Journal of Convergence for Information Technology
Vol. 10. No. 6, pp. 209-216, 2020
e-ISSN 2586-4440
DOI : https://doi.org/10.22156/CS4SMB.2020.10.06.209

Prediction of HLA-A*0201-Restricted Antigenic Epitopes Targeting Multiple Myeloma

Yoon Joong Kang
Professor, Department of Biomedical Science, Jungwon University

다발성 골수종 적용을 위한 HLA-A*0201 제한 항원성 펩타이드 예측

강윤중
중원대학교 의생명과학과 교수

Abstract Protein antigens and their epitopes are targets for epitope based vaccines. There are many prediction servers which can be used for identification of binding peptides to MHC molecules. However, choosing of appropriate prediction servers is difficult. This study compared data obtained from prediction servers and evaluate them in scope of binding affinity to MHC-I molecules. Here we predicted HLA-A2-restricted cytotoxic T lymphocyte epitopes from survivin as a potential target for multiple myeloma. We suggest a procedure for prediction of antigenic peptides which could bind to MHC-I molecule. The results of this study will assist researchers in selection and prediction of noble antigenic peptides.

Key Words : Epitope, MHC-class Ⅰ, Peptide, Prediction, Multiple myeloma

요 약 단백질 항원에 존재하는 에피토프는 에피토프를 기반으로 한 백신 개발의 표적이 되고 있다. 인간의 주조직적합 복합체 (MHC-1)에 결합하는 펩타이드를 확인할 수 있는 여러 서버들이 보고되고 있으나 인간의 MHC-Ⅰ 분자의 수가 매우 많고 각 서버 검색 방법의 표준화 부재 등의 문제로 인해 펩타이드 예측에 적절한 서버를 선정하는 것이 쉽지 않다. 본 논문에서는 MHC-Ⅰ 결합 펩타이드를 예측하는 서버 30 종을 비교하였으며, 다발성 골수종에 적용하기 위해 survivin 단백질로부터 사람의 HLA-A2 제한 펩타이드를 예측하였다. 본 연구의 결과는 MHC-I 결합 예측의 표준화된 방법을 제시하고 펩타이드 에피토프를 예측하는데 도움을 줄 것이다.

주제어 : 에피토프, MHC-Ⅰ, 펩타이드, 예측, 다발성 골수종

1. Introduction

Vaccines have known to be effective tool to control infectious disease [1] and provide promising solutions for controlling infectious disease, cancer, allergies, and other immunologic diseases [2]. Antigenic peptide vaccines provide means for safe immune intervention; several peptide vaccines are currently under developing [3]. Most of these vaccines consist of various forms of peptide antigens. Different strategies of formulation as well as for delivery systems have been studied [4]. Antigenic epitopes, the basic immunogenic units within protein antigens, could be used to initiate immune responses [3]. The formulations of epitope vaccines include peptides, carbohydrates,

*Corresponding Author : Yoon Joong Kang(yjkang@jwu.ac.kr)
Received May 3, 2020
Revised May 28, 2020
Accepted June 20, 2020
Published June 28, 2020

epitope-coding nucleic acids or combinations of those.

Multiple myeloma is B cell malignancy, among the most immune-responsive of human cancers, which has lead to interest in the development of immunotherapy. Most immunotherapy protocols in development for multiple myeloma have used tumour specific protein or whole tumour cell lysates, which are very labour intensive and expensive. Thus, development of more effective immunotherapy is necessary. Among several potential effective immunotherapeutic approaches, peptide epitope based vaccines can successfully induce anti-tumour immune responses.

T cell epitopes means peptides that induce T cell mediated immune responses when bound by MHC molecules. Those T cell epitopes are presented on the cell surface for recognition. Epitope peptides are derived from degradation of internal proteins and bound to MHC-I molecules. Researchers have tried to develop multivalent vaccines that could induce efficient T cell priming, long-lasting and high magnitude of immune response [4]. Antigenic epitopes induce specific responses against foreign cells.

