In silico model for predicting IL-2 inducing peptides in human

Anjali Lathwal[1, #], Rajesh Kumar[2, #], Dilraj kaur[1], and Gajendra P.S. Raghava[1, *]

1- Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi

2- Bioinformatics Centre, Institute of Microbial Technology, Chandigarh

 # - Contributed Equally

* - Corresponding Author

 Mailing Address of Authors

Anjali Lathwal: - anjalil@iiitd.ac.in

Rajesh Kumar: - b.rajesh1130@gmail.com

Dilraj Kaur: - dilrajk@iiitd.ac.in

Gajendra P.S. Raghava: - raghava@iiitd.ac.in

 Corresponding Author

Gajendra Pal Singh Raghava

Head and Professor

Department of Computational Biology

Indraprastha Institute of Information Technology, Delhi

Okhla Industrial Estate, Phase III, (Near Govind Puri Metro Station)

New Delhi, India - 110020

Office: A-302 (R&D Block)

Phone: 011-26907444

Email: raghava@iiitd.ac.in

Website: http://webs.iiitd.edu.in/raghava/

ORCID id - 0000-0002-8902-2876

Abstract

Interleukin-2 (IL-2) based immunotherapy has been already approved to treat certain type of cancers as it plays vital role in immune system. Thus it is important to discover new peptides or epitopes that can induce IL-2 with high efficiency. We analyzed experimentally validated IL-2 inducing and non-inducing peptides and observed differ in average amino acid composition, motifs, length, and positional preference of amino acid residues at the N- and C-terminus. In this study, 2528 IL-2 inducing and 2104 non-IL-2 inducing peptides have been used for traning, testing, traing and validation of our models. A large number of machine learning techniques and around 10,000 peptide features have been used for developing prediction models. The Random Forest-based model using hybrid features achieved a maximum accuracy of 73.25%, with AUC of 0.73 on the training set; accuracy of 72.89% with AUC of 0.72 on validation dataset. A web-server IL2pred has been developed for predicting IL-2 inducing peptides, scanning IL-inducing regions in a protein and designing IL-2 specific epitopes by ranking peptide analogs (https://webs.iiitd.edu.in/raghava/il2pred/).

Keywords: Interleukin-2, Cancer immunotherapy; Prediction model; Peptide Analogs; IL-2 epitopes.

Introduction

The conventionally used cancer treatment strategies include chemotherapy, radiotherapy, and a combination of both. A mounting experimental and clinical outcome suggests that conventional therapies have many side effects ranging from patient discomfort to toxicity, development of resistance [1]. In this regard, the modulation of the body's immune system via immunotherapy is considered the safest and most advanced option for the treatment of cancer [2]. Recent clinical studies have confirmed the efficacy and safety of immunotherapy in managing large cohorts of patients [3]. Thus the use of immunotherapeutic agents becomes a staple of modern oncology treatment regimens [4]. Immunotherapy employs the use of therapeutic modalities such as interleukins, oncolytic viruses, and vaccines. Interleukins are small glycoproteins that can have a pleiotropic effect on the immune system. They bind to receptors on the cell surface that can aid in the differentiation, survival, and proliferation of immune cells [5].

Currently, two interleukins-based drugs are approved by the FDA for the treatment of human malignancies. This include interferon-alpha based drug Roferon-A and interleukin-2 (IL-2) based Aldesleukin or Proleukin. IL-2 is an early cytokine that was commercially approved to treat metastatic renal carcinoma and metastatic melanoma by the FDA. IL-2 is a 15.5 KDa alpha-helical cytokine that belongs to the gamma chain family of immune modulators. CD4+ T-cells chiefly produce IL-2, but they can also be made by CD8+ T-cells, Dendritic cells, and Natural killer cells (NK-cells) [6]. IL-2 acts via the JAK-STAT pathway and helps maintain and differentiate CD4+ T-cells in various subsets, including Th1, Th2, and Th17. It can promote CD8+ T-cell and NK-cell cytotoxic activity and in the generation of memory cells [7].

