國立臺灣大學電機資訊學院資訊工程學系

博士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Doctoral Dissertation

抗體免疫表型與細胞免疫聚型的結構免疫資訊應用

Structural Immunoinformatics Exploits on AMI-Epitopes and CMI-Agretopes

范振杰

Chen-Chieh Fan, M.D.

指導教授：高成炎博士

Advisor: Cheng-Yan Kao, Ph.D.

共同指導教授：張春梵博士

Co-Advisor: Chun-Fan Chang, Ph.D.

中華民國 101 年 7 月

July 2012

國立臺灣大學博士學位論文
口試委員會審定書

細胞免疫聚型與抗體免疫表型的結構免疫資訊應用

Structural Immunoinformatics Exploits on CMI-Agretopes and AMI-Epitopes

本論文係范振杰君（學號 F91922027）在國立臺灣大學資訊工程學系完成之博士學位論文，於民國 101 年 7 月 11 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

（指導教授）

系 主 任

誌謝

終於畢業了，回顧一路走來，心中感觸良多。

當初因為興趣，一頭栽進了生物資訊的領域，在 高成炎教授的鼓勵下，由碩士班而直升博士班，期間因為專科醫師考試，職場生涯的轉換，乃至於自行創業，兩度因分身乏術而休學，然而一路跌跌撞撞，竟然還是畢業了。首先最要感謝的當然是指導教授 高成炎博士及共同指導教授 張春梵博士，感謝老師們的包容及在論文指導上所花費的精力和時間，在我想要放棄時給我支持及鼓勵。還有要感謝 朱學亭老師不厭其煩的協助及指導。也感謝 李盛安博士的大力幫忙，經常提供我許多寶貴的建議，即使畢業後還是常常給我意見及協助。感謝 詹鎮熊博士的指導，讓我對這個領域有更多的認識。還要感謝台大醫院的婁培人教授，提供了我研究上重要的啟發。實驗室的重瑞學弟及所有成員當然也不能忘記，你們的幫忙我銘感在心。

特別要感謝的是我的太太紹儀，她的支持及體諒是我前進的最大動力，感謝她及我的小孩懿文、綱佑，甜蜜的家庭是我最大的驕傲。

最後，謹以此文獻給我在天之靈的雙親。

摘要

鼻咽癌(Nasopharyngeal Carcinoma, NPC)屬於鼻咽部表皮發生的癌症，常見於台灣與中國大陸東南沿海。NPC 治療目前主流療程皆以放射線治療暨合併化學治療，然而局部復發與遠端轉移仍有可能發生，並且治療的副作用常會影響患者的生活品質。於此，醫界對於傳統 NPC 主流療程仍然積極嘗試開發相關的輔助替代療程，其中包括免疫療法。

病毒、遺傳、與環境相關三項因子皆屬高度關連於鼻咽癌 NPC 發生。病毒方面，鼻咽癌 NPC 與 EB 病毒(Epstein Barr virus, EBV)相互密切關聯，多數 NPC 細胞皆可測得 EBV 存在。EBV 攻擊 B 淋巴球與抑制細胞毒殺 T 細胞(cytotoxic T lymphocyte, CTL)；相異於典型病毒感染，EBV 造成潛伏感染時期宿主細胞表面僅只呈現少數抗原(包括 EBNA1 與 LMP1, LMP2)，其中功用相較重要關鍵的潛伏感染時期胞膜蛋白甲(latent infection membrane protein 1, LMP1)，助使潛伏感染細胞同時具備持續增生能力暨突變遁逃細胞免疫(cell-mediated immunity, CMI)的降低 CTL 細胞毒殺辨識，致使長期潛伏感染細胞持續增生癌化。遺傳方面，台灣族群相較歐美族群流行病學研究調查呈現較高的 NPC 發生機率，似與二個族群分別具有的人類白血球組織抗原(human leukocyte antigen, HLA)第一群亞型 A*02:07 與 A*02:01 呈現高度關連的統計要件，台灣族群 NPC 檢體的腫瘤細胞檢測分析發現，相較常見 LMP 存在些微差異的突變型 N-LMP1 似為免疫遁逃結果。

本論文結構免疫資訊學研究目標設定發展弱化 EBV-LMP1 細胞免疫遁逃相關 NPC 免疫療法，實施對策強化 EBV 潛伏感染細胞相關 LMP1 細胞免疫抗原的呈現效率，期能提升細胞免疫系統相關 LMP1 辨識毒殺清除效率。結構免疫資訊演算實作架構建立已知蛋白質結構片段資料庫，利用細胞免疫 CMI 相關細胞免疫抗原的歐米茄型胜肽表型與聚型(epitope & agretope)，並利用遺傳演算法預測抗原胺基酸序列之立體結構，據此進行蛋白質-蛋白質對接(protein-protein docking, PPD)演算，避免使用 LMP1 或 LMP2 完整蛋白質的致癌風險，進行蛋白質-蛋白質對接(protein-protein docking, PPD)演算，預測 MHC 與抗原胺基酸片段之結合強度(即細胞聚型之結合強度)，演算細胞免疫抗原與 A*02:07 功用複合體強化(Action complex enhancement, Ace)，相關佳化胜肽片段與篩選核可藥物，分別採取核酸疫苗轉染與直接藥物投用。

強化 EBV-LMP1 細胞免疫抗原的呈現效率方面，本研究篩選出 LMP1 中胜肽表型表現較佳之片段，根據其聚型表現，尋求其胺基酸序列之最佳化組合。弱化 EBV-LMP1 細胞免疫遁逃方面，本研究評估目前核可之藥物，篩選出在其存在的情況下，MHC 與抗原胺基酸片段之結合強度會因而增強者(即細胞聚型之結合

強度)。經結構免疫資訊演算後得知相關最佳化胜肽片段與篩選核可藥物，可分別採取核酸疫苗轉殖與直接藥物投用之方式以加強免疫療法的效果。

本論文結構免疫資訊學研究發展 EBV LMP1 免疫逋逃相關 NPC 免疫療法二項施行任務，涵蓋活體外(*ex vivo*)活化 CTL 與活體內(*in vivo*)強化 CMI：前者分離個人免疫細胞的巨噬細胞或 B 細胞與 CTL 前驅細胞並以半透網隔分區細胞培養，併用核酸疫苗呈現與直接藥物使用確效活體外活化 CTL；後者個案臨床療程僅採直接藥物使用強化抗原呈現效率暨植回活體外活化 CTL 避免核酸疫苗毒性，或許合併療程期間自體內活化 CTL 達成活體內強化 CMI；預期藉由二項操作提升臨床療程 NPC-CMI 功用，繼而有效毒殺清理 EBV-LMP1 與 A*02:07 免疫逋逃暨持續增生的潛伏感染細胞，達成減低癌化風險展現免疫療法效用。

關鍵詞：鼻咽癌、EB 病毒、細胞媒介免疫、潛伏感染時期胞膜蛋白甲、胜肽聚型、功用/聚型複合體強化

ABSTRACT

Nasopharyngeal carcinoma (NPC) is a squamous cell carcinoma that occurs on the epithelium of nasopharynx. It is a common malignancy in south-east Asia countries including Taiwan, Indonesia, Singapore, Malaysia, and Vietnam in addition to Hong Kong and southern China. Environmental factors, Ebstein-Barr virus (EBV), and genetic susceptibility are thought to play important roles towards the development of NPC. The radiotherapy or concurrent chemoradiotherapy of NPC clinical treatment may still occur local pathologic failure and distant metastasis in many patients despite of some outcome improvements. Moreover, the radiotherapy with chemotherapy often accompanies with acute side effects and long-term sequelae including secondary malignancy. Pursue for novel approaches aiming at improving outcome and reducing demand for conventional cytotoxic therapy seems thus to indicate immunotherapy as of an attractive option under development. The crucial advantage of antigen-specific immunotherapy is the ability to evaluate and monitor immune responses against targeted antigens and to correlate the findings with clinical responses.

NPC shows strong association with EBV infection that attacks B-lymphocytes as primary target towards resulted lifelong latent infection while and as well reveals an observed inhibition on specified cytotoxic T lymphocyte (CTL) populations with EBV antigenecity specificity. Notably, NPC latent infection case expresses only limited EBV viral antigens with less immunogenicity including EBV-encoded nuclear antigen (EBNA1) and latent membrane protein 1 and 2 (LMP1 and 2) which is greatly unlike that regular EBV latent infection case with expression of many EBV viral antigens in symptomatic EBV-related diseases. Both LMP1 and LMP2 may serve potentially as better vaccine targets due to the poor processing efficiency over with EBNA antigen while in antigen-presenting cells (APC) as of the infected B lymphocytes. However, LMP1 and LMP2 are with main shortages both in risky oncogenicity and as well in weak immunogenicity by stringent class I major histocompatibility complex (MHC-I) presentation in the host cell of infected B lymphocytes in order for cytotoxic T lymphocyte (CTL) activations in which as a result shifts the balance towards flexible class II-MHC (MHC-II) presentation of infected B lymphocytes in order for T helper (Th) lymphocyte activations with subtle feedback network to enhance B lymphocyte proliferations towards aberrant tumorigenesis.

Immunotherapeutic vaccination strategy with immunogenic vaccine polypeptides of assembled multiple epitope set is thus preferred whereas the oncogenic full-length

LMP1 and/or LMP2 are therefore not recommended. Promising progress in tumor growth controlling has been exemplified in animal model studies with polyepitope vaccines comprising MHC-I equivalent class I Human Leukocyte Antigen (HLA-I) restricted CTL epitope peptides from LMP1 and LMP2 despite of being with notable restriction in a relatively narrow spectrum of HLA-I alleles as of genetic susceptibility. Thus, prevalent HLA-I alleles in NPC endemic regions as of HLA-A11, A24, B27, and B57 should also be included in designing LMP-based vaccine polyepitopes along with most common HLA-I alleles such as HLA-A2.

Likely, the restricted HLA spectrum of genetic susceptibility may indicate that the overlooked anchoring agretopes of omega-shape vaccine peptide seems to be required for crucial docking onto the MHC-I pocket sub-zones towards adequate antigen presentation of peptide epitope on demonstrating immunogenicity. The design strategy of MHC-I vaccine peptide thus seemingly demands both optimized agretopes and immunogenic epitope to which additional peptide segments for improved APC proteasome processing are attached at both flanking sides. The intended vaccine peptide of epitope and agretope may be delivered in the format of "*in silico* DNA vaccine" which is constructed with expression DNA sequence deduced from the intended vaccine peptide sequence and as well with upstream control sequence of LMP1/2 promoter sequence. The developed "*in silico* DNA vaccine" with intended specific expression in EBV latent infection lymphocytes may be verified with NPC cell line of EBV-latent infected B lymphocytes for immunogenic induction in order to demonstrate the potential ability in shifting cell-mediated immunity (CMI) pathway towards MHC-I CTL while away from MHC-II Th cell.

In this thesis, we verified structure-based immunoinformatic algorithms of implemented in-house bmPDA tool in chapter 1 towards important application aspects of vasopressin bio-mimicry peptide design of known structure, MHC-I binding epitope peptide prediction of unknown structure, and EBV LMP1 related cancer vaccine peptide design of combined structure with adequate agretope and epitope for MHC-I presentation in designed delivery format likely as of DNA vaccine. The implemented algorithm comprises three sections including constructed peptide building blocks database, assembled peptide backbone model of building block candidates, and predicted peptide surface model of functional peptides.