Identification of antigenic peptides helps improve our knowledge of immune specificity. It is key step for discovery of immuno-therapies [3,5,6]. Many protein variants have been identified in viruses and bacteria. Lots of tumor antigens and their variants have been studied [7,8]. Currently, several thousand variants of human HLA have been identified [9]. Due to the large number of antigens and HLA variants, testing of binding capacity of these antigenic peptides is hard to achieve. Therefore, many computational methods have been developed to identify MHC binding peptides [10,11]. More than 30 prediction servers have been generated. These computational screening of antigens for epitopes becomes a standard approach in epitope mining researches. But There are so many prediction servers available, new challenge have attract our notice: selection of the better server; can they be used to predict binding affinity of antigenis peptides? Because there is no standards to develop MHC-I binding prediction servers, we have encountered many servers that have differences in values of predictions. Some comparing methods for prediction of antigenic peptides have been reported [12-14]. Predictions of T cell antigenic epitopes are known to be less accurate. In a recent study, tens of candidate peptides were identified from computational screening of some thousands of peptides. Among those candidate peptides, 21 peptides were identified as T-cell epitopes and 17 peptides were identified as HLA-binders with high-affinity [15]. A new era of predictive methods that fuse predictions of several antigen processing and presentation procedures: HLA binding, peptide binding to transported associated with antigen processing (TAP) [16]. Combination of HLA predictions and TAP predictions improves peptide predictions in some cases [17-19]. Proteasomal cleavage predictions showed much lower accuracy than TAP-binding predictions or HLA-binding [16]. Meanwhile, predictions of HLA binding remain the most valuable computational methods for identifying of T-cell antigenic epitopes.

Peters et al. [13] reported a prediction on a data set consist of huge quantitative peptide-binding affinity. Other researchers compared the ability of prediction servers and combined data [14]. Because there is no appropriate independent test sets, the comparison studies done to date have been based on verifying predictive performance using known sets of peptides, rather than studies of complete full length antigens. In this

study we compared the ability of several servers by normalizing the data to a general scale. These antigenic peptides were derived from a tumor related antigen and from a part of a virus antigen. We compared those servers to find if any of them produce identical data. The main part of the study explored the prediction of HLA binding peptides and binding affinity. We assessed their prediction abilities on two sets of well known T-cell epitopes.

2. Methods

We chose thirty servers which can predict HLA-I binding peptides. The study consisted of several steps: a) test data sets which were independent measured were identified; b) antigenic peptide binding predictions were made; c) the predictions of each servers were compared; d) data were normalized to the general scale; e) estimation of classification was made; f) assessment of the accuracy of binding affinities.

2.1 Data sets

We used data sets generated by the iTopia™ Epitope Discovery System. Those data sets consisted of the data of 134 peptides which overlap the full length of tumor antigen survivin (Swiss-Prot: O15392) and the 42 peptides from cytomegalovirus (CMV) internal matrix protein pp65.

HLA-A2 restricted antigenic epitopes had been obtained from the literature. Some prediction servers did not have specific information on alleles. For example, the results extracted by SMM are binding affinities of peptides to HLA-A2.

2.2 Predictions and comparisons

Those two protein sequences were submitted to each predictors and the data were recorded. For each HLA molecule two experimental applications were performed: identification of binders and prediction of binding affinity. To assess the accuracy, we used the area under the ROC curve (AROC) [20]. The curve was a plot of the real positive rate on the Y axis vs. false positive rate on the X axis. The values AROC≥ 0.9 meaned excellent, 0.9〉AROC≥0.8 meaned good, 0.8〉AROC≥0.7 meaned marginal and 0.7〉AROC indicated poor predictions [21].

3. Results

3.1 Classification binding peptides

Because all of these servers were not designed for peptide binding predictions, all of them have antigenic peptide binding predictions implemented. For example, MAPPP and ProPred1 perform prediction of multiple steps of antigen processing, MULTIPRE shows data of peptide binding to HLA supertypes, and BIMAS shows predictions of peptide binding as well as dissociation. Several servers have other options, for instance, MHCPred tells anchor positions. We used the most simple prediction method available at each predictor. After doing all predictions, we obtained the Pearson correlation coefficient for all the servers and found out that MAPPP (SYFPEITHI) and MAPPP (BIMAS) showed identical results (r=1) to SYFPEITHI and BIMAS, respectively. The BIMAS and ProPred1 results showed r ≥ 0.998. This was expected result because SYFPEITHI and BIMAS were adopted by MAPPP servers. The numbers of the prediction servers we assessed was twenty-seven for A*0201. The analysis of prediction servers by calculating correlation coefficient tells that these servers are independent, and shows different subsets of HLA-I binding antigenic peptides.