The greater immune-stimulating activity of IL-2 has led to the development of several treatment regimens. The IL-2 has been used in clinical settings as a monotherapy agent and also in

combination with several other therapeutics regimes. The low dose of IL-2 combined with interferon-alpha shows more significant therapeutic benefits for patients with renal cell carcinoma [8]. IL-2, in combination with activated killer cells, has been evaluated for the treatment of melanoma patients [9]. IL-2, when combined with chemotherapeutic agents, including cisplatin and dacarbazine has been extensively investigated in patients with metastatic melanoma [10]. The IL-2, combined with targeted therapies such as gefitinib, showed a positive response in NSCLC patients [11]. The combination of IL-2 with a therapeutic cancer vaccine also indicates a synergistic effect in advanced melanoma [12].

Undoubtedly, IL-2 showed great potential in treating cancers. However, its application in the clinic remains relatively restricted due to several shortcomings. The major limitation faced by IL-2 therapy is its toxicity when administered exogenously in high doses. The IL-2 also suffers from a short half-life of several minutes. It is observed that high doses of IL-2 sometimes lead to vascular leakage, hypotension [13]. The clinicians and researchers employed several strategies for overcoming the challenges faced by IL-2 therapy. One such approach is the generation of mutants of IL-2, which can activate immune cells without pro-inflammatory activity and thus free from toxicity-related problems [14]. The mutant version of IL-2 is known as "Superkine" because of its enhanced anti-tumor property.

These points highlight that the mutant version of natural IL-2 possesses a high therapeutic index. Thus the generation of an improved version of IL-2 has become an important area of research. However, the identification and generation of mutants in clinical setup is time-consuming and cost-intensive. The in-silico computation methods can help scientists and clinicians in this regard. Several methods have also been developed in the past that targets and utilized the therapeutic potential of interleukin-based therapy for the management of various human

malignancies. CytoPred is one such method that is developed for the identification and classification of cytokines [15]. In addition, methods have been developed for inducing specific type of cytokines, it include IL4pred [16], IL10Pred [17], IFNepitope [18], IL6pred [19] for IL-4, IL-10, IFN-γ gamma, IL-6 respectively. Despite the huge therapeutic potential, there is no method developed in the literature for the identification and generation of IL-2 inducing peptides. In the present study, we have attempted to develop a computational method for the prediction and identification of peptides that can possess the potential of inducing IL-2 cytokine. To serve the scientific community, the developed computational method has been incorporated in web-server, which is freely available at https://webs.iiitd.edu.in/raghava/il2pred/.

Materials and Methods

Dataset Preparations and Pre-processing

We extracted all the experimentally validated IL-2 inducing and non-IL-2 inducing peptides from the largest repository of Immune epitope (IEDB). We specifically extracted a total of 5427 peptides showing MHC class-II binding affinity and that can trigger IL-2 secretion as measured by different immunological assays. These epitopes were termed IL-2 inducing peptides and grouped under a positive dataset. The negative dataset consists of a total of 3568 epitopes that do not trigger the IL-2 secretion and are termed as non-inducer. All the identified positive and negative datasets are considered only of human host species. Literature evidence suggests that peptide of length between 8-25 is most suitable for MHC antigen processing and presentation. Thus we have removed all peptides having lengths below 8 and above 25. We also removed all the redundant peptides and finally left with 2528 IL-2 inducing and 2104 non-IL-2 inducing peptides. These 2528 peptides and 2104 peptides constitute our final dataset for further analysis.

Length Distribution and Residue Conservation Analysis

The positive and negative dataset was analyzed for preferred peptide length distribution using an in-house developed R script. We computed the average composition of different amino acids present in both positive and negative datasets. The two-sample logo was generated using the R package "ggseqlogo" for both datasets to understand the preference of specific amino acid residue at a particular position [20]. This package processes only fixed-length peptide sequences as input, and since the smallest length peptide is eight residues long, we considered the combination of 8-residue length sequences from both N and C terminals for plotting the two-sample logo. The following example shows the protocol for generating the sequence of 16 residues long peptides.

Peptide Sequence (original): ARGCGHTRKL

N-terminus 8-residues (N->C): ARGCGHTR [1:8]

C-terminus 8-residues (C->N): LKRTHGCG [10:3]

Peptide of 16-residues: N-terminus(N->C) + C-terminus residues (C->N)

Single peptide of 16-residues: ARGCGHTRLKRTHGCG

The first eight positions of the logo represent the N-terminus of peptides, and the last eight residue positions represent the C-terminus of the peptide. We have used all the peptides presents in our positive and negative dataset for creating the two-sample logo.