Basically, with the concept of tri-peptide fragment assembly in chapter 2, we implemented an in-house tool of bio-mimicry peptide design algorithm (bmPDA-tool) for modeling given peptide structure. With the extracted penta peptides (penta-pep, PDB-5mer) from all entries of current protein data bank (PDB) in order for serving as basic bmPDA building blocks, the segmental backbone angles of the 3rd alpha carbon (defined as aC[3]) towards neighboring aC[2] and aC[4] as of the middle aC[2~4] in

each aC[1~5] building blocks are analyzed and constructed into searchable tri-peptides structure string (TPSS-3mer) database which is based on the "structure alphabet" with putative 22 clusters according to the parameter values including defined theta angle and edge distance, rotation axis, and rotation angle in order for the k-mean clustering analysis with bootstrapping 10,000 data entries of tri-peptide structures. With structure string alphabet of TPSS database, the mining task for similar backbone structure of 9-mer vasopressin peptide simply takes less than 1 minute for searching exact matches in entire TPSS database transformed and indexed from entire PDB.

First, to model bio-mimicry peptide structures similar to reference peptide with known backbone structure in chapter 3, the matched aC[2~4] according to serial reference penta peptide structures are mined from in-house penta-pep TPSS database with bmPDA tool in order for assembling peptide structure contig. Specifically, two mined aC[2~4] building blocks exemplified with KAV and VYN are assembled towards KAVYN contig based on superimposing [N\aC/C] co-plane of both [aa] tail-with-head amino acids between two mined aC[2~4] blocks in which the spatial rotation of mined blocks is accomplished by Quaternion-based approach along with the simple spatial shift to avoid potential structural hindrance. All fused peptide conformations in respective block combinations of bio-mimicry structures are evaluated based on minimal free energy (maximal stability) of each conformation or based on maximal structure similarity to reference structure in order for ranking optimal structures by Genetic Algorithm (GA) search strategy and/or third party program such as ProCheck for instability and Ramachandran plot analysis. The yield candidate peptide structures are converted to TPSS data in which vasopressin 9-mer peptide with known backbone structure may normally yield about 400 TPSS data entries. Second, the selected peptide model with surface structure is converted to quantitative structure–activity relationship (QSAR) model which is constructed with TPSS data and quantitative descriptors including peptide surface properties of amino acids such as exposed surface, accessibility, flexibility, hydrophilicity, charge, and so forth towards binary clustering based on structure similarity and/or binding affinity with support vector machine (SVM) according to the surface structure of reference peptide.

On the first algorithm validation in chapter 4 exemplified with known backbone structure of reference peptide vasopressin 9-mer [1YF4] CYF QNC PRG, our bmPDA tool mines bio-mimicry aC[2~4] building blocks from constructed TPSS database with qualified [theta/Ad] values in order for assembling candidate peptide combinations with highly mimicking reference structure. The bmPDA-designed bio-mimicry peptide backbone structures with different amino acid sequences from

vasopressin are exemplified with annotated solution numbers (SN) of KGN SVL AIP (SN.12), DGN SVL AIP (SN.36), and DGN SVL ADS (SN.37) taken from pooled combinations of candidate peptides in which further requires massive computational optimization with GA search strategy. The yield candidate peptide structures are coded as TPSS data in which vasopressin 9-mer peptide with known backbone structure may normally yield about 400 TPSS data entries in addition to the larger epitope peptide TMB-355 with about 3,000 TPSS data entries. Thus, the structure similarity evaluation on respective assembled structure combinations according to reference structures of vasopressin backbone and surface is accomplished by evaluating parameters with GA search strategy in physiochemical property, energy stability, and docking fitness based on accounted reference peptide structures of vasopressin backbone and surface.

On the second algorithm validation in chapter 5 exemplified with unknown epitope structure of EBV LMP1/LMP2 peptide sequence, our bmPDA tool mines bio-mimicry aC[2~4] building blocks from constructed TPSS database with qualified [theta/Ad] values in order for predicting epitope peptide structure for which the pre-processing filtering applies GA search strategy and/or ProCheck analysis in order to preliminarily predict and select stable peptide structures from assembled massive candidate block combinations and subsequently to be used for assembling runs until completion. Again, the predicted candidate epitope backbone and surface structures of assembled peptides are coded as TPSS data in order for full-size immunogenic epitope structure evaluation by GA search strategy with grouped parameters including physiochemical property, energy stability, docking fitness, and so forth. In that, our predicted peptide structures of EBV LMP1/2 contain epitope structure regions which demonstrate high consistency with epitope antigenecity index measured with NetCTL server, Kolaskar and Tongaonkar antigenecity scale, and Bepipred program. Moreover, the peptide design application for NPC cancer vaccine of likely omega shape MHC-I vaccine peptide from EBV LMP1/LMP2 demands both immunogenic epitope of previous session and as well optimal anchoring agretopes onto which respective peptide segments for improving proteasomal processing in antigen presentation cell (APC) are attached at either flanking sides towards integrated exogenous peptide expression in DNA construct.

Along with prediction methods for LMP1/2 epitope structures in previous session, the additional interactions between the potential docking sub-zones in HLA-I antigen presentation pocket and the anchoring agretopes of predicted candidate vaccine peptides are evaluated with converted QSAR models for accurate docking analysis by Molegro Virtual Docker towards mining qualified HLA-specific agretopes. The binding affinity between HLA docking sub-zones and peptide anchoring agretopes is

evaluated with SVM based on the correlations among the docking scores and the quantitative descriptors of amino acid properties. With the reference data set of the used epitopes of NPC vaccines in previous studies, the comparison on the predicted epitope and agretopes with our bmPDA tool of structural immunoinformatic approaches reveals high consistency between the candidate agretope segments and the predicted candidate epitope segments of EBV LMP1/LMP2. Moreover, the highly potential epitope segment without effective agretope segments maybe replaced with proper agretope segments in order to become highly immunogenic epitopes with improved antigenecity index when compared to the original peptide structure as of poor immunogenic epitopes.

In chapter 6, we collected approved drugs from Drugbank. Virtual screening was done by docking with MHC receptor. Drugs with better binding affinity with MHC receptor were collected as possible candidate for adjuvant immunotherapy. Epitopes with better performance of antigenecity were collected by the same procedure in chapter 5. Epitope structure prediction was done by modeling method in chapter 2. MHC receptor and candidate drugs were docked with candidate epitopes. Drugs which could enhance the binding affinity between epitope and MHC receptor were identified. We suggest drugs with ACE (action complex enhancement) to be adjuvant immunotherapy for NPC.

In conclusion, the in-house designed HLA-I cancer vaccine peptide of epitope and agretope flanking with proteasomal processing peptide can be delivered adequately in the likely practical format from "*in silico* DNA vaccine" which is constructed in chapter 6 with expression DNA sequence deduced from the designed vaccine peptide sequence and as well with upstream control sequence of active LMP1/2 promoter sequence. The developed "DNA vaccine of MHC-I cancer peptide *in silico*" with intended specific expression in EBV latent infection lymphocytes can be verified with NPC cell line of EBV latent infection for immunogenic induction which may demonstrate the potential CMI pathway shifting towards MHC-I Tc of CTL while away from MHC-II Th cell.

Keywords: Nasopharyngeal carcinoma, Epstein-Barr virus, cell-mediated immunity, class I human leukocyte antigen, latent membrane protein 1, agretope, action/agretope complex enhancer.

LIST OF ABBREVIATIONS

Abbrev.	Full Term	Page
ACE	Agretope complex enhancement	7
AMA	Anchor modified agretope	5
AMI	Antibody Mediated Immunity	4
APC	antigen-presenting cells	3
CMI	Cellular Mediated Immunity	2
CTL	cytotoxic T lymphocytes	2
EBV	Epstein-Barr virus	1
HLA	Human Lymphocyte Antigen	3
LMP1	Latent Membrane Protein 1	2
LMP2	Latent Membrane Protein 2	2
MHC	Major Histocompatibility Complex	2
NPC	Nasopharyngeal Carcinoma	1
PTLD	post-transplant lymphoproliferative disorders	2
Th	T helper lymphocyte	3
SVM	Support Vector Machine	4
SVR	Support Vector Regression	16

Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Structural Immunoinformatics on Modeling Epitope Variability and Agretope Stability

Immune system is the defense mechanism of our body against infectious agents and other foreign organisms in rather complicated process. Immunoinformatics is the computational method focusing on immune-related interactions in consists of immune-related databases, epitope prediction, vaccine design, and so forth.

In the past, vaccine development depends on biochemical and immunological experiments, such as attenuation of the wild type pathogens by random mutations and serial passages, X-ray crystallography studies of antibody/antigen structure, phage display library, overlapping peptides, NMR, radioimmunoassay, immunofluorescence, ELISA, Western blotting, and immunohistochemistry, which is very expensive, time-consuming, with low immunogenicity and reversible.[1] In recent, high-throughput experiment and computational advance progresses the understanding on immune system greatly. Assisted with epitope prediction approach, we can reduce the spectrum of dry lab target proteins and reduce the cost of wet lab experiments.

1.1 Background Review

1.1.1 Nasopharyngeal carcinoma (NPC)

Nasopharyngeal carcinoma (NPC) is a squamous cell carcinoma that occurs on the epithelium of nasopharynx.[2] It is a common malignancy in south-east Asia countries including Taiwan, Indonesia, Singapore, Malaysia, and Vietnam in addition to Hong Kong and southern China.[3] Environmental factors, Epstein-Barr virus (EBV), and genetic susceptibility are thought to play important roles towards the development of NPC.

The radiotherapy or concurrent chemoradiotherapy of NPC clinical treatment may still occur local pathologic failure and distant metastasis in many patients despite of some outcome improvements. Moreover, the radiotherapy with chemotherapy often accompanies with acute side effects and long-term sequelae including secondary malignancy.[3] Pursue for novel approaches aiming at improving outcome and reducing demand for conventional cytotoxic therapy seems thus to indicate immunotherapy as of an attractive option under development. The crucial advantage of antigen-specific immunotherapy is the ability to evaluate and monitor immune responses against targeted antigens and to correlate the findings with clinical responses.

1.1.2 Epstein-Barr virus (EBV)

Epstein-Barr virus (EBV) is a member of the herpesvirus family.[4] It has a double-stranded DNA genome of 184-kb pairs in length, encoding nearly 100 proteins.[5] It was the first virus to be associated to human cancer. EBV attack B-lymphocyte as primary target, resulting in lifelong infection.[5] Presence of EBV genome is demonstrated virtually in most NPC cells through oncogenesis process of EBV latent infections. Regardless of geographical origin, EBV is uniformly detected in patients with undifferentiated and poorly-differentiated NPC.[6] The EBV-NPC oncogenesis process may equip both proliferation advantage and immune evasion in order to overcome efficient anti-EBV immune clearance mechanisms of antibody-mediated immunity (AMI) of antibody-dependent cell-mediated cytotoxicity (ADCC) as well as cell-mediated immunity (CMI) of cytotoxic T lymphocyte (CTL)-initiative cytotoxic apoptosis during either latent and/or regular EBV infection phases.