The accuracy of binder classification was

assessed using the cutoff of 30 (measured binding affinity of ≥ 30% of the binding affinity of a positive control) for binders. As a results, we had 39 excellent servers , 47 good servers, 33 marginal servers, and 28 poor servers. The AROC values used in these predictions are shown in Fig. 1.

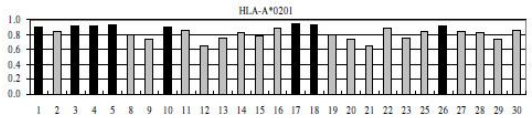

Fig. 1. AROC values used in predictions. Black bars means servers showing the best ability. Y axes show the value of AROC while X axes show individual prediction servers

We also done the analysis of survivin and CMV construct test sets. The results shown in Figures 2 and 3 were very similar to the previous result. About 30% of servers for HLA-A2 showed result of excellent classification.

The best predictors in this result was NETM_ANN, followed by IEDB_SMM and IEDB_ANN also showed good predictions. The best servers we recommend for binder/non-binder classification prediction are shown in Fig. 1 (black bars).

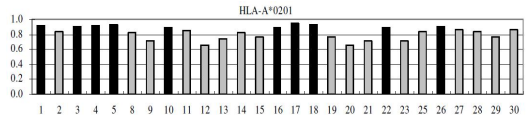

Fig. 2. Predictions using AROC values with survivin test set. Black bars means servers showing the best prediction ability. Y axes show the value of AROC and X axes means each servers.

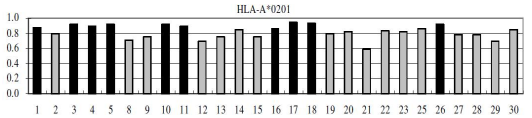

Fig. 3. Predictions using AROC values with CMV test set. Black bars means servers showing the best prediction ability. Y axes show the value of AROC and X axes means each servers.

3.2 Prediction of peptide binding affinity

Prediction values from servers represent a number of measures. Experimental data from the iTopia™ are expressed as the peptide concentration needed for 50% binding. Then the data compared as percentage of binding affinity relative to the control. For instance, IEDB and NETM_ANN servers represent binding affinity on a nanomolar scales, the binding value for BIMAS shows off-rates, MHC I predictor shows "binding energy", while MULTIPRED predictor tells an arbitrary binding score. Large differences are noticed even between predictors from the same prediction server. For instance, the survivin 1.9 peptide MGAPTLPPA is an candidate binder to A2 with estimated 94% binding affinity relative to the control. The respective binding predictions for IEDB_ANN, IEDB_SMM, and IEDB_ARB, are 23441, 3237, and 365 nM, while NETM_ANN predicted score is 8574nM. Across all server predictors a various scales of prediction values have been noticed. Apparently these server predictors have to be treated as different *in silico* assessments and the comparison could be made only by application of relative scales. By using iTopia™ binding assay as the experimental control, we obtained correlation coefficients for all available prediction servers for CMV construct, survivin, and the combined data set. The data represents that a accurate prediction of peptide binding affinity could be achieved for HLA-A2 as shown in Figure 4. NETM_ANN and IEDB_ANN show values of $r > 0.8$ while A*0201 predictors SYFPEITHI, HCI_MM,N ETM_WM, MULTI_SVM, and MULTI_ANN showed a relatively high correlation coefficient of $0.8 < r < 0.7$. Overall, the best peptide binding predictors are NETM_ANN and IEDB_ANN. Using this procedure, peptide binding affinities to HLA-A2

could be predicted and identified *in silico*. The best peptide binding predictors recommended for prediction of antigenic peptides which bind to HLA-A2 molecule are marked with asterisks in Fig. 4.

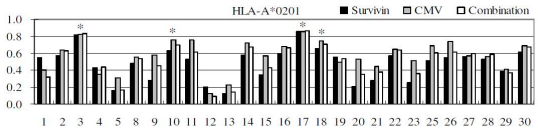

Fig. 4. The correlation coefficients of prediction servers for survivin, CMV, and combined data sets. Black bars represents survivin, gray bars represents CMV construct, and white bars represents the combined set. Y axis shows the value of correlation coefficients, and X axis shows each servers.