Motif Analysis

Identification of the motifs within peptides is very crucial in annotating the function of these peptides. We have MERCI software for the identification of specific motifs in both positive and negative datasets. We implemented the MERCI software in two steps; firstly, we extracted the motifs for a positive dataset by providing the IL-2 inducing peptides as a positive set and non-IL-

2 inducing peptides as a negative set. In the second iteration, we retrieved the motifs for the negative dataset by inputting non-IL-2 inducing peptides as a positive set and IL-2 inducing peptides as a negative set. Using the above-mentioned approach, we calculated the motifs in both positive and negative datasets. We have used two MERCI motif analyses using NONE classification for the identification of motifs in a mutually exclusive manner. We have screened peptides containing unique motifs from both sets to have an idea of the overall coverage of the various motifs in the complete data.

Feature Estimation and Selection

We have calculated the different features of peptides present in both positive and negative datasets. These calculated features are employed in the development of machine learning-based prediction method development. For the feature generation, we have used the Pfeature web-server that can calculate thousands of peptide features [21]. These include Composition based descriptors - Amino Acid Composition (AAC), Dipeptide Composition (DPC), Tripeptide Composition (TPC), Autocorrelation, Conjoint Triad Descriptors (CTDs), and Composition enhanced Transition and Distribution (CeTD); II) Binary profiles – Amino acids, Atom and Bonds, and Residue properties; III) Structural descriptors – SMILES, Surface Accessibility and Secondary; IV) Pattern-based descriptors – Binary profiles, PSSM profiles, and Universal properties from the peptides. With the help of Pfeature we have calculated approximately 10,000 peptide descriptors for both positive and negative datasets. As literature evidence suggests that not all descriptors are good for developing machine learning models. Thus, to remove the unwanted features we have employed the random forest feature selection function available in the R caret package. All the features mentioned above were tested independently as well as in

combination with other features after the feature selection step, for their classification capabilities among IL-2 inducing and non-inducing peptides.

Machine Learning-based Classification Model

We have implemented various machine learning-based algorithms for the classification of IL-2 inducing and non-IL-2 inducing peptides. We have specifically used the caret package of R for the classification and regression. The classification and regression algorithm includes - decision trees (DT), random forest (RF), multi-layer perceptron (MLP), eXtreme gradient boosting (XGBoost), K-nearest neighbors (KNN), support random vector with radial basis (SVR), neural network (NN), Ridge, Lasso, and Elastic Net". Different parameters were optimized using "expand Grid" functionality of the "caret" package of R. The DT algorithm was based on the non-parametric supervised algorithm; RF is an ensemble-based method which fits numerous decision tree to predict the outcome of the dependent variable; KNN is an instance-based learning algorithm and XGBoost is a tree boosting classification algorithm based on iterative search approach for making the final prediction.

Cross-validation

For developing the machine learning-based model we have used the standard protocols as used in different previous studies [22]. We divided our dataset in an 80:20 proportion as the training and testing sets, respectively, similar to several studies in the past. Training data comprises 2020 IL-2 inducing peptides and 1683 non-IL-2 inducing peptides, whereas the testing set had 505 IL-2 inducing peptides and 420 non-IL-2 inducing peptides. The different classifiers were trained and evaluated using a ten-fold cross-validation method, which is a well-accepted technique for the optimization of parameters and performance of the models. In ten-fold cross-validation, the whole training dataset was divided into ten equal parts, where iteratively nine parts were used in

training and one in the validation, to obtain the optimized parameters for the models. This process was repeated ten times to ensure that every set has been used in the training and validation. All these classifiers were implemented using an in-house R script. The performance of the developed model was evaluated using metrics such as specificity, sensitivity, accuracy, Matthews's correlation coefficient (MCC), and area under the curve (AUC). We have explained some of the important performance-evaluation parameters below:

$$\text{Sensitivity} = \frac{TP}{(TP+FN)}$$

$$\text{Specificity} = \frac{TN}{(TN+FP)}$$

$$\text{Accuracy} = \frac{TP+TN}{(TP+TN+FP+FN)}$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TN+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