In spite of being a latent infection in B cells, inhibition by a population of EBV-specific CTLs was observed.[7] Both in vitro and in vivo, these CTLs have been shown to have potent antiviral activity. Growing evidence revealed that cytotoxic T lymphocytes-based immunotherapy is effective in other EBV-linked malignancies, such as post-transplant lymphoproliferative disorders (PTLD). The success of this therapy has encouraged researchers to develop similar strategies for other EBV-positive tumors, such as NPC.

1.1.3 Latent membrane protein 1 and 2 (LMP1/LMP2)

Notably, NPC latent infection case expresses only limited EBV viral antigens with less immunogenicity including EBV-encoded nuclear antigen (EBNA1) and latent membrane protein 1 and 2 (LMP1 & LMP2) which is greatly unlike that regular EBV latent infection

case with expression of many EBV viral antigens in symptomatic EBV-related diseases.[8] Both LMP1 and LMP2 may serve potentially as better vaccine targets due to the poor processing efficiency over with EBNA antigen while in antigen-presenting cells (APC) as of the infected B lymphocytes according to literatures in murine models.[9]

However, LMP1 and LMP2 are with main shortages both in strong oncogenicity and as well in weak immunogenicity by stringent class I major histocompatibility complex (MHC) presentation in the host cell of infected B lymphocytes in order for cytotoxic T lymphocyte (CTL) activations due to that as a result may shift balance towards flexible MHC-II presentation of infected B lymphocytes in order for T helper (Th) lymphocyte activations with subtle feedback network to enhance B lymphocyte proliferations towards aberrant tumorigenesis.[10] Specifically, the basic proliferation advantage is likely from encoding EBV latent infection membrane protein 1 (LMP1) with growth factor receptor-like activity and as well the critical immune evasion is likely from ethnic class I human leukocyte antigen (HLA1) difference with mutating EBV genome for poor immunogenicity responses at AMI-antigen epitopes and CMI-antigen epitopes/agretopes within LMP1/LMP2 and/or EBNA of EBV-encoded proteins.

1.1.4 Epitope variability and agretope stability

Epitope is a part of a protein antigen recognized by either a particular antibody molecule or a particular T-cell/B-cell receptor of the immune system.[12] On the other hand, agretope is histocompatibility complex (MHC) binding motif of a protein antigen.[13] Highly likely, the EBV-NPC immune evasion on ultimatum agretope mutant of CMI maybe the most crucial strategy for oncogenic negative selection against which the host immune system cannot counter-act efficiently as opposed to that epitope mutants of AMI and CMI in oncogenic cells maybe eventually removed with affinity maturations of B cell receptor (BCR) and T cell receptor (TCR) by means of hyper mutations with gene rearrangements during the long-term process of EBV-related NPC oncogenesis.

1.1.5 Immune evasion

Modulation of T-cell recognition is of crucial importance for EBV, because this herpesvirus resides intracellularly for most of its life cycle. During the latent and lytic phase of EBV infection, antigen presentation of host cell via MHC class I and class II is blocked by multiple EBV gene products. Detection of cells harboring latent and replicating EBV by CD8+ and

CD4+ T lymphocytes is thus prevented. These so-called immune-evasive maneuvers prevent the induction of programmed cell death and develop persistent infection and tumor growth eventually. The NPC immune evasion of agretope maybe exemplified with the prevalence difference on ethnic HLA1 spectrum along with EBV-LMP1 mutant assorts. The A*02:07 (common in Taiwan population) shows higher prevalence of EBV-NPC than A*02:01 (common in Caucasian population) while further with additional synergistic B*4601/B*14 and extended haplotype HLA A*3303- B*5801/2- DRB1*0301- DQB1*0201/2- DPB1*0401.

The EBV-NPC biopsies from Taiwan cases reveal variant NPC-related LMP1 (NLMP1) {GenBank: X66863} of immune evasion which shares high amino acid sequence homology prominently with prototype B95.8-LMP1 {GenBank: V01555} and CAO-LMP1 in China population. Importantly, NLMP1 over-expression in Balb/c class I major histocompatibility complex (MHC1)-context towards regressing experimental murine EBV-NPC may be resulted likely from regaining strong agretope presentation of NLMP1 in mice MHC1 context in disregard of selected immune evasion of original NLMP1 in tumor microenvironment of human HLA1 context that was with weak agretope binding in CMI antigen presentation and with immune suppression in local immune suppressive cells. The ethnic A*02:07 difference of genetic susceptibility to EBV-NPC may indicate that omega-shape NLMP1np in CMI antigen presentation is required for crucial docking onto A*02:07 pit while with overlooked under-side agretope of head-anchor and tail-anchor along with overstressed bulge-side epitope in order for inducing adequate immunogenicity presentation towards effective CMI-CTL induction.

1.1.6 Cancer immunotherapy

Despite recent treatment advances that have improved the quality of life of patients with nasopharyngeal cancer, local regional failure and distant metastasis still occur in many patients. Innovative therapies are therefore still under developed. Immunotherapy is an attractive therapeutic option. There are several advantages to make use of the immune system to fight cancer. First, the immune system has the natural ability to specifically identify and kill neoplastic cells while sparing normal tissue. Second, the immune system demonstrates potential to evolve with the cancer cells. Both humoral and cellular immune system involve with cells with a vast array of clonally distributed antigen receptors. The diversity of these receptors enables the immune system to recognize foreign and/or altered antigens and to discriminate self, or normal cells, from non-self, or cancerous cells.[11]

The immunotherapeutic regime against EBV-NPC for instance may conveniently exploit various aspects including AMI-ADCC with vaccine peptides, CMI-CTL with DNA vaccines,

and microenvironment immune suppression with in vitro cell activation towards in vivo adoptive cell transfer. The current challenges except vaccine peptides still install obstacles including weak HLA1-binding agretope in host cell or dendritic cell, and specific delivery of DNA vaccine to host cells without damaging innocent bystander cells. Immunoinformatics is with remarkably high practical potential in feasible application of epitope/agretope binders onto AMI-BCR and CMI-HLA/TCR towards mining putative anchor modified agretope (Ama) and agretope complex enhancer (Ace) with reinforced binding affinity (BAff) of NLMP1 agretope and A*02:07 pit in order to likely improve NPC-CMI specifically while with low adverse cytotoxic effect due to non-specificity.

In this study, we implement bio-mimicry peptide design algorithm (bmPDA) comprising peptide database construction of building blocks, peptide backbone modeling of building block candidates, and quality evaluation on predicted nona-peptide structures. Our bmPDA of structure-based immunoinformatic approach aims at designing EBV immunogenicity-related omega-shape NLMP1 nona-peptide (NLMP1np) structures. We apply in-house bmPDA-tool towards applications of predicting A*02:07-binding EBV-NLMP1np structures in order that the verification on putative epitope and agretope quality may be accomplished with outsourcing tools of NetCTL server and Molegro Virtual Docker (MVD) software. The BAff with designed omega-shape NLMP1np and LMP1np structures on docking both HLA pits of A*02:07 {PDB: 3OXS} and A*02:01 {PDB: 1BD2} may be evaluated with MDV tool towards mining putative Ama and Ace candidates among which may be identified in modified-anchor assorts and FDA-approval drugs based on stable BAff of NLMP1np agretope and A*02:07 pit in order to specifically improve NPC CMI yet likely with low adverse effect due to non-specificity.

1.2 Specific Aims

The aim of this thesis was to develop immunotherapy of NPC via CMI-epitopes and AMI-agretopes by structural immunoinformatic approach. By our bmPDA algorithm and fragment database construction, we hope to predict epitope structure correctly and finding out mimic backbone of specific peptide. And depending on accurate docking software, we hope to correctly predict binding affinity of antigen to MHC class I molecule (agretope), thus better vaccine design and adjuvant drug for immunotherapy will be achieved.

1.3 Thesis Overview

This dissertation is organized as follows. In Chapter 2, we introduce an algorithm for peptide

structure prediction. There are many methods developed for protein structure prediction, and this field is still in progress nowadays. However, there are relatively few methods for short sequence peptide structure prediction. In the immunoinformatic field, epitope is primarily the target of concern, which is consisted of about 8-12 amino acids. After prediction of peptide structure from sequence was done, we need to extract the structural information from predicted structures. Here we followed the concept of structure alphabet; we identified 22 states of the structural alphabet that represent pattern profiles of the backbone fragments based on our block feature definition.

In chapter 3, we proposed a method for peptide block assembly and developed an algorithm for peptide block modeling. Based on the peptide block assembly method and peptide structure observation, features of a block of 3 amino acids were defined.
We developed a methodology to build QSAR models by using SVM. After prediction of peptide structure from sequence was done, we need to extract the structural information from predicted structures. With the concept of structure alphabet; we looked up 22 states of structural alphabet that represent backbone pattern profiles based on defined block feature definition. After the peptide model had been generated, it was converted to structural string specifically.[14] Combined with other physiochemical properties (amino acid symbol, hydropathy, polarity, side chain charge) of peptide blocks, the MHC binding affinity was predicted by SVM.

In chapter 4, we develop a method called “bio-mimicry peptide design“. Follow the concept of inverse folding search; we develop an approach to find possible sequence combinations mimicking target structure. Evaluation on the structure similarity with target peptide, physiochemical property, and structure stability of predicted solutions were done for finding better potential candidates.

In chapter 5, we applied the above structural immunoinformatic approaches for nasopharyngeal carcinoma (NPC) vaccine design. NPC is a common malignancy in southern China, Hong Kong, and south-east Asia countries including Taiwan, Singapore, Malaysia, Indonesia, and Vietnam. It is strongly associated with Epstein-Barr virus (EBV). Immunotherapy for NPC is currently focusing on the tumor-associated antigens called LMP1 and LMP2. However, poor antigenecity of LMP1/LMP2 limited the efficacy of EBV vaccine in NPC immunotherapy. We predicted the structure of every possible epitopes of LMP1/LMP2 from sequence, docked them with MHC-I molecule, and compare the docking result with predicted antigenecity of LMP1/LMP2 from several epitope prediction servers. Epitopes with better performance of antigenecity were collected as candidates for polyepitope regimen. According to the preference observation on known epitopes, residues on specific

position of the candidate epitopes were modified to become epitopes with even better antigenecity. Agretope performance was evaluated by binding affinity prediction from docking with MHC receptor. We suggest epitopes with better performance on epitope and agretope to be candidates of polyepitope regimen on NPC immunotherapy.

In chapter 6, we collected approved drugs from DrugBank. Virtual screening was done by docking with MHC receptor. Drugs with better binding affinity with MHC receptor were collected as possible candidate for adjuvant immunotherapy. Epitopes with better performance of antigenecity were collected by the same procedure in chapter 5. Epitope structure prediction was done by modeling method in chapter 2. MHC receptor and candidate drugs were docked with candidate epitopes. Drugs which could enhance the binding affinity between epitope and MHC receptor were identified. We suggest drugs with ACE (action complex enhancement) to be adjuvant immunotherapy for NPC.

Finally, in chapter 7, we presented summary and future perspectives.