The peptide binding prediction results within survivin, CMV, and combined data sets show apparent consistency for indicating that the most prediction servers predict well among different data sets. In most predictors, the peptide prediction accuracy for survivin was somewhat lower than that of CMV construct, while the predictions of the combined data set were mostly higher. The prediction server BIMAS showed low stability in this test, while recommended prediction servers show high level of consistency to perform prediction across the three different test sets.

3.3 Prediction of T cell antigenic epitopes

We conducted prediction of peptide binding with tumor related antigenic epitope and viral epitope sets. Both tumor related antigenic epitope and viral epitope sets showed patterns of similar prediction. Then we proceeded analysis with merged data sets. For each prediction server, we calculated the binding affinity of all epitopes in the merged set. We determined the threshold at which approximately 90% of the tested antigenic epitopes were predicted as binders. We also measured the threshold at which the first false positive appears. The predictor with higher thresholds was used for further analysis.

Prediction servers can be used for different experimental purposes. We compared the ability of servers in three scopes for each server and the representative data are shown in Tables 1, 2 and 3. These scenarios are shown by the selection of thresholds. The first scenario is the selection of threshold at which ~90% of antigenic epitopes are predicted as binders. In the second scenario, threshold predicts almost correctly the majority of binders, 31 out of 33 peptides. In the third scenario, threshold does not show any non-binding peptides to be predicted as binding peptides. The resulting data apparently tell that the selection of the best prediction server depend on each experimental purpose. For instance, NETM_ANN has been selected as the best server overall HLA-A2 predictor as shown in Figure 1 and Figure 4. This prediction server also represents the best performance for thresholds that could optimize the T cell antigenic epitope selection as shown in Table 1 and the threshold which does not bear false positive as shown in Table 2 and 3. As we can see in Table 2, the distinct best prediction server with highly sensitive threshold is NHP_CP which overall performance has been represented as modest. In summary, considering the balance between false positive and false negative and prediction of antigenic T cell epitopes, NETM_ANN is seems to be the best candidate in most cases. The higher sensitivity of peptide prediction, the larger number of false positives. In other way, with higher specificity of prediction the lower number of true positives peptides.

Overall, our results have represented that the best prediction servers of classification also show the best ability in prediction of HLA-A2

binding affinity. Prediction of T cell antigenic epitopes are more seems to be followed to the peptides with highest binding affinity [15,21].

Table 1. Prediction ability of selected representative servers in order to predict T cell peptide epitopes

Server	Thr1	TP (binding)	TN (Binding)	FP (Binding)	FN (Binding)
BIMAS (A)	2	10	143	0	23
MHCP_1 (A)	100	31	7	136	2
IEDB_SMM (B)	1,000	10	143	0	23
NHP_CP (C)	0	31	126	17	2
IEDB_ANN (D)	10,000	9	143	0	24
MULTI_SVM (D)	5.5	6	141	2	27
NETM_ANN (D)	10,000	15	143	0	18

Table 2. Prediction ability of selected representative servers in order to predict the majority of binding peptides

Server	Thr2	TP (binding)	TN (Binding)	FP (Binding)	FN (Binding)
BIMAS (A)	0.003	31	105	38	2
MHCP_1 (A)	1,000	31	7	136	2
IEDB_SMM (B)	79,000	31	109	34	2
NHP_CP (C)	0	31	126	17	2
IEDB_ANN (D)	39,000	31	75	68	2
MULTI_SVM (D)	3.9	31	101	42	2
NETM_ANN (D)	40,000	31	113	30	2

Table 3. Prediction ability of selected representative servers in order to exclude false positives peptides

Server	Thr3	TP (binding)	TN (Binding)	FP (Binding)	FN (Binding)
BIMAS (A)	2	10	143	0	23
MHCP_1 (A)	10	0	143	0	33
IEDB_SMM (B)	1,000	10	143	0	23
NHP_CP (C)	0.5	6	143	0	27
IEDB_ANN (D)	10,000	9	143	0	24
MULTI_SVM (D)	5.8	4	143	0	29
NETM_ANN (D)	10,000	15	143	0	18

4. Discussion

Although improvement in therapeutic strategies for treating multiple myeloma, novel effective therapeutic approaches are urgently needed. Cancer peptide vaccines or T cell-based immunotherapy suggest promising treatment options to induce anti-tumour immunity. An promising therapeutic method is to use immunogenic peptides, which would have ease of production, low toxicity, broader applications, specificity for tumour cell targets. Efficacy of peptide-based immunotherapy could be further enhanced by targeting multiple antigens on tumour cells, thereby achieving epitope spreading in the patients. However, there are very few reports in multiple myeloma peptide vaccines.

