneficial effects by not only restoring aberrant TNF mediated immune mechanisms but also by deactivating the pathogenic fibroblast-like mesenchymal cells [35].

As reported in the literature, TNF-α is a key cytokine involved in several diseases and their progression. Therefore, it can act as a primary target cytokine in disease progression. This creates a need to develop a computational tool for predicting TNF-α inducing peptides using sequence information. In the present study, we have developed an in-silico method to classify the TNF-α inducing and non-inducing epitopes. We have developed this tool using experimentally validated TNF-α inducing and non-inducing peptides from the human and mouse hosts. Additionally, we have also used randomly generated peptides from the SwissProt database [36] as the negative dataset. We have developed prediction models using various machine learning classifiers and evaluated their performance on the independent dataset.

2. Material and methods

2.1. Overall workflow

The complete workflow of the current study is illustrated in Fig. 1.

2.2. Dataset collection and preprocessing

In this study, we have collected experimentally validated TNF-α inducing peptides from the immune epitope database (IEDB [37]). After pre-processing, we observed that 3177 out of 3635 TNF-α inducing peptides are belong to human or mouse hosts, and only a few epitopes were available for other hosts. So, we worked with only two major hosts (i.e., human and mouse). We found most of the peptides lie within the range of 8–20 amino-acid residues; hence we fixed the length and removed the redundant peptides from our final dataset. Finally, we obtained 1215 and 539 TNF-α inducing epitopes for humans and mouse host, respectively. After that, we generated two separate negative datasets for both human and mouse hosts. The first negative dataset was collected from IEDB, containing 2383 experimentally validated TNF-α non-inducing epitopes for both hosts. After preprocessing, we obtain 1312 unique TNF-α non-inducing epitopes within a range of (8–20 amino acids) for human host. On the other hand, we have 539 unique TNF-α non-inducing epitopes for the mouse within the similar length range.

Finally, the main dataset for human incorporates 1215 TNF-α inducing and 1312 TNF-α non-inducing peptides. On the other side, the mouse dataset incorporates a total of 539 TNF-α inducing and 539 non-inducing peptides in the main dataset. The alternate negative dataset incorporated random peptides generated using the Swiss-Prot database [36]. The alternate dataset for human incorporates 1215 TNF-α inducing and 1215 randomly generated peptides. Similarly, in case of mouse we have a total of 539 TNF-α inducing and 539 randomly generated peptides. The final datasets for both human and mouse hosts were divided into training and independent dataset. Here, the complete dataset was split into 80:20 ratio, where 80% data was used to train the models and 20% data was kept aside for external validation.

2.3. Composition-based analysis

We have used Pfeature [38] to calculate the amino acid composition (AAC) of main and alternate datasets. Using the compositional analysis, we attempt to understand the similarity between the different peptide sequences taken from positive and negative datasets. Using the following equation 1, we have generated a feature vector of length 20, which specify the percent composition of 20 amino-acid residues.

$$AAC_i = \frac{AAR_i}{Total\ number\ of\ residues} \times 100$$

where AAC_i and AAR_i are the percentage composition and number of residues of type i in a peptide, respectively.

2.4. WebLogo

In order to understand the positional preference of amino-acid residues, we have generated sequence logos using WebLogo software [39] (http://weblogo.threeplusone.com). In the WebLogo, the x-axis represents the amino-acid residues, and the y-axis presents the bit-score, which shows the importance of a particular residue at a given

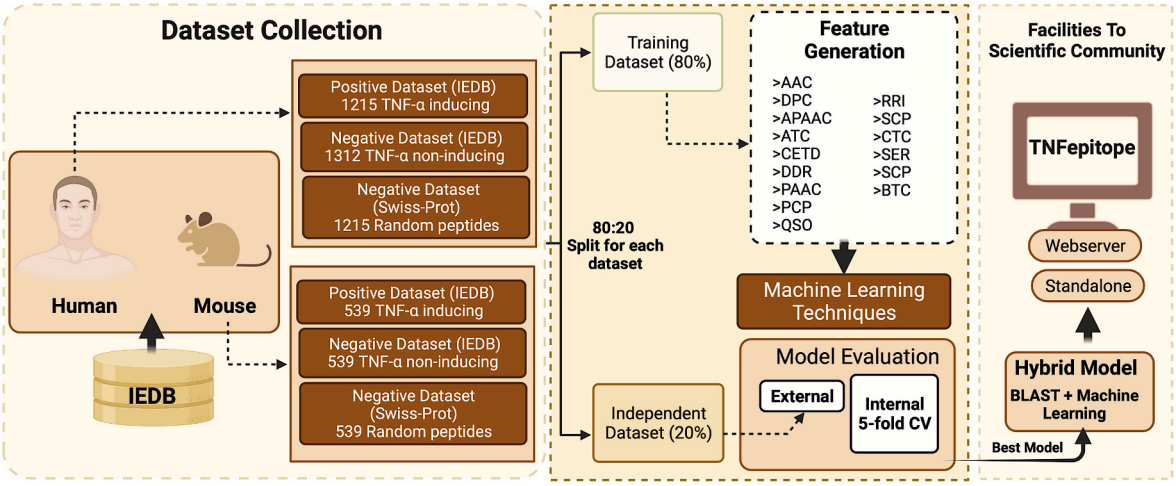


Fig. 1. Overall architecture of the study representing the dataset creation, feature generation, model building using machine learning algorithms and development of webserver.

position. WebLogo takes a fixed-length vector of input peptide sequences. To create a fixed length vector, we have considered eight amino acids from the N-terminal and C-terminal of the peptides, as eight is the minimum length of peptides in our dataset and merged them to generate a fixed length vector of sixteen residues for both human and mouse TNF-α inducing epitopes.

2.5. Generation of composition-based features

In the current study, we have calculated a wide range of features using the sequence information of the peptide sequences. We have used Pfeature [38] standalone package in order to calculate the composition-based features. We have computed different types of descriptors/features such as AAC (Amino Acid Composition), DPC (Di-Peptide Composition), APAAC (Amphiphilic Pseudo Amino Acid Composition), ATC (Atomic Composition), CETD (Composition-Enhanced Transition Distribution), DDR (Distance Distribution of Residues), PAAC (Pseudo Amino Acid Composition), PCP (Physico--Chemical Properties composition), QSO (Quasi-Sequence Order), RRI (Residue Repeat Information), SPC (Shannon entropy of Physico-Chemical properties), CTD (Conjoint Triad Descriptors), etc. In this study, we have developed prediction models using each feature as well as a combination of all the features.

2.6. Machine learning and cross-validation techniques

In order to develop prediction models, we have used various machine learning algorithms such as Random Forest (RF), Decision Tree (DT), Gaussian Naive Bayes (GNB), Logistic Regression (LR), Support Vector Classifier (SVC), K-Nearest Neighbor (KNN) and Extra Tree (ET). We have trained the parameters on training dataset and predictions were made on the independent dataset. Scikit-learn [40] python library was used in the study for the implementation of various classifiers. We have employed five-fold cross validation technique in order to evade the curse of biasness and overfitting. In the five-fold cross-validation technique, first the training dataset was divided into five equal sets; where four sets were used for training and fifth set was used for testing. This process is repeated five times where each part gets utilized for testing of the model as shown in some previous studies [41–47]. Of note, the final performance is the mean of the performances resulted after each iteration.