Chapter 2 Peptide Structure Indexing with Tri-mer Blocks and Structure Strings

2.1 Introduction

Description of 3D information of protein structures is the first step for structural bioinformatics. Protein structure is defined by four distinct levels: primary, secondary, tertiary, and quaternary structures. There are further description systems in each specific level. In this chapter, we develop a protein backbone description system based on features of 3mers peptide block. Fragment databases are constructed based on our method. Further simplified description system called "structural alphabets" was developed. Based on structural alphabet and other physiochemical properties of peptides, a structure-based epitope prediction was also achieved.

2.1.1 Peptide structure features

Description of local protein structures is essential for structural bioinformatics. Traditionally, protein backbone structures can be described by secondary structure such as α-helix, β-strand and coil. However, description of protein structures by only three states is oversimplified. There are some other definitions developed for describe protein structure. Different strategies had been used such as dihedral angle, backbone curvature and torsion. In this study, we develop a protein backbone description based on features of 3mers peptide block.

2.1.2 Structure alphabet

By clustering certain features of protein structures in a chosen number of states, we can define a prototype which is representative of the local structures in each cluster. These libraries of local structures prototypes are called "structural alphabets".[15] They complement each other to form a 'universal code' of local conformations. There are several structural alphabets being developed by different methods such as cluster analysis, Kohonen maps and

Hidden Markov Models.[16] By encoding protein structure as 1D sequence, structural alphabets has been applied on many fields such as decoy generation, local structure prediction, structural comparison, alignments, and mining, structure reconstruction from Ca, and so on.[16]

2.2 Method

2.2.1 Defined structure features of tri-peptide building blocks

Features of a 3mers block were defined as (1) theta angle: by Cα[-1]:Cα[0]:Cα[+1], (2) arm distances between Cα[-1]:Cα[0] and Cα[0]:Cα[+1], (3) Rotation angle and rotation axis of two planes formed respectively by Cα[-1]:N[-1]:C[-1] and Cα[+1]:N[+1]:C[+1] (**Figure 2-1**).

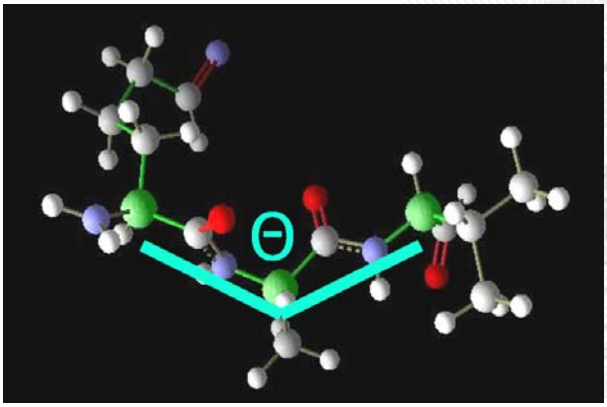


Fig. 2-1 Tri-peptide building blocks with defined features of theta angle and arm distances [tA/aD] of middle aC[2~4] grouped into 22 clusters.

2.2.2 Basic ideas of bio-mimicry peptide design algorithm (bmPDA)

For manipulate the peptide blocks, an algorithm for peptide blocks assembly was developed. The basic principle is (1) make use of the actual existing structures as building blocks , (2) use 3mers as building blocks and one more residue at both end as conformational limiting constraint (3) replace the last residue of one block with the leading residue of another block. Under the above principles, bio-mimicry peptide design algorithm (bmPDA) was developed. Our implemented in-house bmPDA tool for modeling omega-shape nona-peptide structures of class I HLA pit structure comprises two sectors: preparation of tri-peptide building blocks and optimization of predicted candidate structures.

2.2.3 Construction of tri-peptide structure part (TPSP-5mer) library

Based on the concept of assembled penta-peptide from fusing identical tail and head residues of serial tri-peptide candidates in realistic presence structures from protein data bank (PDB), we implemented bmPDA tool for preparing required tri-peptide structure part (TPSP) denoted as alpha carbon 2~4 (aC{2~4}) in order to show source templates of penta-peptide denoted as aC{1~5} which are entirely extracted from all available protein structure entries in RCSB PDB (www.pdb.org).

The aC{1~5} penta-peptide library at 40GB data size extracted from all known PDB entries until 22 JAN, 2010 comprises 45,141,909 blocks with middle portion for preparing middle segment aC{2~4} TPSP database. In addition, the applied TPSP features of tri-alpha-carbon (3aC{-1~0~+1}) backbone are respectively annotated with convenient parameters including theta-angle and arm-distances {tA/aD} and rotation index of bilateral terminal residues.

2.2.4 Construction of tri-peptide structure string (TPSS-3mer) database

Following the concept of structure alphabet, we construct the tri-peptide structure string (TPSS-3mer) database with identified 22 cluster states of the structural alphabet that represent pattern profiles of tri-peptide backbone structures based on our building block feature definition including theta angle, edge distances, and rotation angles and distances. Having done with peptide structure prediction from peptide sequence, we convert all PDB entries into structural alphabet strings in order for backbone searching or extracting blocks from known PDB structure files (**Table 2-1**).

The tri-peptides structure string (TPSS-3mer) database comprising putative 22 “structure alphabet” clusters is based on tA/aD values of respective aC{2~4} TPSP in order for comparing structure similarity among given TPSP sets. The statistical analysis is accomplished towards 22 k-mean clusters with about 223,050 tA/aD values of aC{2~4} TPSP from applied penta-peptide blocks extracted from 1,000 protein structures randomly selected from PDB. In that, the TPSS database of converted structure alphabet for entire PDB structure entries may take at less than 50 MB data size based on compact 22 tA/aD cluster values and may take 500MB data size with detailed annotation contents. The searching performance of structure similarity on given nona-peptide structure through TPSS database of entire PDB library may take less than 50 seconds with regular NoteBook PC for mining exact match TPSS candidate structures within overall penta-peptide blocks.

Reference epitope structures of omega-shape octa-peptides, nona-peptides, and deca-peptides extracted from PDB as reference templates are converted as TPSS structure sequences to become loose structure reference for inferring unknown structure epitopes towards loose structure similarity models in order for subsequent virtual docking with HLA1 pit structure. The HLA1 pit structures of both A*02:07 { 3OXS} of Taiwan population and A*02:01 {1BD2} of Caucasian population are converted as spatial regression axis to assist the virtual docking with designed nona-peptide structure upon evaluation. Immune Epitope Database (IEDB) inferred position-specific preference on HLA1-binding nona-peptide at position {2nd; 9th} amino acids of agretope head-anchor and tail-anchor may indicate respective preference with strong agretope anchor {L/M; I/L/V} rather than with tolerated {I/Q/V; A/M}.

Table 2-1 TPSS database of 22 clustering matrix assigned in English alphabet characters except [J,U,O,X] for structural string data conversion and illustrated with cluster distribution of percentage and instance number.

Cluster symbol	C alpha distance arm 1	C alpha distance arm 2	C alpha angle	Ca-N vector angle	plane rotation angle	Percentage (%)	Instance Number
A	4.4187	4.4094	1.7115	2.1612	2.2219	3%	11998
B	3.8990	3.8913	1.5860	1.6774	1.8083	15%	59657
C	4.5855	4.4426	1.6357	0.7546	2.1371	2%	8779
D	3.8299	3.8295	2.0735	0.8649	1.0812	5%	20472
E	3.8967	3.8932	2.0996	1.6037	1.9411	3%	13310
F	4.0889	4.0638	1.6465	0.4079	1.5149	2%	8836
G	3.8370	3.8518	2.1091	1.0285	1.8692	4%	15751
H	4.3305	4.2964	1.6116	1.0598	1.2222	5%	17820
I	3.8564	3.8670	1.6132	1.8624	1.7653	15%	57255
K	3.8342	3.8502	2.2842	0.4272	2.0378	2%	8341
L	3.9041	3.9356	1.9001	0.5015	0.6959	5%	17972
M	3.9737	3.9763	2.3243	2.0920	1.4676	2%	8427
N	4.4889	4.5327	1.8651	2.1660	0.8041	2%	6727
P	3.8148	3.8132	2.2800	0.3220	1.3125	4%	14350
Q	4.1179	4.1199	1.6206	1.2939	1.8264	4%	13883
R	4.0264	3.9265	2.2301	2.3018	2.5706	2%	8121
S	3.8618	3.8670	2.1328	1.3833	1.2832	4%	17400
T	3.9361	3.9526	1.9107	0.3189	2.6691	2%	6858
V	3.7942	3.7953	2.1049	0.1378	0.2468	6%	23338
W	4.1290	4.1868	1.6202	1.6354	1.2376	5%	19498
Y	4.1302	4.1443	1.6750	2.1397	1.4272	3%	13065
Z	3.8116	3.812	2.2908	0.3062	0.6892	5%	19261

2.3 Result

2.3.1 Penta-peptide (PDB-5mer) structure library

The PDB-5mer Library at 40GB data size (data not shown) comprises 45,141,909 penta peptide blocks extracted from all known PDB entries until 22 JAN, 2010. The tri-peptide of

PDB-5mer mid blocks (shown in **Figure 2-1**) are annotated with theta angle and arm distances [tA/aD] of middle aC[2~4] segment backbone in each aC[1~5] building blocks.

2.3.2 Structural alphabet

The aC[2~4] block backbones are analyzed and constructed into tri-peptides structure string (TPSS) database with assigned "structure alphabet" characters for putative 22 clusters based on aC[2~4] tA/aD values. The clustering analysis is accomplished towards 22 k-mean clusters with about 223,050 PDB-5mer blocks extracted from randomly selected 1,000 PDB sequences (**Table 2-2**).

Table 2-2 TPSS database implemented similarity distance matrix.