This study represents major outcomes that have recently been made in the area of computational immunology and applicable for development of peptide vaccines against multiple myeloma. These are mainly the results of the collaborative efforts that concentrate on the accumulation of computational data for immunology, such as ImmunoGrid or IEDB. The development of large data sets of antigenic T-cell epitopes and enhanced algorithms enabled the establishment of *in silico* methodology that assist practical research.

We have studied key lessons about the algorithms that are used to show interactions between HLA molecule and peptide. Recently developed algorithms are usually work to be pursued, since *in silico* assessments that practical accuracy are available only for single HLA-A2 peptides. We have also noticed the problems of some prediction methods as shown in Figure 6. Group A predictors shows low sensitivity and can be improved by machine training with new data, particularly data of binding peptides. Group B shows low specificity and these servers could be improved by machine training with larger number of non-binders. Group C can be further ameliorated by machine retraining with larger

number of data, on the other hand group D can be enhanced by further improvement of algorithms. The combination of predictions with highly accurate peptide predictors is seems to be a right direction for amelioration of antigenic peptide predictions [14]. A large number of prediction servers, in particular, those from groups A and B can be enhanced by post-processing of raw peptide prediction data.

Our data also tells that normalization of output data sets by scaling onto a common scale would be helpful to the field by providing a standard *in silico* scale. In this way, the negative control could be mapped to 0, while the positive control mapped to 100. Binder peptides of higher affinity than the positive control peptide will have binding value larger than 100. The interpretation of the normalized values is clearer than the raw values in data as shown in Table 1, 2, 3. For those that belong to prediction server groups A, B, or C, it is quite hard to determine the best threshold because the threshold area between "good" and "poor" predictions is so narrow, compared to group D which has wide threshold area.

The study of computational immunology are rapidly growing field [13]. Combining practical and *in silico* methods is key step to solve complex problems associated with developing of peptide vaccines. While identification of antigenic T cell peptide epitopes is only a step in the whole process of peptide vaccine research into clinical applications, it is a promising sign of significant advances that can be used to develop noble antigenic peptide vaccine.

It is expected that the rapidly increasing power of computational techniques will have a tremendous impact on the unraveling of the antigenic peptide. With the identification of more T cell epitopes effectively with helps of computational research, we will in principle be able to develop effective immunotherapies that are applicable in treatment of B cell malignancies including multiple myeloma. Together with upcoming improvements in the immunogenic context of peptide vaccines and computational approaches this is expected to result in a significantly enhanced efficacy of anti-cancer immunotherapies.

REFERENCES

[1] J. Ehreth (2003). The value of vaccination: a global perspective. *Vaccine, 21(27-30)*, 4105-4117.
DOI : 10.1016/s0264-410x(03)00377-3

[2] V. Brusic, J. T. August & N. Petrovsky. (2005). Information technologies for vaccine research. *Expert Rev Vaccines, 4(3),* 407-417.
DOI : 10.1586/14760584.4.3.407

[3] A. W. Purcell, J. McCluskey & J. Rossjohn. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nat Rev Drug Discov, 6(5),* 404-414.
DOI : 10.1038/nrd2224

[4] P. Riedl, J. Reimann & R. Schirmbeck. (2006). Complexes of DNA vaccines with cationic, antigenic peptides are potent, polyvalent CD8(+) T-cell-stimulating immunogens. *Methods in molecular medicine, 127,* 159-169.
DOI : 10.1385/1-59745-168-1:159

[5] S. H. van der Burg, M. S. Bijker, M. J. Welters, R. Offringa & C. J. Melief. (2006). Improved peptide vaccine strategies, creating synthetic artificial infections to maximize immune efficacy. *Advanced drug delivery reviews, 58(8)*, 916-930.
DOI : 10.1016/j.addr.2005.11.003

[6] A. Muzzi, V. Masignani & R. Rappuoli. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials, *Drug Discov Today, 12(11.12)*, 429-439.
DOI : 10.1016/j.drudis.2007.04.008

[7] P. Van Der Bruggen et al. (2002). Tumor specific shared antigenic peptides recognized by human T cells. *Immunol Rev, 188*, 51-64.
DOI : 10.1034/j.1600-065X.2002.18806.x

[8] G. Parmiani, A. De Filippo, L. Novellino & C.