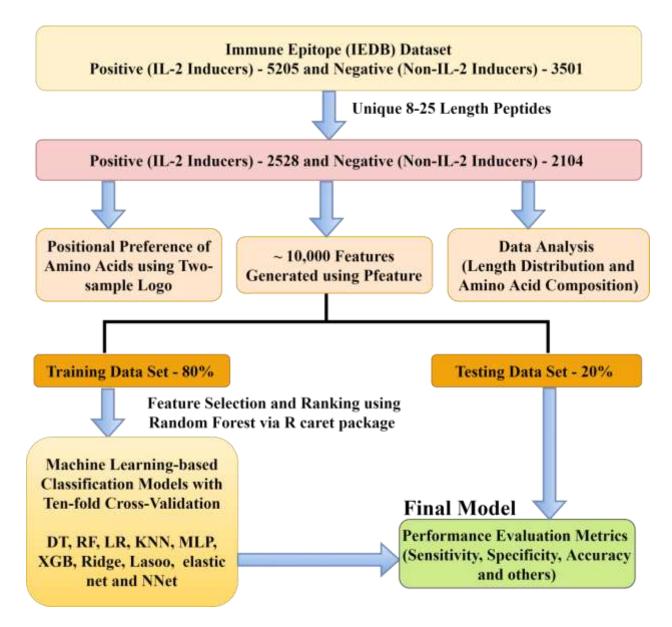We have explained the overall workflow of the study in **Figure 1.1.**

Figure 1.1: Schematic representation of workflow of the study

Architecture of Web-server

A web-server named "IL2Pred" (https://webs.iiitd.edu.in/raghava/il2pred) is developed to predict

the IL-2 inducing and non-inducing peptides. The front end of the web server was developed by

using HTML5, JAVA, CSS3, and PHP scripts. It is based on responsive templates that adjust the

screen based on the size of the device. It is compatible with almost all modern devices such as mobile, tablet, Mac, and desktop.

Results

In this study, we have utilized the 2528 IL-2 inducing peptides and 2104 non-IL-2 inducing peptides for developing the prediction algorithm. All the analysis and model building was done on these datasets. Based on the overall analysis, different prediction models were built, trained, and tested for their performance in predicting IL-2 inducing and IL-2 non-inducing peptides.

Datasets Length Distribution Analysis

We have analyzed both positive and negative datasets for length distribution analysis. The distribution analysis reveals that positive/IL-2 inducing peptides predominately consists of peptide length ranging from 12-20 while 15 lengths are most abundant. Whereas in negative/non-IL-2 inducing datasets, the peptide comprises 15, 17, and 20 length amino-acid sequences. The length distribution analysis of both the positive and negative datasets is provided in **Figure 1.2.**
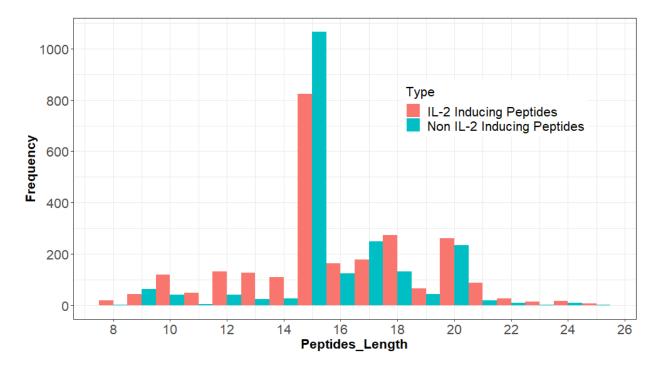
Figure 1.2: Visual representation of the frequency distribution of peptides length in the dataset.

Average Amino Acid Composition Analysis

We have also analyzed both positive and negative datasets for their average amino acid composition analysis. In the positive dataset, the amino acid residue A, L, S, and Y are predominant while in the negative dataset residues D and M have higher composition. The average amino acid composition analysis is provided in **Figure 1.3** along with the p-value, value of significance**.**
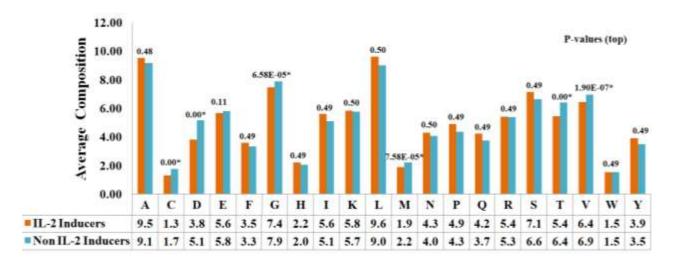
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
IL-2 Inducers	9.5	1.3	3.8	5.6	3.5	7.4	2.2	5.6	5.8	9.6	1.9	4.3	4.9	4.2	5.4	7.1	5.4	6.4	1.5	3.9
Non IL-2 Inducers	9.1	1.7	5.1	5.8	3.3	7.9	2.0	5.1	5.7	9.0	2.2	4.0	4.3	3.7	5.3	6.6	6.4	6.9	1.5	3.5

Figure 1.3: Bar plot of the average single amino acid composition of the peptides in the IL-2 inducers and non-IL2 inducers.