2.7. Similarity search approach

We have used BLAST [48] to implement similarity search or alignment-based approach; where we classify the epitopes as TNF-α inducing and non-inducing on the basis of the sequence similarity. Here, we have used NCBI-BLAST + version 2.2.29 (blastp suite) for similarity search and makeblastdb suite of NCBI-BLAST + for the creation of custom database. We have created a custom database using the training dataset; and sequences of validation dataset were queried against the created database. Based on the hits and their similarity with the customized database, we assign the class as TNF-α inducer or non-inducer. Currently we have considered only the top-hit of BLAST (i.e., if the top-hit of BLAST is against the TNF-α inducer peptide then the query sequence was assigned as TNF-α inducing peptide or vice-versa). To identify the optimal value of e-value; we ran BLAST at various e-values cut-offs varying from 1e-6 to 1e+3.

2.8. Hybrid model

In order to improve the performance of prediction models, we have applied a hybrid approach in which we merge alignment-based (BLAST) and alignment-free (machine learning based prediction) methods. Here, first we classify the peptide/epitope based on the BLAST query. After that, we add '0.5' score for the positive prediction i.e., TNF-α inducing peptide, '-0.5' score integrated for the negative predictions i.e., TNF-α non-inducing peptide and '0' score if no-hit was found. Further, we incorporate the prediction score calculated using machine learning based models. Finally, we combine the BLAST score and machine learning prediction score to make final predictions.

2.9. Performance evaluation

The performance of different models were evaluated using standard performance evaluation parameters sensitivity, specificity, accuracy, Area Under Receiver Operating Characteristics (AUROC) curve, Area Under the Precision-Recall Curve (AUPRC), Matthews Correlation Coefficient (MCC), and F1-score. We have computed both threshold-dependent (including sensitivity, specificity, accuracy, F1-score, and MCC) and independent parameters such as AUROC and AUPRC. The equations of evaluation parameters is provided in equations (2)–(6).

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad [2]$$

$$Specificity = \frac{T_N}{T_N + F_P} \quad [3]$$

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad [4]$$

$$F1 - Score = \frac{2T_P}{2T_P + F_P + F_N} \quad [5]$$

$$MCC = \frac{(T_P * T_N) - (F_P * F_N)}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad [6]$$

Where, F_P is false positive, F_N is false negative, T_P is true positive, and T_N is true negative.

3. Results

3.1. Compositional analysis

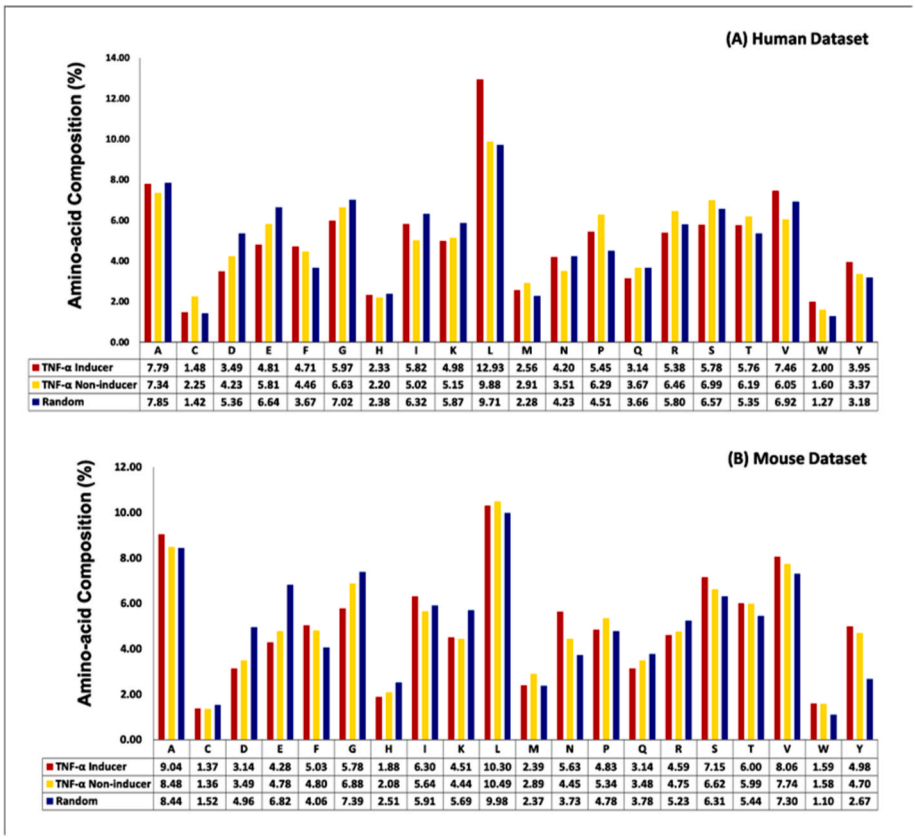
We have computed amino acid composition for the main and alternate datasets for human and mouse hosts. After that, we have calculated the average compositions for each amino acid residues in TNF-α inducing and non-inducing peptides. As depicted in Fig. 2A, in case of human dataset, amino acids such as leucine (L), valine (V), tyrosine (Y), and tryptophan (W) have higher composition in the TNF-α inducing peptides in comparison with the TNF-α non-inducing and random peptides. Similarly, the average composition of residues like alanine (A), isoleucine (I), asparagine (N), and serine (S) are more abundant in TNF-α inducing peptides of mouse dataset (See Fig. 2B).

3.2. Positional conservation analysis

In this analysis, we study the preference of residues at particular positions in the TNF-α inducing epitopes for human and mouse dataset. In the case of human TNF-α inducing epitopes, residues 'L' is highly conserved at most of the positions, whereas 'V' is preferred at 9th and 16th positions; 'A' is located on 7th, 9th, 10th, 11th, 12th, 13th and 16th positions (See Fig. 3A). In the case of mouse TNF-α inducing epitopes, 'L' is highly dominated on 2nd, 3rd, 8th, 9th, 12th, 13th and 16th positions; similarly residue 'N' is highly conserved at 5th and 13th positions; however, 'A' is predominated on 5th, 8th, 9th, 13th, 16th positions, as shown in Fig. 3B.

3.3. Machine learning based predictions

We have developed prediction models using different classifiers such as DT, RF, GNB, KNN, SVC, LR and ET on main and alternate datasets of both human and mouse hosts. For this, we have generated 15 different



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
TNF-α Inducer	7.79	1.48	3.49	4.81	4.71	5.97	2.33	5.82	4.98	12.93	2.56	4.20	5.45	3.14	5.38	5.78	5.76	7.46	2.00	3.95
TNF-α Non-inducer	7.34	2.25	4.23	5.81	4.46	6.63	2.20	5.02	5.15	9.88	2.91	3.51	6.29	3.67	6.46	6.99	6.19	6.05	1.60	3.37
Random	7.85	1.42	5.36	6.64	3.67	7.02	2.38	6.32	5.87	9.71	2.28	4.23	4.51	3.66	5.80	6.57	5.35	6.92	1.27	3.18

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
TNF-α Inducer	9.04	1.37	3.14	4.28	5.03	5.78	1.88	6.30	4.51	10.30	2.39	5.63	4.83	3.14	4.59	7.15	6.00	8.06	1.59	4.98
TNF-α Non-inducer	8.48	1.36	3.49	4.78	4.80	6.88	2.08	5.64	4.44	10.49	2.89	4.45	5.34	3.48	4.75	6.62	5.99	7.74	1.58	4.70
Random	8.44	1.52	4.96	6.82	4.06	7.39	2.51	5.91	5.69	9.98	2.37	3.73	4.78	3.78	5.23	6.31	5.44	7.30	1.10	2.67

Fig. 2. Plots representing the average amino-acid composition of each amino acid in TNF-α inducing, non-inducing and random peptides generated using Swiss-Prot database in (A) Human and (B) Mouse host.