Castelli. (2007). Unique human tumor antigens: immunobiology and use in clinical trials. *J Immunol, 178(4)*, 1975-1979.
DOI : 10.4049/jimmunol.178.4.1975

[9] J. Robinson, M. J. Waller, S. C. Fail & S. G. Marsh (2006). The IMGT/HLA and IPD databases. *Hum Mutat, 27(12)*, 1192-1199.
DOI : 10.1093/nar/gkz950

[10] V. Brusic, V. B. Bajic & N. Petrovsky. (2004). Computational methods for prediction of T-cell epitopes. a framework for modelling testing, and applications. *Methods, 34(4)*, 436-443.
DOI : 10.1016/j.ymeth.2004.06.006

[11] B. Korber, M. LaBute & K. Yusim. (2006). Immunoinformatics comes of age. *PLoS Comput Biol, 2(6)*, e71.
DOI : 10.1371/journal.pcbi.0020071

[12] K. Yu, N. Petrovsky, C. Schonbach, J. Y. Koh & V. Brusic. (2002). Methods for prediction of peptide binding to MHC molecules: a comparative study, *Molecular medicine, 8(3)*, 137-148.
DOI : doi.org/10.1007/BF03402006

[13] B. Peters et al. (2006). A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol, 2(6)*, e65.
DOI : 10.1371/journal.pcbi.0020065

[14] B. Trost, M. Bickis & A. Kusalik. (2007). Strength in numbers: achieving greater accuracy in MHC-I binding prediction by combining the results from multiple prediction tools. *Immunome Res, 3(1)*, 5.
DOI : 10.1186/1745-7580-3-5

[15] V. Pasquetto et al. (2005). HLA-A* HLA-A*1101, and HLA-B*0702 transgenic mice recognize numerous poxvirus determinants from a wide variety of viral gene products. *J Immunol, 175(8)*, 5504-5515.
DOI : 10.4049/jimmunol.175.8.5504

[16] C. undegaard C, O. Lund, C. Kesmir, S. Brunak & M. Nielsen (2007). Modeling the adaptive immune system: predictions and simulations. *Bioinformatics, 23(24)*, 3265-3275.
DOI : 10.1093/bioinformatics/btm471

[17] N. Tiwari, N. Garbi, T. Reinheckel, G. Moldenhauer, G. J. Hammerling & F. Momburg (2007). A transporter associated with antigen-processing independent vacuolar pathway for the MHC class I-mediated presentation of endogenous transmembrane proteins. *J Immunol, 178(12)*, 7932-7942.
DOI : 10.4049/jimmunol.178.12.7932

[18] O. Demirel et al. (2007). Identification of a lysosomal peptide transport system induced during dendritic cell development. *J Biol Chem, 282(52)*, 37836-37843.
DOI : 10.1074/jbc.M708139200

[19] T. Kurotaki et al. (2007). Efficient cross-presentation by heat shock protein 90-peptide complex-loaded dendritic cells via an endosomal pathway. *J Immunol, 179(3)*, 1803-1813.
DOI : 10.4049/jimmunol.179.3.1803

[20] J. A. Swets. (1998). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285-1293.
DOI : 10.1126/science.3287615

[21] Y. Louzoun, T. Vider & M. Weigert. (2006). T-cell epitope repertoire as predicted from human and viral genomes. *Mol Immunol, 43(6)*, 559-569.
DOI : 10.1016/j.molimm.2005.04.017

강 윤 중(Yoon-Joong Kang) 정회원]

· 1995년 2월 : 경북대학교 유전공학과(이학사)
· 1997년 2월 : 경북대학교 유전공학과(이학석사)
· 2003년 2월 : 경북대학교 유전공학과(이학박사)
· 2009년 9월 ~ 현재 : 중원대학교 의생명과학과 교수
· 관심분야 : 면역학, 분자세포생물학
· E-Mail : yjkang@jwu.ac.kr