Positional Preference Analysis

In this analysis, we have generated the two-sample logo of both positive and negative datasets. This analysis was done to investigate the positional preference of specific amino acid residues in both positive and negative datasets. The two sample logo analysis reveals that amino acid residues G and S are predominately preferential at 16th and 11th position in the IL-2 inducing peptides. It means the residues G and S are predominant at the C-terminus of the IL-2 inducing peptides. Similarly, amino acid residues such as A and L are most predominant in non-IL-2 inducing peptides at both C and N-terminus end. Overall we conclude that polar residues are more prominent in the positive datasets, and hydrophobic ones are the abundant ones in the negative dataset. The results of amino acid positional preference analysis are provided in **Figure 1.4.**
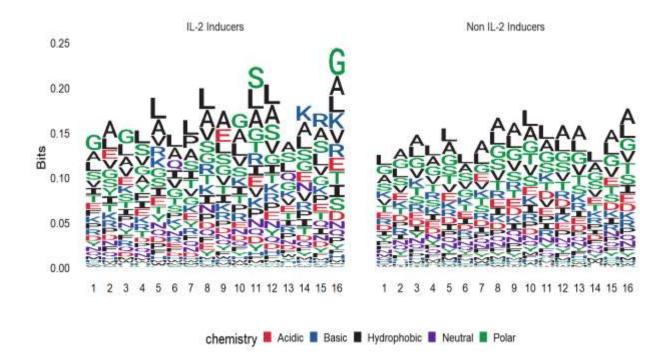
Figure 1.4: Two-sample logo depicting the relative abundance of single amino-acid residues at different positions in the positive and negative datasets.

Motif Analysis

We have utilized MERCI software to mine the motifs present exclusively in the IL-2 inducing peptides but not in non-IL-2 inducing peptides. Similarly, we have computed the motifs that are exclusive to non-IL-2 inducing peptides. The motif analysis reveals that A, L, S, and G are more prominent residues in the positive dataset, which was also found from the two-sample logo analysis. The results of motif analysis were presented in **Table 1.1.** These motifs have an odd ratio of infinity since are found in a mutually exclusive manner i.e. motifs found in one set are not fund in the other set. This explains the usage of these motifs as unique identifies in identifying IL-2 inducers and non-IL2 inducers.

Table 1.1: Motifs exclusive to interleukin-2 inducing and non-inducing peptides along with the frequency of their occurrence in the peptides searched using MERCI software.

MERCI MOTIF NONE			
IL-2 Inducers - Motifs	**Freq**	**Non-IL-2 Inducers - Motifs**	**Freq**
L-E-G-S	99	D-C-L	19
A-L-E-G	98	K-D-C	19
A-L-E-G-S	97	C-L-F	18
E-G-S-L-Q	96	E-A-Y-F	18
L-A-L-E-G	96	F-K-D-C	18
S-L-Q-K	96	L-E-A-Y	18
-		Y-F-K-D	18

Development of Prediction Models

We have developed various machine-learning-based prediction models such as DT, RF, KNN, MLP, Ridge, Lasso, ElasticNet. Firstly, we have computed the features of IL-2 and non-IL-2 inducing peptides using the Pfeature. The feature selection was done by using SVC-LI through random forest functionality available in the "caret" package of R. Apart from these selected features, classification models were built and tested separately on independent features that were widely followed in several past studies [17]. With the feature selection process, we end up with 10 and 100 most suitable relevant features. With these feature sets, we built different machine-learning algorithms. The statistical detail of each feature along with the best classifiers has been provided in **Table 1.2.** It can be observed from Table 1.2 that the DPC-based RF model performed well in classifying the IL-2 inducing peptides versus non-IL-2 inducing peptides with overall accuracy, MCC, and AUC of 71.29%, 0.42, and 0.71, respectively.

Table 1.2: Show the performance of different classifiers developed using various type of peptide features; only best models are shown.