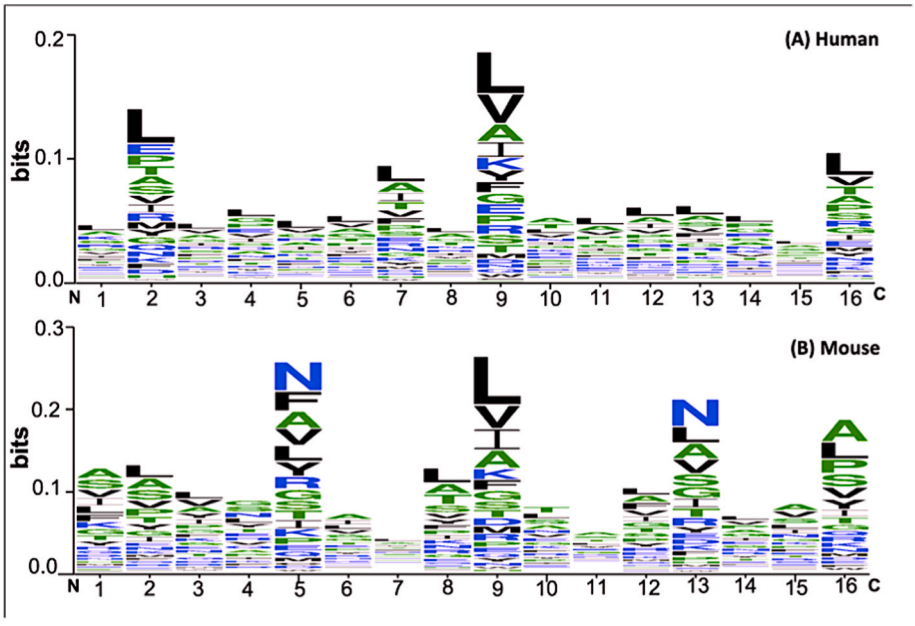


Fig. 3. Sequence logos generated using WebLogo software depicting the positional preference of residues in TNF-α inducing peptides in (A) Human and (B) Mouse host.

types of composition-based features using Pfeature standalone. We evaluated the performance on different features as well as combining all the features.

3.4. Performance of composition-based features

Here, we have computed performance on 15 different features. We have observed that RF and ET classifiers performed best among the other classifiers (See Supplementary Table S1). As shown in Table 1, in the

Table 1
Various performance metrics for the best performing models developed using 15 different types of composition-based features for human independent datasets.

Feature Type	Main Dataset						Alternate Dataset					
	Sens	Spec	Acc	AUROC	AUPRC	MCC	Sens	Spec	Acc	AUROC	AUPRC	MCC
AAC	55.97	58.56	57.31	0.63	0.61	0.15	63.37	66.26	64.82	0.70	0.72	0.30
DPC	72.02	72.62	72.33	0.79	0.76	0.45	68.72	61.73	65.23	0.71	0.73	0.31
ATC	55.97	58.56	57.31	0.63	0.61	0.15	59.67	58.03	58.85	0.61	0.62	0.18
APAAC	68.31	74.91	71.74	0.78	0.75	0.43	63.37	67.49	65.43	0.70	0.73	0.31
BTC	69.55	68.82	69.17	0.69	0.64	0.38	55.97	50.62	53.29	0.55	0.53	0.07
CETD	66.67	70.34	68.58	0.74	0.72	0.37	61.32	61.32	61.32	0.64	0.64	0.23
CTD	61.32	66.92	64.23	0.70	0.65	0.28	62.14	61.73	61.93	0.66	0.68	0.24
DDR	72.02	73.76	72.93	0.77	0.74	0.46	62.55	64.61	63.58	0.70	0.71	0.27
PAAC	68.31	74.14	71.34	0.78	0.75	0.43	65.02	65.43	65.23	0.70	0.72	0.31
PCP	64.61	67.68	66.21	0.73	0.72	0.32	62.96	63.37	63.17	0.67	0.67	0.26
QSO	62.55	71.86	67.39	0.72	0.71	0.35	63.79	65.43	64.61	0.69	0.71	0.29
RRI	62.55	68.06	65.42	0.73	0.70	0.31	62.96	57.20	60.08	0.66	0.69	0.20
SEP	63.37	60.84	62.06	0.69	0.67	0.24	43.62	57.61	50.62	0.51	0.50	0.01
SER	67.08	73.38	70.36	0.78	0.75	0.41	64.61	67.90	66.26	0.70	0.73	0.33
SPC	66.67	73.38	70.16	0.74	0.73	0.40	65.02	62.14	63.58	0.68	0.70	0.27
ALL_COMP	68.31	74.91	71.73	0.77	0.74	0.433	65.43	65.02	65.22	0.71	0.73	0.30

*Sens: Sensitivity; Spec: Specificity; Acc: Accuracy; AUROC: Area Under the Receiver Operating Characteristics curve; AUPRC: Area Under the Precision Recall Curve; MCC: Matthews Correlation Coefficient; AAC: Amino Acid Composition; DPC: Di-peptide Composition; ATC: Atomic Composition; APAAC: Amphiphilic Pseudo Amino Acid Composition; BTC: Bond Composition; CETD: Composition-Enhanced Transition Distribution; CTD: Conjoint Triad Descriptors; DDR: Distance Distribution of Residues; PAAC: Pseudo Amino Acid Composition; PCP: Physico-Chemical Properties composition; QSO: Quasi-Sequence Order; RRI: Residue Repeat Information; SEP: Shannon-Entropy of Peptide; SER: Shannon-Entropy of Residues; SPC: Shannon-entropy of Physico-Chemical properties; ALL_COMP: Combination of All Composition based features.

case of human host, we achieved maximum performance on main dataset with an AUROC of 0.79 and MCC of 0.45 on the independent dataset using DPC based features. APAAC and SER based features also performed quite well on independent dataset with an AUROC of 0.78 and AUPRC of 0.75. In the case of alternate dataset, we attained a maximum AUROC of 0.71, AUPRC of 0.73 and MCC of 0.31 using DPC based features. Upon combining all the features, we got an AUROC of 0.77 and 0.71 on main and alternate dataset for human host, respectively. Other composition-based features, performed poorly on both main and alternate datasets. The complete results of all the classifiers for each feature type are shown in Supplementary Table S2.

In case of mouse dataset, RF-based classifier performed well with an AUROC of 0.74, AUPRC of 0.76 and MCC of 0.34 on alternate dataset using DPC as the input feature (See Table 2). Similarly, we achieved similar performance (i.e., AUROC = 0.72, AUPRC = 0.73, and MCC = 0.30) using AAC-based features on the alternate dataset. In addition, RRI, DDR and APAAC also perform quite well with AUROC>0.72 on the alternate dataset. However, the performance of machine learning models is comparatively poor on the main dataset. The complete results on training and independent dataset is provide in Supplementary Tables S3 and S4.

3.5. Performance of hybrid models

In this study, we have developed a hybrid model to classify TNF-α inducing and non-inducing peptides. At first, we have used the similarity search approach (BLAST) for the prediction of positive and negative peptides. As shown in Tables 1 and 2, DPC based features outperformed other features, on both human and mouse prediction models. Hence, we combined BLAST similarity scores and machine learning scores

Table 2
Performance measures for best performing classifiers developed using 15 different types of composition-based features for mouse independent datasets.