	A	B	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Z
A	0	0.73	0.90	1.59	1.41	0.55	1.05	5.03	0.46	1.47	0.76	1.58	1.51	1.80	0.49	2.49	1.27	0.90	1.56	4.60	0.82	1.72
B	0.73	0	0.58	1.21	1.15	0.39	1.38	4.67	0.40	1.55	1.10	1.86	1.94	2.04	0.24	2.81	1.13	1.01	1.91	4.10	1.40	1.98
C	0.90	0.58	0	0.92	0.84	0.59	1.41	4.77	0.83	1.42	1.15	1.81	1.92	2.05	0.60	2.90	0.89	0.95	2.01	4.13	1.59	1.98
D	1.59	1.21	0.93	0	0.33	1.12	1.51	4.20	1.41	1.10	1.38	1.67	2.01	1.85	1.28	2.77	0.54	1.03	2.02	3.53	2.05	1.81
E	1.41	1.15	0.84	0.33	0	1.02	1.25	4.36	1.29	0.88	1.14	1.43	1.74	1.64	1.16	2.59	0.28	0.79	1.80	3.76	1.82	1.59
F	0.55	0.39	0.59	1.12	1.02	0	1.07	4.62	0.34	1.25	0.77	1.54	1.64	1.74	0.29	2.52	0.92	0.70	1.62	4.12	1.16	1.66
G	1.05	1.38	1.41	1.51	1.25	1.07	0	4.61	1.13	0.69	0.32	0.57	0.66	0.79	1.20	1.61	0.99	0.54	0.67	4.36	0.81	0.71
H	5.03	4.67	4.77	4.20	4.36	4.62	4.61	0	4.72	4.27	4.64	4.43	4.73	4.23	4.77	4.23	4.36	4.49	4.43	1.36	4.96	4.27
I	0.46	0.40	0.83	1.41	1.29	0.34	1.13	4.72	0	1.45	0.84	1.65	1.68	1.82	0.25	2.53	1.19	0.89	1.61	4.26	1.04	1.75
K	1.47	1.55	1.42	1.10	0.88	1.25	0.69	4.27	1.45	0	0.79	0.60	1.02	0.80	1.45	1.76	0.63	0.60	1.06	3.93	1.45	0.75
L	0.76	1.10	1.15	1.38	1.14	0.77	0.32	4.64	0.84	0.79	0	0.84	0.89	1.05	0.91	1.84	0.89	0.37	0.91	4.31	0.72	0.97
M	1.58	1.86	1.81	1.67	1.43	1.54	0.57	4.43	1.65	0.60	0.84	0	0.52	0.33	1.71	1.25	1.17	0.90	0.58	4.26	1.24	0.27
N	1.51	1.94	1.92	2.00	1.74	1.64	0.66	4.73	1.68	1.03	0.89	0.52	0	0.66	1.75	1.22	1.49	1.08	0.54	4.61	0.98	0.62
P	1.80	2.04	2.05	1.85	1.64	1.74	0.79	4.23	1.82	0.80	1.05	0.33	0.66	0	1.91	0.97	1.39	1.12	0.5	4.15	1.38	0.09
Q	0.49	0.24	0.60	1.28	1.16	0.29	1.20	4.77	0.25	1.45	0.91	1.71	1.75	1.91	0	2.67	1.1	0.88	1.74	4.25	1.18	1.84
R	2.49	2.81	2.90	2.77	2.59	2.52	1.61	4.23	2.53	1.76	1.84	1.25	1.22	0.97	2.67	0	2.34	2.01	0.98	4.45	1.83	1.03
S	1.27	1.13	0.89	0.54	0.28	0.92	0.99	4.36	1.19	0.63	0.89	1.17	1.49	1.39	1.10	2.34	0	0.55	1.54	3.85	1.60	1.34
T	0.90	1.01	0.95	1.03	0.79	0.7	0.54	4.49	0.89	0.60	0.37	0.90	1.08	1.12	0.88	2.01	0.55	0	1.12	4.09	1.07	1.05
V	1.56	1.91	2.01	2.02	1.80	1.62	0.67	4.43	1.62	1.06	0.91	0.58	0.54	0.50	1.74	0.98	1.54	1.12	0	4.38	0.99	0.46
W	4.60	4.10	4.13	3.53	3.76	4.12	4.36	1.36	4.26	3.93	4.31	4.26	4.61	4.15	4.25	4.45	3.85	4.09	4.38	0	4.75	4.17
Y	0.82	1.40	1.59	2.04	1.82	1.16	0.81	4.96	1.04	1.45	0.72	1.24	0.98	1.38	1.18	1.83	1.60	1.07	0.99	4.75	0	1.31
Z	1.72	1.98	1.98	1.81	1.59	1.66	0.71	4.27	1.75	0.75	0.97	0.27	0.62	0.09	1.84	1.03	1.34	1.05	0.46	4.17	1.31	0

2.3.3 Tri-peptide structure string (TPSS-3mer) database

The TPSS database of converted structure alphabet for the entire PDB structure data entries may take at less than 50 MB data size based on compact 22 tA/aD cluster values and may take 500MB data size with detailed annotation contents. The searching performance on any given 9mer peptide structure thru the TPSS database of entire PDB library may take less than 50 seconds with regular PC for mining TPSS-3mer exact match structures of overall PDB-5mer candidate blocks.

2.3.4 Application

Based on the Penta-peptide (PDB-5mer) structure library, Tri-peptide structure string (TPSS-3mer) database, and structural alphabet, several tools were developed for manipulating protein blocks (**Table 2-3**). Combining these different tools, we can fulfill many structure-related tasks such as extract arbitrary protein blocks, search PDB for peptide blocks with similar backbone structure of specific peptide, replace some part of protein with other blocks, or even doing structure BLAST over PDB, like 3D-BLAST[14] does.

Table 2-3 Tools implemented for manipulating protein blocks in the initial parts of bmPDA.

No.	Type	Task	Name of implemented tool
1	extractor	peptide sequence	Protein block extractor (by peptide sequence)
2	extractor	structure string	Protein block extractor (by structure string)
3	searcher	peptide sequence	Search PDB by peptide sequence
4	searcher	structure string	Search PDB by structure
5	calculator	superimpose RMSD	Block superimpose and RMSD calculation
6	aligner	structure alignment	Structure alignment of proteins
7	merger	block merging	Protein block merging

2.4 Discussion

Knowledge of the 3D structure of proteins is important in clarifying their properties, behavior and almost all biological condition mediated by proteins, including protein-ligand and protein-protein interactions. It is also helpful in drug discovery and protein design.

Comparing with the rapid increasing number of reported protein sequences, protein structure determination by experiment is far behind. Thus informational technology for protein structure prediction and manipulation are necessary.[17]

Several structural alphabets system had been developed using different methods such as cluster analysis, Kohonen maps and Hidden Markov Models. After encoding protein structure into structure string, complicated 3D coordinate information is more easily manipulated for machine learning or other analysis. The potential of structural alphabets has been shown by application on decoy generation[18], local structure prediction[19-21], sequence-based structural comparison[22], combined sequence-structure alignments[23], 3D structure alignment[24], structure mining[25-29], structural reconstruction from Ca[30], fold classification[26], fold prediction[31], structure generation[32], de novo prediction[33, 34], de novo backbone design[35], and so on.[16]

Chapter 3 Peptide Structure Modeling with Backbone Quaternion-GA and Surface QSAR-SVR

3.1 Introduction

3D protein structures are critical for understanding biology at both molecular and system level. However, the speed of sequence publishing into databanks considerably exceeds that of structure determination despite the advances in experimental structural biology. Protein structure prediction has been a very challenging problem. There are two categories for protein structure prediction: Ab initio method and Knowledge based method. Knowledge based method can be further classified into comparative modeling, fold recognition and other new fold methods. Comparative modeling and fold recognition method are sometimes being called “template based modeling”.[36] In this study, we developed an algorithm for short peptide structure prediction.

3.1.1 Quaternion (Q4)

Quaternion method was introduced by Hamilton in the mid-nineteenth century as an extension of complex numbers and as a tool for manipulating 3-dimensional vectors.[37] Quaternion is a convenient tool for handling spatial rotation problems. It has compact representation of rotations, easy to maintain a quaternion’s unit normalization, and derive many important results concerning rotations in a simple coordinate-free way. Compare with other rotation methods, such as Euler axis system and rotation matrix, quaternion can avoid gimbal lock in Euler system, has less floating-point round-off errors than matrix, and normalizing a quaternion is computationally less expensive. It is widely used in three-dimensional computer graphics and computer vision.

3.1.2 Genetic algorithm (GA)

Inspired from the principles of biological evolutionary theory, genetic algorithm is a stochastic computational model to solve optimization problems. GA model the natural phenomenon of genetic inheritance based on the principle of "survival of the fittest". GA had been widely used to solve sequential decision process for function optimization, machine learning and general optimization problems.[38]

3.1.3 Epitope and agretope prediction

In the past, vaccine development depends on biochemical and immunological experiment, such as phage display library, overlapping peptides, ELISA, NMR, immunofluorescence, radioimmunoassay, Western blotting, immunohistochemistry, X-ray crystallography studies of antibody/antigen structure and attenuation of the wild type pathogens by random mutations and serial passages, which is very expensive, time-consuming, with low immunogenicity and reversible. Under the help of epitope prediction approach, we can narrow the spectrum of target proteins, and reduce the cost of wet experiments.

The most predictable part of T cell epitope generation is peptide-MHC binding. MHC-I and MHC-II genes are highly polymorphic, and the most of their variable part are located in binding pockets that restrict peptide interactions to those with particular amino acids at characteristic positions.[39]

There are four approaches being applied to predict epitopes: sequence-based methods, structure based methods, hybrid methods and consensus methods.[1] The majority of epitope prediction methods are currently data-driven sequence-based, and they are more reliable than structure-based methods. On the other hand, there are several advantages for structured-based methods. First, only a smaller dataset is necessary for training. Second, it can predict peptides for alleles that have not been extensively studied. Third, discontinuous epitopes are only possible to predict by structure-based method. Last, even sequence-based approaches depend on structure information to make reliable predictions. However, the development of structure-based approach is still greatly limited due to high computational cost, development complexity and scarcity of 3D protein structures.[1]

3.1.4 Quantitative structure–activity relationship (QSAR)

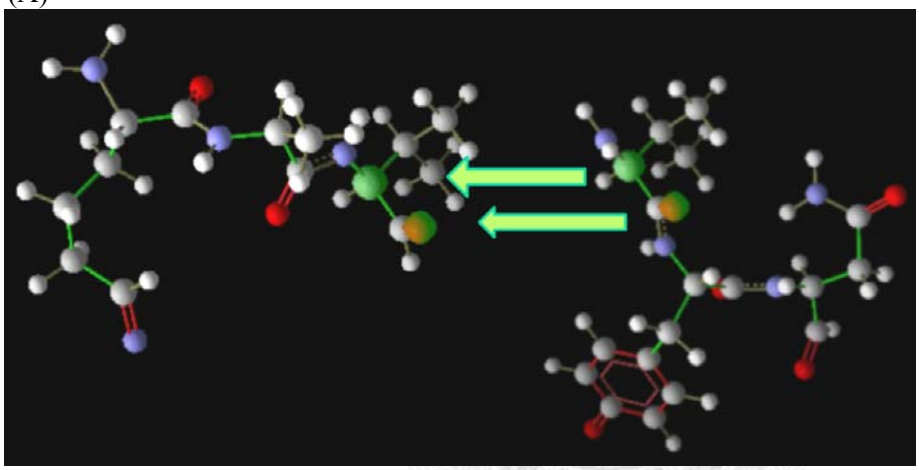
The quantitative structure–activity relationship (QSAR) model combines structure alphabet

string and physiochemical properties (amino acid symbol, hydropathy, polarity, and side chain charge) of predicted peptide structures.

3.1.5 Support vector machine (SVM) / Support vector regression (SVR)

Support vector machine (SVM) is a supervised learning method that can be applied on classification or regression. It was developed by Vapnik since 1963 based on the Structural Risk Minimization Principle. LibSVM is an group of software for support vector classification, (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM)[40].

(A)



(B)

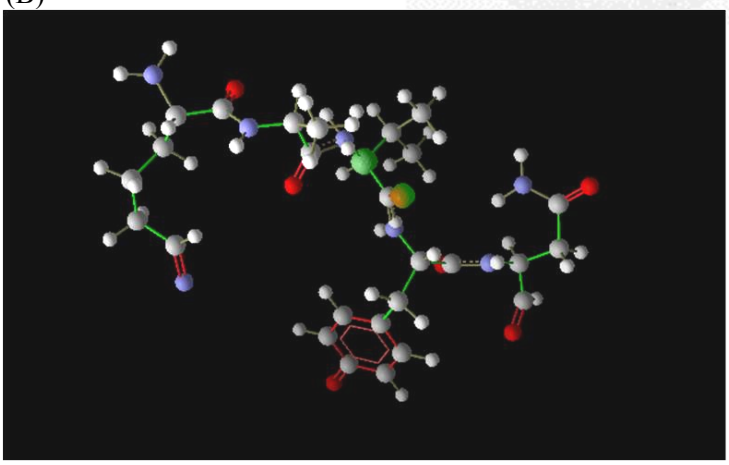


Fig. 3-1 Spatial rotation and aligned fusion with mined TPSS-3mer within similar PDB-5mer blocks. (A) rotation and shift for fusion at super-imposed tail-under-head residue thru Quaternion and GA; and (B) assembled penta-peptide extended with two tri-peptide building blocks.