Features [Classifier]	Sensitivity (%)	Sensitivity (%)	Accuracy (%)	MCC	AUC
Training					
AAC [RF]	74.70	64.13	69.90	0.39	0.69
DPC [RF]	**80.07**	**60.72**	**71.29**	**0.42**	**0.71**
PCP [RF]	70.05	57.42	64.32	0.27	0.64
Adv PCP [RF]	69.96	44.66	58.46	0.15	0.58
Struct PCP [XGB]	66.11	56.23	61.62	0.22	0.61
RRI [RF]	76.43	51.60	65.16	0.29	0.65
CeTD [XGB]	67.34	56.41	62.38	0.24	0.62
DOOR [RF]	70.15	54.81	63.18	0.25	0.63
Pfeature_top10 [RF]	67.63	54.04	61.46	0.22	0.61
Pfeature_top100 [RF]	77.56	57.48	68.45	0.35	0.68
Testing					
AAC [RF]	72.72	63.33	68.47	0.36	0.68
DPC [RF]	**78.81**	**60.95**	**70.70**	**0.40**	**0.70**
PCP [RF]	70.94	55.23	63.82	0.26	0.63
Adv PCP [RF]	72.52	40.95	58.21	0.14	0.57
Struct PCP [XGB]	59.28	56.19	57.88	0.15	0.57

RRI [RF]	74.30	51.90	64.14	0.26	0.63
CeTD [XGB]	70.94	52.85	62.74	0.24	0.62
DOOR [RF]	68.37	57.14	63.28	0.25	0.63
Pfeature_top10 [RF]	69.16	53.09	61.87	0.22	0.61
Pfeature_top100 [RF]	79.44	52.62	67.28	0.33	0.67

Bold faced are the best models

The AAC-based RF model also performs comparably to the DPC-based method. From this analysis, it is visible the importance of amino acid composition in the classification of IL-2 inducing and non-IL-2 inducing peptides.

Models based on Hybrid Features

Length of peptide plays a crucial role in MHC antigen presentation and thus in the induction of IL-2 inducing peptides. Thus we have developed models using hybrid features that combine DPC and length of the peptide. The statistical details of the different models using various classifiers have been presented in **Table 1.3.** Our RF based model developed using hybrid features achieve maximum accuracy of 73.25%, with AUC of 0.73 on the training set.

Table 1.3: The performance of different classifies developed using hybrid features (DPC and length of peptides).

Classifiers	Sensitivity (%)	Sensitivity (%)	Accuracy (%)	MCC	AUC
Training					
DT	55.28	66.80	60.52	0.22	0.61

RF	**74.60**	**71.61**	**73.25**	**0.46**	**0.73**
KNN	72.43	64.96	69.04	0.37	0.69
MLP	64.77	60.27	62.79	0.25	0.62
SVR	72.97	59.02	66.63	0.32	0.66
XGB	73.12	64.20	69.07	0.37	0.68
LASSO	67.69	53.92	61.43	0.22	0.61
Testing					
DT	58.30	68.09	62.74	0.26	0.63
RF	**74.11**	**71.42**	**72.89**	**0.45**	**0.72**
KNN	73.91	64.76	69.76	0.39	0.69
MLP	62.84	65.47	64.03	0.28	0.64
SVR	73.91	58.57	66.95	0.33	0.67
XGB	73.12	66.67	70.19	0.39	0.69
LASSO	66.99	52.61	60.47	0.20	0.60

*Bold faced are the best models

The robustness of a developed model is best calculated on an independent dataset. We have tested the performance of the developed model on the 20% dataset which we have not used in the training process. The model showed comparable performance on the testing set with an overall accuracy of 72.89% with AUC of 0.72.

Modules and Functionalities of IL2pred Server

This web server is compatible with all sort of devices viz. Desktop, tablets, and phones and hence provide an interactive and better experience to the users. The webserver has three main

modules – 1) Prediction; 2) Analogue; 3) Protein Scan. The "Prediction module" allows users to identify the peptides that have IL-2 inducing potential; results can be can be obtained in CSV format. The "Analogue module" module provides users with the opportunity to generate all possible mutants of given query peptide sequence and then to rank them based on their IL-2 inducing potential. "Protein Scan" module allow users to predict IL-2 inducing regions within a given protein sequence. The user-friendly web server is available at https://webs.iiitd.edu.in/raghava/il2pred.