Feature Type	Main Dataset						Alternate Dataset					
	Sens	Spec	Acc	AUROC	AUPRC	MCC	Sens	Spec	Acc	AUROC	AUPRC	MCC
AAC	62.18	60.56	61.37	0.67	0.66	0.23	64.82	64.82	64.82	0.72	0.73	0.30
DPC	58.47	59.86	59.17	0.63	0.62	0.18	66.67	67.59	67.13	0.74	0.76	0.34
ATC	51.97	50.35	51.16	0.54	0.53	0.02	55.56	62.04	58.80	0.65	0.62	0.18
APAAC	62.18	60.09	61.14	0.65	0.63	0.22	63.89	65.74	64.82	0.72	0.73	0.30
BTC	51.51	52.44	51.97	0.55	0.53	0.04	51.85	58.33	55.09	0.56	0.55	0.10
CETD	56.15	58.24	57.19	0.62	0.63	0.14	63.89	66.67	65.28	0.70	0.73	0.31
CTD	51.51	53.13	52.32	0.56	0.57	0.05	65.74	63.89	64.82	0.68	0.68	0.30
DDR	56.85	59.86	58.35	0.62	0.63	0.17	69.44	67.59	68.52	0.74	0.75	0.37
PAAC	60.79	61.02	60.91	0.65	0.64	0.22	67.59	65.74	66.67	0.72	0.73	0.33
PCP	57.77	61.49	59.63	0.61	0.59	0.19	56.48	69.44	62.96	0.70	0.70	0.26
QSO	58.01	58.47	58.24	0.60	0.59	0.17	61.11	70.37	65.74	0.73	0.74	0.32
RRI	59.86	60.79	60.33	0.63	0.62	0.21	65.74	66.67	66.20	0.75	0.74	0.32
SEP	55.68	54.06	54.87	0.57	0.56	0.10	36.11	51.85	43.98	0.45	0.46	−0.12
SER	60.56	62.41	61.49	0.67	0.66	0.23	67.59	69.44	68.52	0.73	0.74	0.37
SPC	57.77	58.47	58.12	0.61	0.59	0.16	60.19	69.44	64.82	0.69	0.66	0.30
ALL_COMP	62.96	62.96	62.96	0.67	0.67	0.26	64.81	68.51	66.67	0.73	0.73	0.33

*Sens: Sensitivity; Spec: Specificity; Acc: Accuracy; AUROC: Area Under the Receiver Operating Characteristics curve; AUPRC: Area Under the Precision Recall Curve; MCC: Matthews Correlation Coefficient; AAC: Amino Acid Composition; DPC: Di-peptide Composition; ATC: Atomic Composition; APAAC: Amphiphilic Pseudo Amino Acid Composition; BTC: Bond Composition; CETD: Composition-Enhanced Transition Distribution; CTD: Conjoint Triad Descriptors; DDR: Distance Distribution of Residues; PAAC: Pseudo Amino Acid Composition; PCP: Physico-Chemical Properties composition; QSO: Quasi-Sequence Order; RRI: Residue Repeat Information; SEP: Shannon-Entropy of Peptide; SER: Shannon-Entropy of Residues; SPC: Shannon-entropy of Physico-Chemical properties; ALL_COMP: Combination of All Composition based features.

computed using DPC features to make the final predictions. As shown in Supplementary Table S2, RF- and ET-based models performed well on main and alternate human datasets, respectively. We have used DPC features and the best models to calculate the performance for the hybrid model at different e-value cutoffs on independent datasets as exhibit in Table 3 for human host. We obtained the best performance at e-value (1.00E-01) with AUROC of (0.83 and 0.79), AUPRC of (0.80 and 0.84), MCC of (0.52 and 0.41) on main and alternate dataset, respectively (See Table 3). The complete results for training and independent datasets are provided in Supplementary Table S3.

We have applied a similar approach on mouse dataset, as provided in Supplementary Table S4, RF-based model outperforms the other classifiers on both main and alternate human datasets with DPC-based features. Using hybrid model, we achieved highest performance at e-value (1.00E-01) with AUROC of (0.70 and 0.77), AUPRC of (0.69 and 0.81), MCC of (0.28 and 0.34) on main and alternate dataset, respectively (See Table 4). The comprehensive results for training and independent datasets are given in Supplementary Table S5.

3.6. Services to scientific community

We have developed a web-server named 'TNFepitope' for the prediction of TNF-α inducing and non-inducing epitopes using sequence information. The best prediction models for human and mouse hosts were integrated in the webserver. We have incorporated five major modules in the server (i) Predict; (ii) Design; (iii) Scan; (iv) Blast Search; and (v) Standalone. 'Predict' module facilitates the users to stratify TNF-α inducing peptides from the non-inducing peptides. The 'Design' module provide the facility to the user to design/create all possible mutants of query sequence and predict if that can induce the TNF-α release or not. The 'Scan' module allows the user to map/scan the TNF-α secreting segment in the query protein sequence. The 'BLAST Search' module is entirely based on similarity search algorithm where the input sequence is hit against the customized database created using the known TNF-α inducing and non-inducing peptides. The submitted amino-acid sequence is predicted as TNF-α inducer/non-inducer based on the similarity with the sequences in the database. 'TNFepitope' server was developed using HTML, JAVA and PHP scripts; it is compatible with a number of devices such as laptops, iPhone, tablets, etc. The webserver (https://webs.iiitd.edu.in/raghava/tnfepitope), standalone package (https://webs.iiitd.edu.in/raghava/tnfepitope/package.php) and GitLab (https://gitlab.com/raghavalab/tnfepitope) are freely-accessible. Fig. 4 depicts all the major modules of TNFepitope webserver.

3.7. Case study

In order to demonstrate the application of our work, we predicted TNF-α inducing epitopes using 'Scan' module of TNFepitope webserver with default parameters (i.e., length of peptide 15 and threshold 0.45 with the hybrid method). Here, we have used three viral proteins (envelope glycoprotein of HIV-1, HIV-2, and surface glycoprotein/spike protein of SARS-CoV-2), two human proteins (insulin protein and insulin receptor protein) and food protein (rice Q10 MI4). As depicted in Table 5, we did not find any BLAST hits against rice protein, meaning that it does not activate/induce TNF-α production. This strategy can be used to scan TNF-α inducing regions in other foods or Genetically modified (GM) foods. Similarly, in the case of human insulin receptor protein, we did not find any hits. Interestingly, we discovered that human insulin hormone which is a small protein contains the highest percentage of TNF-α inducing regions i.e., 55.21% (See Table 5). This shows that elevation in insulin levels is responsible for the production of TNF-α peptides/epitopes. This observation is in agreement with previous studies where they have demonstrated that insulin resistant patients have higher levels of TNF-α [49,50].

In addition, various studies have reported that elevated levels of TNF-α is associated with the pathogenesis of viral infections such as human immunodeficiency virus (HIV) and SARS-CoV-2 [27,51–53]. As shown in Table 5, the envelope proteins of HIV-1 and HIV-2 possesses 24.82% and 26.48% TNF-α inducing regions, while the spike protein of SARS-CoV-2 has 36.38% TNF-α inducers, which supports the previous studies where severity in COVID-19 patients is associated with the high levels of TNF-α. In Supplementary Table S7, we have provided the top-most TNF-α inducing epitopes of HIV-1, HIV-2, spike protein and human insulin protein. The complete results for each protein in provided in Supplementary Tables S8–S13. These results indicates that our study can be used to predict the TNF-α inducing capabilities of different viral proteins. We hope that our findings can assist the scientific community, working in the era of subunit vaccine designing against deadly viruses and other autoimmune diseases that can be proliferated by the elevation of TNF-α.