3.2 Method

3.2.1 Quaternion (Q4) spatial rotation and super-imposition shift to assemble peptide from TPSP-5mer building blocks

For rotation of the blocks, we use Quaternion (Q4) calculation aiming at (1) avoiding gimbal lock in Euler system, (2) reducing floating-point round-off errors than matrix, and (3) normalizing a quaternion is computationally less expensive. For superimposing the identical amino acid residue, firstly we matched two residues by matching Cα atom, alignment was done with Cα-N vector, and then rotate with N-Cα-C plane matched (**Figure 3-1A**).

For modeling of a peptide sequence, the sequence was firstly transformed to contig sequences based on overlapping amino acid subsequences, such as the KAV and VYN towards KAVYN sequence contig. For each contig sequences, candidate building blocks were identified by searching bmPDA 5mer building blocks database (PDB-5mers database) with the middle 3mers of the penta-pep matched with the contig sequence. Modeling was done with the building blocks assembled by our bmPDA tool (**Figure 3-1B**).

3.2.2 Peptide modeling

Epitope structure prediction was done by our modeling method of bmPDA-tool (**Figure 3-2**). Specifically, we mined aC[2~4] building blocks exemplified with KAV and VYN are assembled towards KAVYN contig based on superimposing [N\aC/C] co-plane of tail-under-head identical residues between two mined aC[2~4] consecutive blocks in which spatial rotation is accomplished by Quaternion-based approach along with simple spatial shift to avoid potential structural errors. All merged peptide combinations of bio-mimicry structures are evaluated either based on free energy of each conformation or based on structure similarity to reference structure in order for ranking optimal structures by Genetic Algorithm (GA) search strategy (**Figure 3-2**). [17,18,19] In that, the structure similarity between reference structure and bmPDA-tool predicted peptide structures are further verified with SuperPose in order to estimate root mean square deviation (RMSD). [20]

3.2.3 Structural mining on reference peptide TPSS-3mer conversion

Following the concept of structural alphabet by Yang, based on our block feature definition, we identified 22 states of the structural alphabet that represent pattern profiles of the backbone fragments. After the peptide model had been generated, it was converted to structural string specifically (**Figure 3-2**). [14]

3.2.4 Conformation evaluation and GA optimization

The resulting decoy conformations are then evaluated according to free energy of each conformation or structure similarity to target structure. Optimal solution is found by Genetic algorithm (GA) search strategy. The similarity evaluation on merged structures with target structure is accomplished by checking physiochemical property, energy stability, docking fitness with counter-structures of target peptide (**Figure 3-2**).

3.2.5 Structural QSAR descriptors and data encoding

The QSAR descriptors comprise structure alphabet string and physiochemical properties (amino acid symbol, hydropathy, polarity, and side chain charge) of predicted peptide structures (**Figure 3-2**).

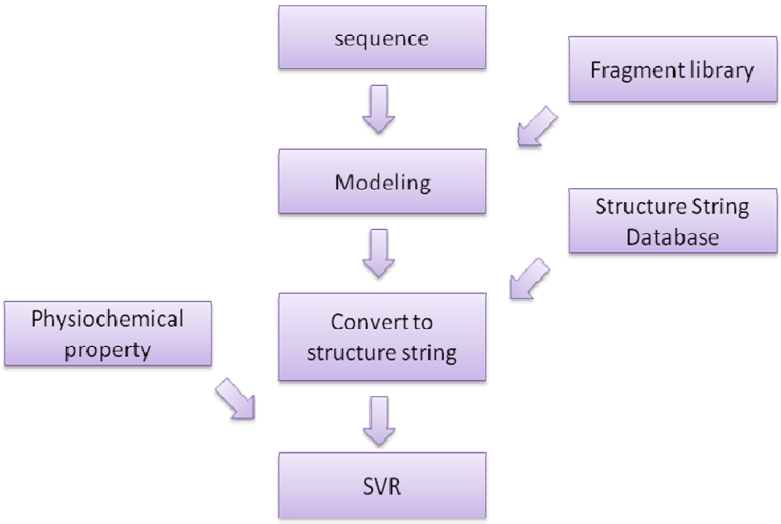


Fig. 3-2 Work flow of epitope prediction by bio-mimicry peptide design algorithm (bmPDA) based on QSAR.

3.2.6 Support vector machine / regression (SVM / SVR)

SVM was employed to predict the possibility of a sequence of 9 amino acids being an epitope of MHC-I. SVM is a supervised learning method developed by Vapnik in 1963 based on the Structural Risk Minimization Principle. [19,21,22] SVM had been applied widely in the field of computational biology and is a promising technique for data classification. SVR (support vector regression) is a version of SVM for regression (**Figure 3-2**).

The applied LibSVM software is an integrated version on classification or regression software comprises support vector classification (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation (one-class SVM). [40]

3.2.7 Modeling omega-shape conformation on HLA1-binding nona-peptide structure

In this study, our bmPDA tool aims at designing omega-shape nona-peptide structures with epitope-stemside bulge and agretope-rootside anchors from NLMP1np {X66863} and LMP1np {V01555} towards docking HLA1 pit structures of A*02:07 {3OXS} and A*02:01 {1BD2}. Reference nona-peptide epitopes as of training dataset extracted from IEDB (www.immuneepitope.org) contains 3,886 A*02:01-binding epitopes of biological binding assay results. Our bmPDA tool models omega-shape nona-peptide epitope structure for docking HLA1 pit structure with both tri-peptide finding kernel (TFK) and geometric hashing kernel (GHK) respectively for mining and fusing TPSP and for filtering and fitting designed candidate structures.

3.2.8 Mining and fusing tri-peptide structure parts into designed combinatorial nona-peptide structures

To model nona-peptide epitope structures of NLMP1 and LMP1 with similar omega-shape backbone to reference epitope structure in TPSS format, the TFK module of bmPDA tool recursively mines the TPSP database for potential tri-peptide aC{2~4} building blocks

according the penta-peptide aC{1~5} sliding window of amino acid sequence within NLMP1 and LMP1 genomes of EBV strain B95-8 in order for fusing into omega-shape nona-peptide structure. Exemplified with NLMP1np- 032/035: LLL-ALL FWL YIV structure with number 032/035 indicating initial amino acid position, the bmPDA tool respectively mines all the aC{1~5} penta-peptide structure candidates of 034: LAL LF and 036: LLF WL according to structure similarity of reference epitope structures TPSS in order for retrieving and fusing all the aC{2~4} TPSP structure blocks of 035: AL(L) and 037: (L)FW into potential aC{1~5} structure combinations of 035: AL(L)FW. The indicated bmPDA design procedure of epitope structure on HLA1-binding nona-peptide may recursively process until completion of structure modeling covering all serial nona-peptide segments within NLMP1 and LMP1.

For fusing the aC{2~4} TPSP blocks exemplified with 035: AL(L) candidates and 037: (L)FW candidates into potential aC{1~5} combinatorial structure candidates of NLMP1 035: AL(L)FW, our bmPDA tool applies quaternions (Q4) approach for spatial rotation and shift towards tail(L)-under-head(L) superimposition 035: AL(L)FW based on co-planar overlapping of {N\aC/C} amino acid plane with Q4-based spatial processing to avoid potential deadly error with Euler angles such as gimbal lock while loss of a degree of freedom during spatial rotation. Fused nona-peptide structures with every TPSP blocks in numerous combinatorial structure candidates onto converted TPSS format are evaluated either based on free energy of each designed structure or based on structure similarity to reference structure of omega-shape epitope templates in TPSS format in order for ranking optimal structures by genetic algorithm (GA) search strategy. Inspired with biological evolutionary principles, GA is a computational stochastic model to solve optimization problems.

Specifically, our Q4-GA module of bmPDA tool evaluates structure similarity on respective combinations of merged nona-peptide structure based on collective IEDB omega-shape epitope templates in TPSS structure clusters which is accomplished by verifying parameters with GA search strategy upon physiochemical property, energy stability, and docking fitness according to backbone and surface structure of reference epitope templates. In that, the structure similarity between reference structure and predicted nona-peptide structures are further verified with Q4-GA based on SuperPose server in order to estimate root mean square deviation (RMSD). Thus, Q4-GA module of bmPDA tool complies with natural phenomenon of genetic inheritance based on “survival of fittest” principle which is widely applied in solving sequential decision process for function optimization, machine learning and general optimization problems.

3.2.9 Filtering and fitting designed candidate structures into putative

omega-shape nona-peptide structures

To evaluate nona-peptide epitope structures of NLMP1 and LMP1 with optimal omega-shape epitope structure with reference template in detailed TPSS format, the GHK module of bmPDA tool filters and fits selected cases of predicted nona-peptide structures in detailed TPSS format of quantitative structure–activity relationship (QSAR) model for support vector regression (SVR). The evaluation in all the predicted HLA1-binding nona-peptide structures of NLMP1 and LMP1 to A*02:07 and A*02:01 pit structures depends upon optimal omega-shape conformation with two agretope-rootside anchors at either end and as well one epitope-stemside bulge at middle segment.

Based on the concept of "structure alphabet" in detailed TPSS conversion format, the QSAR-SVR module of bmPDA tool encodes QSAR descriptors of peptide surface properties of amino acids such as exposed surface, accessibility, flexibility, hydrophilicity, charge, and so forth towards binary clusters on structure similarity and/or binding affinity along with SVR evaluation according to the reference nona-peptide surface structure. The applied LibSVM software set as of a supervised learning method on classification or regression comprises support vector classification (C-SVC, nu-SVC), support vector regression (epsilon-SVR, nu-SVR) and distribution estimation of support vector machine (one-class SVM). Basically, Vapnik developed SVM in 1963 while based on the structural risk minimization principle.

Table 3-1 Structural RMSD difference between bmPDA predicted structures and reference structures of HLA 0201 epitopes based on TPSS similarity distance matrix..