Discussion

The high-throughput studies have significantly enriched our knowledge regarding the tumor onset to progression. However, the survival of patients is dismal as a whole, and finding novel therapeutics remains a major challenge. Recently, evidence from pre-clinical and clinical studies advocates the activation of the immune system for the treatment of human malignancies [5]. Immunotherapy based on cytokines is considered a novel means of activating and manipulating the immune system in the fight against cancer. IL-2 is one such cytokine that can boost the antitumor immune response by maintaining and promoting the activity of CD4+ T-cell and cytotoxic CD8+ T-cells. Thus, the identification of peptides or the antigenic regions that can boost the production of IL-2 would become an important consideration while designing cancer treatment regimens and vaccine candidates.

The present study is a systematic attempt to make an in-silico model for predicting the IL-2 inducing capability of the peptide/antigenic region. The in-silico model is developed on the sequence features of non-redundant, IL-2 inducing and non-inducing peptides extracted from the largest repository of experimentally validated immune epitopes database. The non-redundant dataset ensures that no overfitting/bias/noise can affect the developed model due to the presence

of multiple instances of the same peptide. Literature evidence suggests that the length of the peptide can affect its binding to the MHC complex and thus immune activation [23]. The length distribution analysis reveals that IL-2 inducing peptides predominately consists of length ranging from 12-20 while for non-inducer it varies from 15-17. The compositional analysis reveals that IL-2 inducing peptides are enriched in A, L, S, and Y residues. Past studies also reveal that peptides enriched in S, L and A residues have increased anti-tumor capabilities [24] as they can induce apoptosis in cancer cells via modulation of autophagy [25]. Thus, we can suggest that sequence composition can also be used in the discrimination of IL-2 inducing peptides from the non-IL-2 inducing peptides to some extent.

In addition to the above, the development of the IL-2 prediction models was also carried out using several other sequence features on multiple machine learning-based algorithms. It is observed that DPC based RF machine-learning model can discriminate the IL-2 inducing and non-inducing peptides with good accuracy. However, the hybrid model developed on DPC and length of peptides performed well than the DPC-based model alone in terms of balanced model performance measures. This may be due to the increased vector size while developing the prediction model. The external validation of the developed model further ensures that the performance of the model is not due to over-optimization or over-fitting. The best-performing machine-learning-based model was incorporated in the web server. We believe that the developed tool will help the scientific community in better understanding IL-2 inducing/non-inducing peptides. However, with the increased number of organism-specific datasets, the applicability of the developed model can further be improved. We anticipated that the developed tools may be of great use to the scientific community for identifying IL-2 inducing/non-inducing peptides from the proteomes for improving the current cancer immunotherapeutics. Also, the

developed tools can be used for generating the mutant peptides that are predicted to have IL-2 inducing potential.

Conclusion

Past studies strongly reveal that direct administration of IL-2 may have several toxic effects. The capability of an antigenic region or mutant peptide to induce the IL-2 response is of great significance for activating the immune response towards malignant tissue. However, identifying such mutant peptides with conventional experimental techniques is time-consuming and costly. Therefore, a computational tool in the form of a web server is provided to the scientific community for predicting and identifying the IL-2 inducing/non-inducing peptides from the given proteomes. The web server is freely available to the scientific community at https://webs.iiitd.edu.in/raghava/il2pred/. We hypothesize that the developed tools will be highly utilized by the scientific community for identifying and prioritizing the potential immunotherapeutic candidates

Author's Contribution

Anjali Lathwal: Conceptualization, Methodology, Validation, Formal analysis, Web server implementation, Software and Coding, Investigation, Data curation, Writing original draft, Reviewing and editing original draft. **Rajesh Kumar:** Conceptualization, Methodology, Validation, Formal analysis, Web server implementation, Software and Coding, Investigation, Data curation, Writing original draft, Reviewing and editing original draft. **Dilraj Kaur:** Web server implementation, Reviewing and editing original draft. **Gajendra Pal Singh Raghava:**

Supervision, Conceptualization, Project administration, Reviewing and editing original draft, Investigation, Validation.

Funding

Data Availability

All the data is freely available on IL2pred website in the download section.

Declaration of competing interest

There is no potential conflict of interest among the authors of the manuscript.