4. Discussion

Major histocompatibity complex region encodes numbers of proteins including human leukocyte antigen (HLAs) which are necessary for self-recognition, cytokine genes like TNF, LTA, LTB, which are responsible for the inflammations [54]. TNF-α is a significant inflammatory cytokine produced by T cells and macrophages that regulates several immune cell signaling pathways that result in necrosis or cell death [3,4]. These pathways are involved in a range of biological responses, such as cell proliferation, differentiation, and survival. TNF-α cytokine employed for cancer treatment and perform anti-cancer activities by inducing inflammation, immune response, and tumor cell apoptosis [55–57]. However, improper and excessive activation of TNF signalling pathway may results in the emergence of pathological diseases such as HIV-I, anorexia, cachexia, obesity, autoimmune disorders including rheumatoid arthritis, diabetes, inflammatory bowel disease, and Crohn's

Table 3
Performance measures for the hybrid model on different E-values, build by integrating alignment-based (BLAST) and alignment-free (machine learning) approach on independent datasets for human host.

E-value	Main Dataset						Alternate Dataset					
	Sens	Spec	Acc	AUROC	AUPRC	MCC	Sens	Spec	Acc	AUROC	AUPRC	MCC
1.00E-06	72.43	76.34	74.46	0.82	0.79	0.49	65.02	65.02	65.02	0.72	0.76	0.30
1.00E-05	73.66	77.48	75.64	0.81	0.77	0.51	67.49	65.84	66.67	0.73	0.77	0.33
1.00E-04	72.84	75.57	74.26	0.81	0.76	0.48	66.26	69.14	67.70	0.73	0.77	0.35
1.00E-03	72.43	77.10	74.85	0.81	0.77	0.50	65.02	69.14	67.08	0.73	0.78	0.34
1.00E-02	74.90	76.72	75.84	0.82	0.77	0.52	68.72	69.55	69.14	0.78	0.83	0.38
1.00E-01	76.13	75.95	76.04	0.83	0.80	0.52	70.37	70.78	70.58	0.79	0.84	0.41
1.00E+00	76.54	75.95	76.24	0.83	0.81	0.53	68.72	67.90	68.31	0.77	0.81	0.37
1.00E+01	73.25	74.81	74.06	0.82	0.79	0.48	67.49	68.31	67.90	0.74	0.78	0.36
1.00E+02	72.84	72.14	72.48	0.82	0.79	0.45	67.08	67.49	67.28	0.73	0.78	0.35

*Sens: Sensitivity; Spec: Specificity; Acc: Accuracy; AUROC: Area Under the Receiver Operating Characteristics curve; AUPRC: Area Under the Precision Recall Curve; MCC: Matthews Correlation Coefficient.

Table 4
Performance measures for the hybrid model on different E-values, build by integrating alignment-based (BLAST) and alignment-free (machine learning) approach on independent datasets for mouse host.

E-value	Main Dataset						Alternate Dataset					
	Sens	Spec	Acc	AUROC	AUPRC	MCC	Sens	Spec	Acc	AUROC	AUPRC	MCC
1.00E-06	61.68	59.81	60.75	0.64	0.61	0.22	65.42	65.42	65.42	0.73	0.74	0.31
1.00E-05	69.16	52.34	60.75	0.63	0.61	0.22	64.49	68.22	66.36	0.73	0.75	0.33
1.00E-04	58.88	59.81	59.35	0.64	0.61	0.19	66.36	66.36	66.36	0.73	0.74	0.33
1.00E-03	62.62	64.49	63.55	0.67	0.65	0.27	66.36	67.29	66.82	0.73	0.74	0.34
1.00E-02	61.68	62.62	62.15	0.68	0.66	0.24	67.29	65.42	66.36	0.75	0.79	0.33
1.00E-01	62.62	65.42	64.02	0.70	0.69	0.28	66.36	67.29	66.82	0.77	0.81	0.34
1.00E+00	61.68	62.62	62.15	0.66	0.64	0.24	68.22	69.16	68.69	0.76	0.78	0.37
1.00E+01	60.75	60.75	60.75	0.66	0.64	0.22	65.42	65.42	65.42	0.71	0.74	0.31
1.00E+02	60.75	60.75	60.75	0.65	0.64	0.22	66.36	65.42	65.89	0.71	0.74	0.32

*Sens: Sensitivity; Spec: Specificity; Acc: Accuracy; AUROC: Area Under the Receiver Operating Characteristics curve; AUPRC: Area Under the Precision Recall Curve; MCC: Matthews Correlation Coefficient.

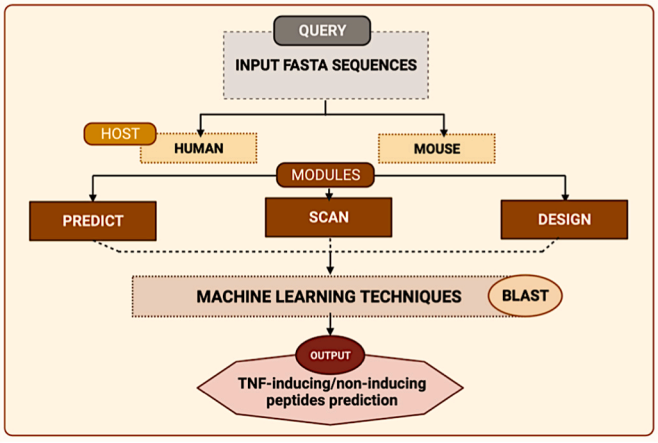


Fig. 4. Schematic representation of different modules of TNFepitope server which includes predict, scan, and design modules.

diseases [58–66]. Several TNF-α inhibitors such as infliximab, etanercept, golimumab, and certolizumab and adalimumab have been developed and approved for clinical use to cure diseases which are associated with abnormal/excessive TNF-α secretion [61,67].

Mortaz et al., also reported the higher level of soluble TNF-α in the patients of COVID-19 in comparison with the healthy control [68]. Therefore, it is crucial to check for the existence of TNF-α inducing epitopes while using anti-TNF therapy in a variety of diseases. In the current study, we have attempted to understand the nature of TNF-α inducing peptides and built a prediction model to recognize the epitopes which can induce TNF-α secretion. We have gathered experimentally confirmed TNF-inducing and non-inducing peptides for human and mouse hosts because datasets are crucial in the development of machine learning algorithms. Dataset plays a major role in developing machine learning models, hence we have collected experimentally validated TNF-α inducing and non-inducing peptides for human and mouse hosts. We created random peptides for the alternate negative dataset using the Swiss-Prot database. Sequence-logo and compositional analysis were done to look into the composition and placement preference. We found that TNF-α inducing epitopes are rich in the amino acid residue "L" in human and "N" in mouse datasets. Features/descriptors play important role in the development of machine learning models. In this study, we employed 'Pfeature' to compute 15 different types of composition-based features using the standalone package.