Epitope Sequence	Complex PDB ID	Chain	RMSD(superpose)
ALWGFFPVL	1LP9	C	0.27
ALWGFFPVL	1B0G	C	0.35
FAPGFFPYL	1I7R	C	0.05
GILGFVFTL	1B0R	C	0.87
GILGFVFTL	2VLR	C	0.39
GILGFVFTL	2VLK	C	0.19
GILGFVFTL	1HHI	C	0.38
GLMWLSYFV	3I6G	C	2.52
IISAVVGIL	1QR1	C	0.76
ILKEPVHGV	1P7Q	C	1.06
ILKEPVHGV	1HHJ	C	1.00
ILSALVGIL	1EEZ	C	1.55
ILSALVGIV	1EEY	C	2.73
IMDQVPFSV	1TVH	C	0.15
ITDQVPFSV	1TVB	C	0.15
LLFGKPVYV	2GIT	C	0.73
LLFGKPVYV	2GJ6	C	1.06
LLFGYAVYV	1QRN	C	0.13
LLFGYPRYV	1QSE	C	0.16
LLFGYPVAV	1QSF	C	0.18
LLFGYPVYV	1AO7	C	0.15
NLVPMVAAV	3GSW	P	2.25

NLVPMVATV	3GSO	P	0.30
NLVPMVAVV	3GSX	P	0.83
NLVPQVATV	3GSV	P	0.54
NLVPSVATV	3GSQ	P	0.96
NLVPTVATV	3GSU	P	0.67
NLVPVVATV	3GSR	P	0.89
RQASLSISV	3BGM	C	2.20
SLLMWITQA	1S9X	C	0.92
SLLMWITQC	2F53	C	0.58
SLLMWITQS	1S9Y	C	0.90
TLTSCNTSV	1HHG	C	1.12
VLHDDLLEA	3FT3	P	0.70
Average			**0.814412**

3.3 Result and Dataset

3.3.1 Epitope structure prediction by bmPDA

Epitopes with known structure of HLA A*0201 were collected from IEDB (Immune Epitope Database) and PDB databases. Peptide structures predicted by bmPDA were compared with the actual structures from PDB. Alignment was done by Superpose and RMSD score was calculated. **Table 3-1** shows the comparison results from SuperPose.[41] An alignment example was illustrated in **Figure 3-3**.

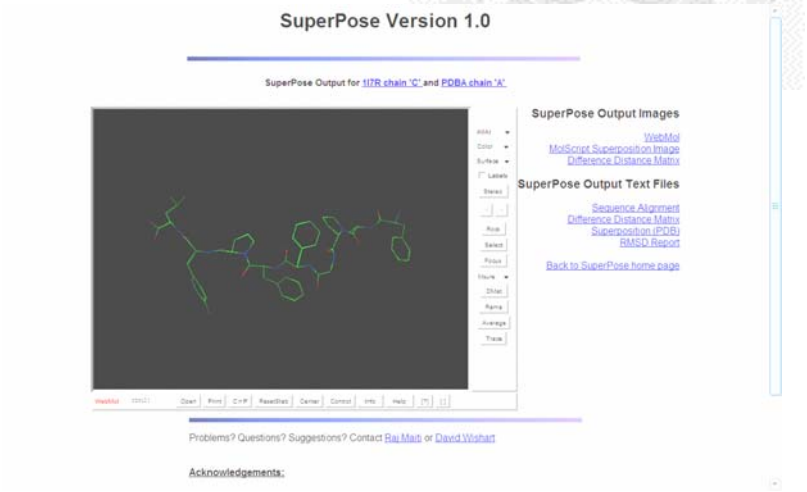


Figure 3-3 Illustrated comparison of SuperPose alignment between bmPDA predicted structure and reference structure.

Peptide structures predicted by Pepstr were also compared with the actual structures from PDB (**Table 3-2**). Alignment was done by Superpose and RMSD score was calculated.

Table 3-2 Structural RMSD difference between Pepstr predicted structures and reference structures of

HLA 0201 epitopes.

Epitope Sequence	Complex PDB ID	Chain	RMSD
ALWGFFPVL	1LP9	C	3.01
FAPGFFPYL	1I7R	C	5.97
GILGFVFTL	1B0R	C	4.35
GILGFVFTL	2VLR	C	4.81
GLMWLSYFV	3I6G	C	3.29
IISAVVGIL	1QR1	C	5.20
ILKEPVHGV	1P7Q	C	3.02
ILKEPVHGV	1HHJ	C	3.14
ILSALVGIL	1EEZ	C	5.23
ILSALVGIV	1EEY	C	6.06
IMDQVPFSV	1TVH	C	5.37
ITDQVPFSV	1TVB	C	3.60
LLFGKPVYV	2GIT	C	5.45
LLFGKPVYV	2GJ6	C	5.66
LLFGYAVYV	1QRN	C	4.37
LLFGYPRYV	1QSE	C	5.83
LLFGYPVAV	1QSF	C	6.08
LLFGYPVYV	1AO7	C	2.26
Average			**4.594444**

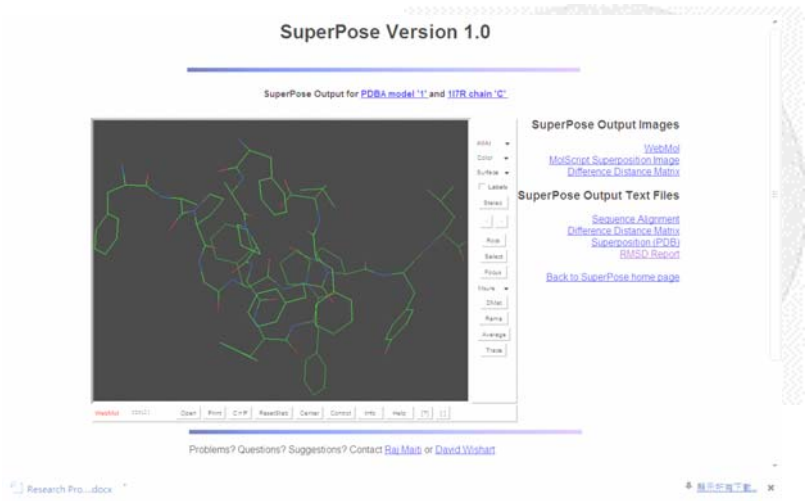


Figure 3-4 Illustrated comparison of SuperPose alignment between Pepstr predicted structure and reference structure.

The RMSD between structure predicted by bmPDA and actual structure was 0.81 (**Table 3-1**). The RMSD between structure predicted by Pepstr and actual structure was 4.59(**Table 3-2**). The performance of bmPDA is better than Pepstr (**Figures 3-3 & 3-4**).

3.3.2 Prediction efficiency of bmPDA on HLA1-binding omega-shape nona-peptide structures

We extract MHC-I HLA binding data from IEDB (Immune Epitope Database) as our training dataset.[42] It contains 3886 binding assay results about HLA-A*0201. The ROC analysis results are shown in **Figures 3-5**, **3-6**, and **3-7**. The difference between our QSAR-SVR model and NetCTL is shown in **Table 3-3**. Compared with other current epitope prediction NetCTL servers which ROC curves shown in **Figures 3-8**, the performance of our QSAR-SVR is statistically no difference with NetCTL (P>0.05).

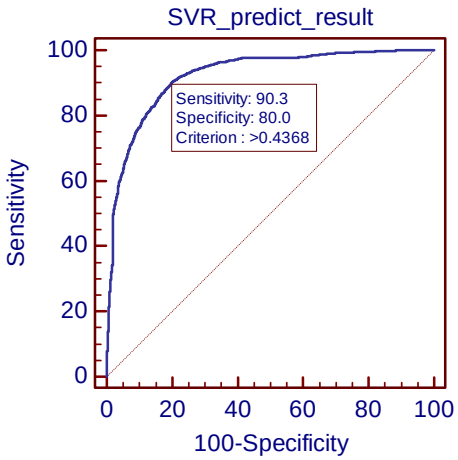


Fig. 3-5 Epitope prediction efficiency of bmPDA QSAR-SVR measured with ROC curve.

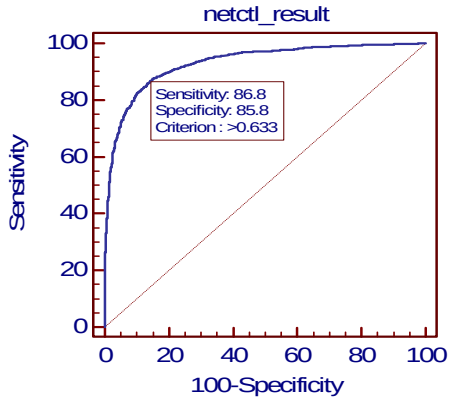


Fig. 3-6 Epitope prediction efficiency of NetCTL server measured with ROC curve.

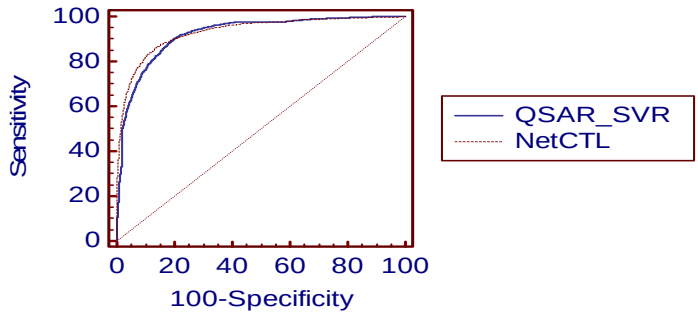


Fig. 3-7 Pairwise comparison on epitope prediction efficiency with ROC curves of bmPDA QSAR-SVR and NetCTL server.

Table 3-3 Summary of pairwise comparison on epitope prediction efficiency with ROC curves of bmPDA QSAR-SVR and NetCTL server.

	Area under curve AUC	Standard Error [a,c] SE	95% Confidence Interval [b] 95% CI
bmPDA QSAR_SVR	**0.924**	**0.00461**	**0.915 to 0.932**
NetCTL server	**0.932**	**0.00436**	**0.923 to 0.940**
Difference of pairwise AUC	**0.00804**	**0.00453**	**-0.000839 to 0.0169**
z statistic	**1.775**		
Significance level	**P = 0.076**		

a Hanley & McNeil, 1982
b Binomial exact
c Hanley & McNeil, 1983

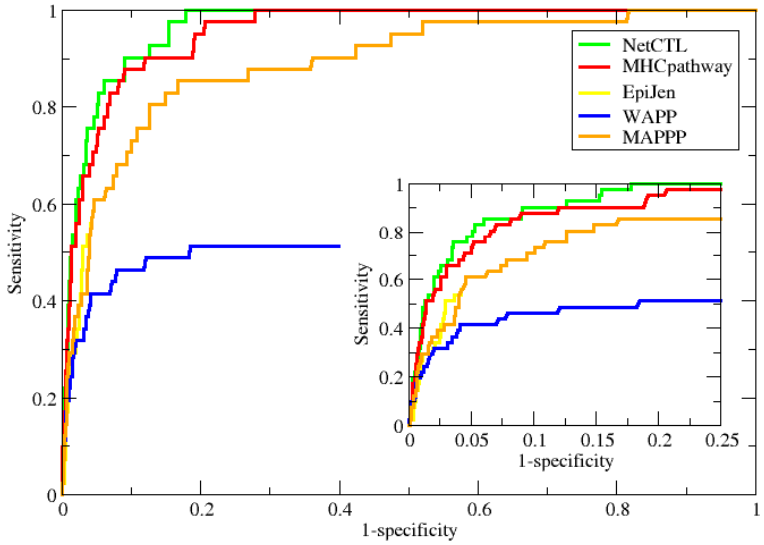


Fig. 3-8 Collective comparison on epitope prediction efficiency among five ROC curves of various web servers based on 41 A3 restricted epitope-protein pairs from the HIV dataset. [43]

The prediction efficiency of bmPDA QSAR-SVR tool and NetCTL server is respectively measured with statistical receiver operating characteristic (ROC) curves based on 3,886 A*02:01-binding assay results from IEDB as of the training dataset. The {bmPDA, NetCTL} pairwise comparison of ROC curves with pairwise values of sensitivity {90.3, 86.6}, specificity {80.0, 86.8}, criterion {> 0.4368, > 0.6330}, AUC {0.924, 0.932} of area under curve, SE {0.00461, 0.00436} of standard error, and 95% CI {0.915~ 0.932, 0.923~ 0.940} of 95% confidence interval as well as with pairwise difference value of AUC: 0.00804, SE: 0.00453, and 95% CI: (-0.000839~ 0.0169) in order for inferring without pairwise performance difference between bmPDA and NetCTL based on resulted significance level in P value at 0.076 and z statistic value at 1.775, .