Acknowledgment

References

1. Lettieri-Barbato D, Aquilano K. Pushing the limits of cancer therapy: The nutrient game. Front. Oncol. 2018; 8:

2. Lathwal A, Kumar R, Raghava GPS. OvirusTdb: A database of oncolytic viruses for the advancement of therapeutics in cancer. Virology 2020; 548:109–116

3. Kruger S, Ilmer M, Kobold S, et al. Advances in cancer immunotherapy 2019 - Latest trends. J. Exp. Clin. Cancer Res. 2019; 38:

4. D'Errico G, Machado HL, Sainz B. A current perspective on cancer immune therapy: step-by-step approach to constructing the magic bullet. Clin. Transl. Med. 2017; 6:

5. Jiang T, Zhou C, Ren S. Role of IL-2 in cancer immunotherapy. Oncoimmunology 2016; 5:

6. . Simultaneous production of IL-2, IL-4, and IFN-gamma by activated human CD4+ and CD8+ T cell clones - PubMed.

7. Boyman O, Sprent J. The role of interleukin-2 during homeostasis and activation of the immune system. Nat. Rev. Immunol. 2012; 12:180–190

8. Atkins MB, Sparano J, Fisher RI, et al. Randomized phase II trial of high-dose interleukin-2 either alone or in combination with interferon Alfa-2b in advanced renal cell carcinoma. J. Clin. Oncol. 1993; 11:661–670

9. Restifo NP, Dudley ME, Rosenberg SA. Adoptive immunotherapy for cancer: Harnessing the T cell response. Nat. Rev. Immunol. 2012; 12:269–281

10. Atkins MB. Cytokine-based therapy and biochemotherapy for advanced melanoma. Clin. Cancer Res. 2006; 12:

11. Bersanelli M, Buti S, Camisa R, et al. Gefitinib plus interleukin-2 in advanced non-small cell lung cancer patients previously treated with chemotherapy. Cancers (Basel). 2014; 6:2035–2048

12. Smith FO, Downey SG, Klapper JA, et al. Treatment of metastatic melanoma using Interleukin-2 alone or in conjunction with vaccines. Clin. Cancer Res. 2008; 14:5610–5618

13. . Role of asialo-GM1-positive lymphoid cells in mediating the toxic effects of recombinant IL-2 in mice - PubMed.

14. Heaton KM, Grimm EA. Differential inhibition of lymphokine-activated killing, proliferation, and cytokine secretion by humanized antibodies against the low- and intermediate-affinity interleukin-2 receptors. A novel model for activation of human peripheral blood mononuclear cells by interleukin 2. Hum. Immunol. 1995; 42:274–280

15. Lata S, Raghava GPS. CytoPred: A server for prediction and classification of cytokines. Protein Eng. Des. Sel. 2008; 21:279–282

16. Dhanda SK, Gupta S, Vir P, et al. Prediction of IL4 inducing peptides. Clin. Dev. Immunol. 2013; 2013:263952

17. Nagpal G, Usmani SS, Dhanda SK, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. Sci. Rep. 2017; 7:

18. Dhanda SK, Vir P, Raghava GPS. Designing of interferon-gamma inducing MHC class-II binders. Biol. Direct 2013; 8:

19. Dhall A, Patiyal S, Sharma N, et al. Computer-aided prediction and design of IL-6 inducing peptides: IL-6 plays a crucial role in COVID-19. Brief. Bioinform. 2021; 22:936–945

20. Wagih O. Ggseqlogo: A versatile R package for drawing sequence logos. Bioinformatics 2017; 33:3645–3647

21. Pande A, Patiyal S, Lathwal A, et al. Computing wide range of protein/peptide features from their sequence and structure. bioRxiv 2019; 599126

22. Lathwal A, Arora C, Raghava GPS. Prediction of risk scores for colorectal cancer patients from the concentration of proteins involved in mitochondrial apoptotic pathway. PLoS One 2019; 14:

23. Chang ST, Ghosh D, Kirschner DE, et al. Peptide length-based prediction of peptide - MHC class II binding. Bioinformatics 2006; 22:2761–2767

24. Tyagi A, Kapoor P, Kumar R, et al. In silico models for designing and discovering novel anticancer peptides. Sci. Rep. 2013; 3:

25. Zamudio-Vázquez R, Ivanova S, Moreno M, et al. A new quinoxaline-containing peptide induces apoptosis in cancer cells by autophagy modulation. Chem. Sci. 2015; 6:4537–4549