Our analysis revealed that, the amino-acid composition and dipeptide composition based features performed best for the classification of TNF-α inducing and non-inducing peptides. After that, we have applied two-sample *t*-test for each dipeptide composition in the TNF-α inducing and non-inducing peptides dataset. In order to check the significance of the important DPC features (See Supplementary Table S14). Bases on the analysis, the top-10 dipeptides features for human dataset (LL, AL, IL, KL, TL, LN, VA, LK, VV, SL) and mouse dataset (KA, NF, IN, VY, KS, AG, YK, VG, AN, NY), which are considerably more frequent DPC-features in the TNF-α associated peptides as compared to non-TNF-α associated peptides. While the atom & bond, Shannon entropy of peptide based composition feature performed poor in the case of main and alternate datasets for both human and mouse models. Using dipeptide composition based features, we have achieved highest AUROC of 0.79 and 0.74 on the human and mouse independent dataset. Of note, our hybrid model outperformed others with an AUROC of 0.83 and 0.77 on the human and mouse independent dataset. We have used the best models and created a web server TNFepitope (https://webs.iiitd.edu.in/raghava/tnfepitope) and a standalone package.

Our proposed method can be utilized by the researchers and scientific community for identification of suitable vaccine candidates against number of diseases including cancer. However, one of the major challenges while designing subunit vaccine candidate is the toxicity, allergenicity, haemolytic potential, and half-life of the peptides. Therefore, it is essential to check the important therapeutic properties of the predicted TNF-α inducing peptides before considering it as a subunit vaccine candidate. The experimental techniques are time-consuming and labour intensive, therefore several in-silico tools such as Algpred2.0, AllerHunter, ToxiPred, PEPlife, HemoPred, HemoPI, HemoPI-MOD [44,

Table 5
Potential TNF-α inducing epitopes predicted by "Protein Scan" module of TNFepitope server in 3 viral proteins (HIV-1, HIV-2, and SARS-CoV-2), 2 human proteins (insulin and insulin receptor) and 1 food protein (rice Q10 MI4).

Protein Name	Length	TNF-α inducing epitopes (Score>0.45)		TNF-α inducing epitopes (Score>0.70)		BLAST Hit (Positive)	
		Number of epitopes	Percentage (%)	Number of epitopes	Percentage (%)	Number of epitopes	Percentage (%)
Envelope glycoprotein(HIV-1)	834	207	24.82%	12	1.43%	9	1.08%
Envelope glycoprotein(HIV-2)	846	224	26.48%	7	0.82%	7	0.83%
Spike Protein(SARS-CoV-2)	1259	458	36.38%	251	19.94%	251	19.94%
Insulin protein (Human)	96	53	55.21%	52	54.16%	26	27.08%
Insulin receptor protein (Human)	1368	211	15.42%	0	0.00%	0	0.00%
Food(Rice protein Q10MI4)	881	167	18.96%	0	0.00%	0	0.00%

69–74], are available in the literature for the prediction of general therapeutic properties of peptides.

5. Conclusion

Designing a vaccine or immunotherapy against various diseases using peptide/epitope technology is a viable approach. TNF-α is a versatile cytokine plays major biological processes including cell survival, proliferation, differentiation, and death. Several clinical trials are being carried out to understand the effects of TNF-based therapy for the treatment of cancer patients. In the past, several in-silico approaches for predicting T cell epitopes were developed, however, there was no specific computational method for predicting TNF-α inducing epitopes/peptides. In this study, we have developed a host-specific computational tool for the prediction of TNF-α inducing or non-inducing peptides. We believe that the scientific community will benefit from this work in the development of peptide-based subunit vaccines.

Funding source

The current work has received grant from the Department of Bio-Technology (DBT), Govt. of India, India.

Authors' contributions

AD and GPSR collected and processed the datasets. AD, SP, KN and GPSR implemented the algorithms and developed the prediction models. AD, SP and GPSR analysed the results. SC, AD and SP created the web server. AD, SJ, SP and SC and GPSR penned the manuscript. GPSR conceived and coordinated the project. All authors have read and approved the final manuscript.

Declaration of competing interest

The authors declare no competing financial and non-financial interests.

Acknowledgements

Authors are thankful to the Department of Bio-Technology (DBT) and Department of Science and Technology (DST-INSPIRE) for fellowships and the financial support and Department of Computational Biology, IIITD New Delhi for infrastructure and facilities.

Biorxiv Link: https://doi.org/10.1101/2022.08.02.502430.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2023.106929.

References

[1] J.K. Sethi, G.S. Hotamisligil, Metabolic Messengers: tumour necrosis factor, Nat. Metab. 3 (2021) 1302–1312.
[2] B.B. Aggarwal, Signalling pathways of the TNF superfamily: a double-edged sword, Nat. Rev. Immunol. 3 (2003) 745–756.
[3] H.T. Idriss, J.H. Naismith, TNF alpha and the TNF receptor superfamily: structure-function relationship(s), Microsc. Res. Tech. 50 (2000) 184–195.
[4] J. Holbrook, S. Lara-Reyna, H. Jarosz-Griffiths, M. McDermott, Tumour Necrosis Factor Signalling in Health and Disease, F1000Res, 2019, p. 8.
[5] A.B. Adams, C.P. Larsen, T.C. Pearson, K.A. Newell, The role of TNF receptor and TNF superfamily molecules in organ transplantation, Am. J. Transplant. 2 (2002) 12–18.
[6] B. Wang, N. Song, T. Yu, L. Zhou, H. Zhang, L. Duan, W. He, Y. Zhu, Y. Bai, M. Zhu, Expression of tumor necrosis factor-alpha-mediated genes predicts recurrence-free survival in lung cancer, PLoS One 9 (2014), e115945.
[7] J.D. Janowska, C1q/TNF-related protein 1, a multifunctional adipokine: an overview of current data, Am. J. Med. Sci. 360 (2020) 222–228.
[8] T. Wang, C. He, Pro-inflammatory cytokines: the link between obesity and osteoarthritis, Cytokine Growth Factor Rev. 44 (2018) 38–50.
[9] G. Fantuzzi, Adipose tissue, adipokines, and inflammation, J. Allergy Clin. Immunol. 115 (2005) 911–919. ; quiz 920.
[10] K. You, H. Gu, Z. Yuan, X. Xu, Tumor necrosis factor alpha signaling and organogenesis, Front. Cell Dev. Biol. 9 (2021), 727075.
[11] L.J. Old, Tumor necrosis factor, Sci. Am. 258 (59–60) (1988) 69–75.
[12] S.I. Grivennikov, M. Karin, Inflammatory cytokines in cancer: tumour necrosis factor and interleukin 6 take the stage, Ann. Rheum. Dis. 70 (Suppl 1) (2011) i104–i108.
[13] J. Saklatvala, W. Davis, F. Guesdon, Interleukin 1 (IL1) and tumour necrosis factor (TNF) signal transduction, Philos. Trans. R. Soc. Lond. B Biol. Sci. 351 (1996) 151–157.
[14] N. Parameswaran, S. Patial, Tumor necrosis factor-alpha signaling in macrophages, Crit. Rev. Eukaryot. Gene Expr. 20 (2010) 87–103.
[15] L. Zhang, Y. Huang, A.R. Lindstrom, T.Y. Lin, K.S. Lam, Y. Li, Peptide-based materials for cancer immunotherapy, Theranostics 9 (2019) 7807–7825.
[16] J. Dey, S.R. Mahapatra, T.K. Raj, T. Kaur, P. Jain, A. Tiwari, S. Patro, N. Misra, M. Suar, Designing a novel multi-epitope vaccine to evoke a robust immune response against pathogenic multidrug-resistant Enterococcus faecium bacterium, Gut Pathog. 14 (2022) 21.
[17] F. Zhu, S. Ma, H. Wen, M. Rao, P. Zhang, W. Peng, Y. Cui, H. Yang, C. Tan, J. Chen, P. Pan, Development of a novel circular mRNA vaccine of six protein combinations against Staphylococcus aureus, J. Biomol. Struct. Dyn. (2022) 1–21.
[18] J. Dey, S.R. Mahapatra, S. Patnaik, S. Lata, G.S. Kushwaha, R.K. Panda, N. Misra, M. Suar, Molecular characterization and designing of a novel multiepitope vaccine construct against Pseudomonas aeruginosa, Int. J. Pept. Res. Therapeut. 28 (2022) 49.
[19] S.R. Mahapatra, J. Dey, A. Jaiswal, R. Roy, N. Misra, M. Suar, Immunoinformatics-guided designing of epitope-based subunit vaccine from Pilus assembly protein of Acinetobacter baumannii bacteria, J. Immunol. Methods 508 (2022), 113325.
[20] P. Sahoo, J. Dey, S.R. Mahapatra, A. Ghosh, A. Jaiswal, S. Padhi, S. C. Prabhuswamimath, N. Misra, M. Suar, Nanotechnology and COVID-19 convergence: toward new planetary health interventions against the pandemic, OMICS 26 (2022) 473–488.
[21] P. Probst, M. Stringhini, D. Ritz, T. Fugmann, D. Neri, Antibody-based delivery of TNF to the tumor neovasculature potentiates the therapeutic activity of a peptide anticancer vaccine, Clin. Cancer Res. 25 (2019) 698–709.
[22] T.C. van der Sluis, S. van Duikeren, S. Huppelschoten, E.S. Jordanova, E. Beyranvand Nejad, A. Sloots, L. Boon, V.T. Smit, M.J. Welters, F. Ossendorp, B. van de Water, R. Arens, S.H. van der Burg, C.J. Melief, Vaccine-induced tumor necrosis factor-producing T cells synergize with cisplatin to promote tumor cell death, Clin. Cancer Res. 21 (2015) 781–794.
[23] T.N. Bullock, TNF-receptor superfamily agonists as molecular adjuvants for cancer vaccines, Curr. Opin. Immunol. 47 (2017) 70–77.
[24] B.S. Cain, D.R. Meldrum, C.A. Dinarello, X. Meng, K.S. Joo, A. Banerjee, A. H. Harken, Tumor necrosis factor-alpha and interleukin-1beta synergistically depress human myocardial function, Crit. Care Med. 27 (1999) 1309–1318.
[25] D. Bryant, L. Becker, J. Richardson, J. Shelton, F. Franco, R. Peshock, M. Thompson, B. Giroir, Cardiac failure in transgenic mice with myocardial expression of tumor necrosis factor-alpha, Circulation 97 (1998) 1375–1381.
[26] U. Muller-Werdan, M. Buerke, H. Ebelt, K.M. Heinroth, A. Herklotz, H. Loppnow, M. Russ, F. Schlegel, A. Schlitt, H.B. Schmidt, G. Soffker, K. Werdan, Septic cardiomyopathy - a not yet discovered cardiomyopathy? Exp. Clin. Cardiol. 11 (2006) 226–236.
[27] Y. Guo, K. Hu, Y. Li, C. Lu, K. Ling, C. Cai, W. Wang, D. Ye, Targeting TNF-alpha for COVID-19: recent advanced and controversies, Front. Public Health 10 (2022), 833967.
[28] C. Halim, A.F. Mirza, M.I. Sari, The association between TNF-alpha, IL-6, and vitamin D levels and COVID-19 severity and mortality: a systematic review and meta-analysis, Pathogens (2022) 11.
[29] D.M. Del Valle, S. Kim-Schulze, H.H. Huang, N.D. Beckmann, S. Nirenberg, B. Wang, Y. Lavin, T.H. Swartz, D. Madduri, A. Stock, T.U. Marron, H. Xie, M. Patel, K. Tuballes, O. Van Oekelen, A. Rahman, P. Kovatch, J.A. Aberg, E. Schadt, S. Jagannath, M. Mazumdar, A.W. Charney, A. Firpo-Betancourt, D. R. Mendu, J. Jhang, D. Reich, K. Sigel, C. Cordon-Cardo, M. Feldmann, S. Parekh, M. Merad, S. Gnjatic, An inflammatory cytokine signature predicts COVID-19 severity and survival, Nat. Med. 26 (2020) 1636–1643.
[30] A. Santa Cruz, A. Mendes-Frias, A.I. Oliveira, L. Dias, A.R. Matos, A. Carvalho, C. Capela, J. Pedrosa, A.G. Castro, R. Silvestre, Interleukin-6 is a biomarker for the development of fatal severe acute respiratory syndrome coronavirus 2 pneumonia, Front. Immunol. 12 (2021), 613422.
[31] L. Dreyer, L. Mellemkjaer, M.L. Hetland, [Cancer in arthritis patients after anti-tumour necrosis factor therapy], Ugeskr Laeger 171 (2009) 506–511.
[32] S. Menegatti, E. Bianchi, L. Rogge, Anti-TNF therapy in spondyloarthritis and related diseases, impact on the immune system and prediction of treatment responses, Front. Immunol. 10 (2019) 382.
[33] C. Plasencia, D. Pascual-Salcedo, S. Garcia-Carazo, L. Lojo, L. Nuno, A. Villalba, D. Peiteado, F. Arribas, J. Diez, M.T. Lopez-Casla, E. Martin-Mola, A. Balsa, The immunogenicity to the first anti-TNF therapy determines the outcome of switching to a second anti-TNF therapy in spondyloarthritis patients, Arthritis Res. Ther. 15 (2013) R79.
[34] L. Peyrin-Biroulet, Anti-TNF therapy in inflammatory bowel diseases: a huge review, Minerva Gastroenterol. Dietol. 56 (2010) 233–243.
[35] G. Evangelatos, G. Bamias, G.D. Kitas, G. Kollias, P.P. Sfikakis, The second decade of anti-TNF-a therapy in clinical practice: new lessons and future directions in the COVID-19 era, Rheumatol. Int. (2022).

[36] A. Bairoch, R. Apweiler, The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000, Nucleic Acids Res. 28 (2000) 45–48.
[37] V. R, M. S, O. Ja, D. Sk, M. S, C. Jr, W. Dk, S. A, P. B, The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343.
[38] A. Pande, S. Patiyal, A. Lathwal, C. Arora, D. Kaur, A. Dhall, G. Mishra, H. Kaur, N. Sharma, S. Jain, S.S. Usmani, P. Agrawal, R. Kumar, V. Kumar, G.P.S. Raghava, Computing Wide Range of Protein/peptide Features from Their Sequence and Structure, bioRxiv, 2019, p. 599126, 599126.
[39] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.
[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[41] A. Dhall, S. Patiyal, N. Sharma, S.S. Usmani, G.P.S. Raghava, Computer-aided prediction and design of IL-6 inducing peptides: IL-6 plays a crucial role in COVID-19, Briefings Bioinf. 22 (2021) 936–945.
[42] A. Dhall, S. Patiyal, G.P.S. Raghava, HLAncPred: a Method for Predicting Promiscuous Non-classical HLA Binding Sites, Brief Bioinform, 2022.
[43] A. Dhall, S. Patiyal, N. Sharma, N.L. Devi, G.P. Raghava, Computer-aided Prediction of Inhibitors against STAT3 for Managing COVID-19 Associate Cytokine Storm, 2021.
[44] N. Sharma, S. Patiyal, A. Dhall, A. Pande, C. Arora, G.P.S. Raghava, AlgPred 2.0: an Improved Method for Predicting Allergenic Proteins and Mapping of IgE Epitopes, Brief Bioinform, 2020.
[45] V. Kumar, S. Patiyal, A. Dhall, N. Sharma, G.P.S. Raghava, B3Pred: a random-forest-based method for predicting and designing blood-brain barrier penetrating peptides, Pharmaceutics 13 (2021).
[46] A.D. Sumeet Patiyal, Gajendra P.S. Raghava*, DBpred: A Deep Learning Method for the Prediction of DNA Interacting Residues in Protein Sequences, bioRxiv, 2021.
[47] S. Patiyal, P. Agrawal, V. Kumar, A. Dhall, R. Kumar, G. Mishra, G.P. Raghava, NAGbinder: an approach for identifying N-acetylglucosamine interacting residues of a protein from its primary sequence, Protein Sci. 29 (2020) 201–210.
[48] S. McGinnis, T.L. Madden, BLAST: at the core of a powerful and diverse set of sequence analysis tools, Nucleic Acids Res. 32 (2004) W20–W25.
[49] M.S.H. Akash, K. Rehman, A. Liaqat, Tumor necrosis factor-alpha: role in development of insulin resistance and pathogenesis of type 2 diabetes mellitus, J. Cell. Biochem. 119 (2018) 105–110.
[50] J.J. Swaroop, D. Rajarajeswari, J.N. Naidu, Association of TNF-alpha with insulin resistance in type 2 diabetes mellitus, Indian J. Med. Res. 135 (2012) 127–130.
[51] R. Planes, M. Serrero, K. Leghmari, L. BenMohamed, E. Bahraoui, HIV-1 envelope glycoproteins induce the production of TNF-alpha and IL-10 in human monocytes by activating calcium pathway, Sci. Rep. 8 (2018), 17215.
[52] S. Pasquereau, A. Kumar, G. Herbein, Targeting TNF and TNF receptor pathway in HIV-1 infection: from immune activation to viral reservoirs, Viruses (2017) 9.
[53] V. Ablamunits, C. Lepsy, Blocking TNF signaling may save lives in COVID-19 infection, Mol. Biol. Rep. 49 (2022) 2303–2309.
[54] T. Shiina, K. Hosomichi, H. Inoko, J.K. Kulski, The HLA genomic loci map: expression, interaction, diversity and disease, J. Hum. Genet. 54 (2009) 15–39.
[55] J. Shen, Z. Xiao, Q. Zhao, M. Li, X. Wu, L. Zhang, W. Hu, C.H. Cho, Anti-cancer therapy with TNFalpha and IFNgamma: a comprehensive review, Cell Prolif. 51 (2018), e12441.
[56] X. Wang, Y. Lin, Tumor necrosis factor and cancer, buddies or foes? Acta Pharmacol. Sin. 29 (2008) 1275–1288.
[57] A. Montfort, C. Colacios, T. Levade, N. Andrieu-Abadie, N. Meyer, B. Segui, The TNF paradox in cancer progression and immunotherapy, Front. Immunol. 10 (2019) 1818.
[58] B.R. Lane, D.M. Markovitz, N.L. Woodford, R. Rochford, R.M. Strieter, M.J. Coffey, TNF-alpha inhibits HIV-1 replication in peripheral blood monocytes and alveolar macrophages by inducing the production of RANTES and decreasing C-C chemokine receptor 5 (CCR5) expression, J. Immunol. 163 (1999) 3653–3661.
[59] V.C. Vaughan, P. Martin, P.A. Lewandowski, Cancer cachexia: impact, mechanisms and emerging treatments, J. Cachexia Sarcopenia Muscle 4 (2013) 95–109.
[60] E.H. Choy, G.S. Panayi, Cytokine pathways and joint inflammation in rheumatoid arthritis, N. Engl. J. Med. 344 (2001) 907–916.
[61] D.I. Jang, A.H. Lee, H.Y. Shin, H.R. Song, J.H. Park, T.B. Kang, S.R. Lee, S.H. Yang, The role of tumor necrosis factor alpha (TNF-alpha) in autoimmune disease and current TNF-alpha inhibitors in therapeutics, Int. J. Mol. Sci. 22 (2021).
[62] S.O. Adegbola, K. Sahnan, J. Warusavitarne, A. Hart, P. Tozer, Anti-TNF therapy in crohn's disease, Int. J. Mol. Sci. 19 (2018).
[63] A.D. Levin, M.E. Wildenberg, G.R. van den Brink, Mechanism of action of anti-TNF therapy in inflammatory bowel disease, J. Crohns Colitis 10 (2016) 989–997.
[64] B.E. Sands, G.G. Kaplan, The role of TNFalpha in ulcerative colitis, J. Clin. Pharmacol. 47 (2007) 930–941.
[65] K. Lis, O. Kuzawinska, E. Balkowiec-Iskra, Tumor necrosis factor inhibitors - state of knowledge, Arch. Med. Sci. 10 (2014) 1175–1185.
[66] S.P. Raychaudhuri, S.K. Raychaudhuri, Biologics: target-specific treatment of systemic and cutaneous autoimmune diseases, Indian J. Dermatol. 54 (2009) 100–109.
[67] P. Stenvinkel, M. Ketteler, R.J. Johnson, B. Lindholm, R. Pecoits-Filho, M. Riella, O. Heimburger, T. Cederholm, M. Girndt, IL-10, IL-6, and TNF-alpha: central factors in the altered cytokine network of uremia–the good, the bad, and the ugly, Kidney Int. 67 (2005) 1216–1233.
[68] E. Mortaz, P. Tabarsi, H. Jamaati, N. Dalil Roofchayee, N.K. Dezfuli, S. M. Hashemian, A. Moniri, M. Marjani, M. Malekmohammad, D. Mansouri, M. Varahram, G. Folkerts, I.M. Adcock, Increased serum levels of soluble TNF-alpha receptor is associated with ICU mortality in COVID-19 patients, Front. Immunol. 12 (2021), 592727.
[69] H.C. Muh, J.C. Tong, M.T. Tammi, AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins, PLoS One 4 (2009), e5861.
[70] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, C. Open Source Drug Discovery, G.P. Raghava, In silico approach for predicting toxicity of peptides and proteins, PLoS One 8 (2013), e73957.
[71] D. Mathur, S. Prakash, P. Anand, H. Kaur, P. Agrawal, A. Mehta, R. Kumar, S. Singh, G.P. Raghava, PEPlife: a repository of the half-life of peptides, Sci. Rep. 6 (2016), 36617.
[72] T.S. Win, A.A. Malik, V. Prachayasittikul, S.W. Je, C. Nantasenamat, W. Shoombuatong, HemoPred: a web server for predicting the hemolytic activity of peptides, Future Med. Chem. 9 (2017) 275–291.
[73] K. Chaudhary, R. Kumar, S. Singh, A. Tuknait, A. Gautam, D. Mathur, P. Anand, G. C. Varshney, G.P. Raghava, A web server and mobile app for computing hemolytic potency of peptides, Sci. Rep. 6 (2016), 22843.
[74] V. Kumar, R. Kumar, P. Agrawal, S. Patiyal, G.P.S. Raghava, A method for predicting hemolytic potency of chemically modified peptides from its structure, Front. Pharmacol. 11 (2020) 54.

Anjali Dhall is currently working as Ph.D. in Computational Biology from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Sumeet Patiyal is currently working as Ph.D. in Computational biology from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Shubham Choudhury is currently working as Ph.D. in Computational Biology from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Shipra Jain is currently working as Ph.D. in Computational Biology from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Kashish Narang is currently working as MTech. (CSE) at Indraprastha Institute of Information Technology, New Delhi, India.

Gajendra P. S. Raghava is currently working as Professor and Head of Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.