3.4 Discussion

There are four major categories for the method of epitope prediction: sequence-based methods, structure based methods, hybrid methods and consensus methods.[1] The majority of epitope prediction methods are currently data-driven sequence-based, and they are more reliable than structure-based methods. However, there are several advantages for structured-based methods. First, only a smaller dataset is necessary for training. Second, it can predict peptides for alleles that have not been extensively studied. Third, discontinuous epitopes is only possible to predict by structure-based method. Last, even sequence-based approaches depend on structure information to make reliable predictions. However, the development of structure-based approach is still greatly limited due to high computational cost, development complexity and scarcity of 3D protein structures.[1]

In this study, we developed a method to predict MHC-I binding based on SVM. The prediction accuracy by ROC analysis is comparative as the best sequence-based method. In our method, relative small training set was employed, while other sequence-based method usually based on a huge data-mining process. However, the time-consuming peptide modeling process was the rate-determine step in our method. Because the relatively low demand for training dataset, our method may apply to other field such as MHC-II epitope prediction, IgE epitope prediction, etc.

The prediction efficiency of bmPDA with pairwise ROC comparison to NetCTL may suggest the compatible performance as outstanding discriminators in which our bmPDA may show slightly better sensitivity yet with slightly compromised specificity. In extension to NetCTL server capabilities, our bmPDA tool of structural prediction on omega-shape nona-peptide conformation and BAff value may offer appropriate power to assay simulated

binding onto HLA1 pit among nona-peptide candidates without public known structure information. Likely, our bmPDA tool may act as an extension tool for mining HLA1-binding peptides within NetCTL server related collaborative style despite that NetCTL server has established best performance efficiency with ROC analysis while among same category servers in efficiency order including MHC-pathway, EpiJen and MAPPP, and WAPP servers.

Chapter 4 Bio-mimicry Peptide Design on Reference Peptide Structure Modeling

4.1 Introduction

To design bio-mimicry peptide at functional portion of bioactive protein is an important bioinformatic task towards intended clinical applications with either agonistic or antagonistic activities. Traditional algorithms compare object surface structure at free rotations which may cause great time complexities through mining many putative structures that may not even exist. Instead, we develop a mining approach in this study based on existing known PDB peptide structures.

4.2 Method

The peptide structure generator software (**Figure 4-1**) is based on our implemented bmPDA tool of conformational anchor spacer hinge (CASH) algorithm in which bmPDA tool penta-pep (penta peptides, 5mer) database of is constructed from retracing segmental 5mer structures of all current PDB (protein data bank) entries as of basic building blocks of bmPDA tool database. Further, the segmental backbone angle of 5mer building blocks is exemplified with the aC[3] (alpha carbon 3rd) angle towards neighboring aC[2] and aC[4] as of the middle 3mer in each 5mer building blocks.

With reference 12mer oligo peptide for finding candidate peptides with mimicking structure of different sequences, the serial penta peptides from reference oligo peptide are generated for calculating the described aC[3] angle. Importantly, both frontal and coda aC[3] angle are respectively replaced with aC[2] angle and aC[4] angle. All aC[n] angles of reference penta peptides are applied for searching bmPDA 5mer building blocks with similar aC[n] angle of backbone yet with different amino acid sequence. The resulted bmPDA 5mer building blocks with serially similar aC[n] angles to reference penta peptides are accordingly assembled to form contig sequences based on overlapping amino acid subsequences, such as the KAV and VYN towards KAVYN sequence contig.

For forming structural contig based on sequence contig, the identified bmPDA 5mer building blocks within different coordination systems are unified into identical coordination system by means of shift and rotation as of Quaternion system yet with no rescaling due to constant atom distances. The exemplified KAVYN merge from the described KAV and VYN with most similar backbone structure with reference oligo peptide is accomplished on superimposing both amino acids of both KAV and VYN efficiently by means of matching at 3 points including N[n], aC[n], and C[n].

Towards mimicking structure of bmPDA structural contig versus reference peptide based on similar backbone structure, the distance between side chain and backbone aC[n] as of conformational index is subsequently applied for mining best contig structures. The similarity evaluation on merged structures with reference structure is accomplished by checking physiochemical property, energy stability, docking fitness with counter-structures of reference peptide.

The overall workflow of bio-mimicry peptide design algorithm is shown in **Figure 4-1**.

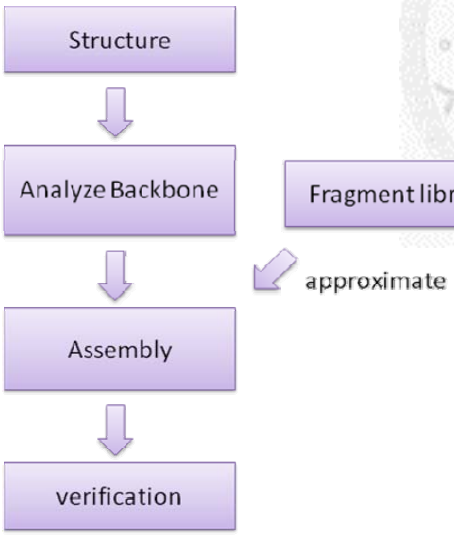


Fig. 4-1 Workflow of bio-mimicry peptide design algorithm (bmPDA).

4.3 Result

The implemented in-house bio-mimicry peptide design algorithm tool (bmPDA-tool) comprises three sections including constructed peptide building blocks database, assembled peptide model of building block candidates, and predicted peptide model of functional

peptides.

4.3.1 Bio-mimicry peptide structure design: Reference vasopressin

Part of Bio-mimicry prediction results of vasopressin (**Figures 4-2**) is listed in **Table 4-1**. Illustrations of the structure of better solutions from bmPDA-prediction are shown in **Figures 4-3** and **4-4**. The backbone distance of Solution no. 12 (KGN-SVL-AIP) is 1.736.

The backbone distance of Solution no. 12 (KGN-SVL-AIP) is 1.736. The backbone distance of Solution no. 37 (DGN-SVL-ADS) is 2.300. The structure of solution no. 36 (DGN-SVL-AIP) and no. 279 (SEA-SKQ-TAA) are also shown in **Figures 4-5** and **4-6**.

Table 4-1 Bio-mimicry peptide sequences of bmPDA designed vasopressin structure sorted in backbone distance order along with appropriate peptide values of Morris class and G-factor.

Solution no.	Structure Sequence	Backbone Distance	Morris class	G-factors
@289	CYF-QNC-PRG	0	2 3 1	0.17
3	GEA-SGS-SQV	1.723	4 1 1	0.08
* 12	KGN-SVL-AIP	1.736	1 1 1	0.24
13	KGN-SVL-ADS	1.755	1 1 1	0.23
221	WKG-RTW-EPA	1.852	1 3 1	-0.19
278	SEA-SGS-SQV	2.148	4 1 1	0.12
272	SEA-SGS-STP	2.233	4 1 1	-0.10
* 37	DGN-SVL-ADS	2.300	1 1 1	0.13
36	DGN-SVL-AIP	2.306	1 1 1	0.13

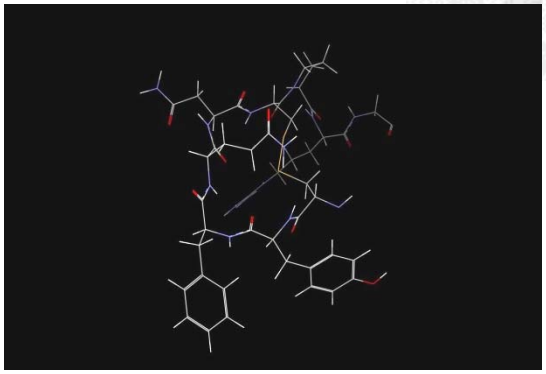


Fig. 4-2 Actual reference peptide structure of vasopressin 9-mer [1YF4] CYF-QNC-PRG in bar.

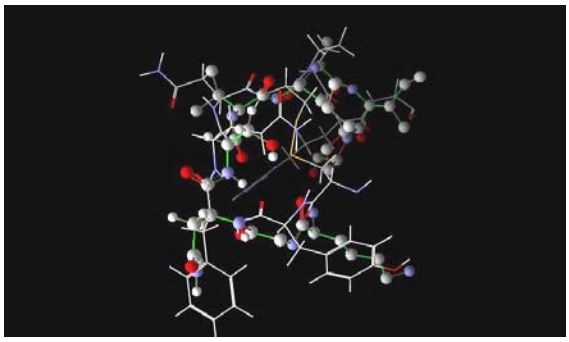


Fig. 4-3 Bio-mimicry reference peptide structure of vasopressin 9-mer [1YF4] CYF-QNC-PRG with selected bmPDA designed solution number 12 in ball-stick format aligned with Vasopressin overall structure in bar.

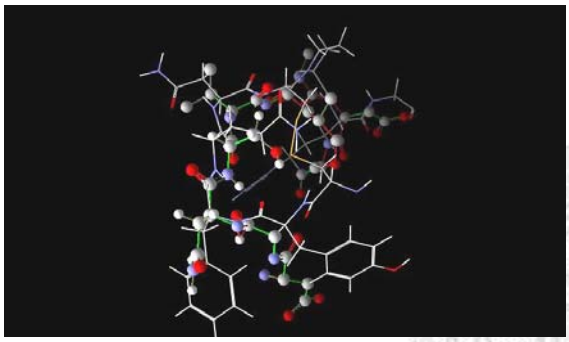


Fig. 4-4 Bio-mimicry reference peptide structure of vasopressin 9-mer [1YF4] CYF-QNC-PRG with selected bmPDA designed solution number 37 in ball-stick format aligned with Vasopressin overall structure in bar.

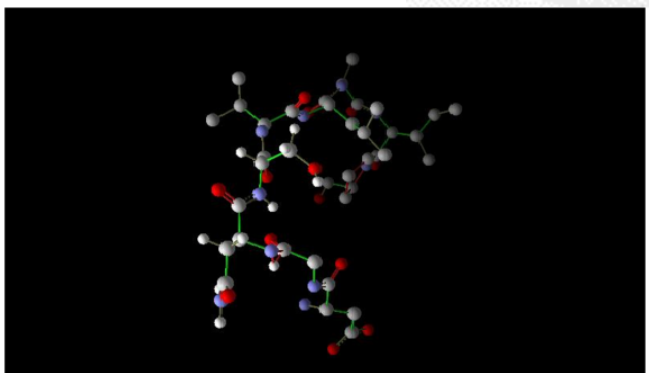


Fig. 4-5 Bio-mimicry reference peptide structure of vasopressin 9-mer [1YF4] CYF-QNC-PRG with selected bmPDA designed solution number 36 in ball-stick format.

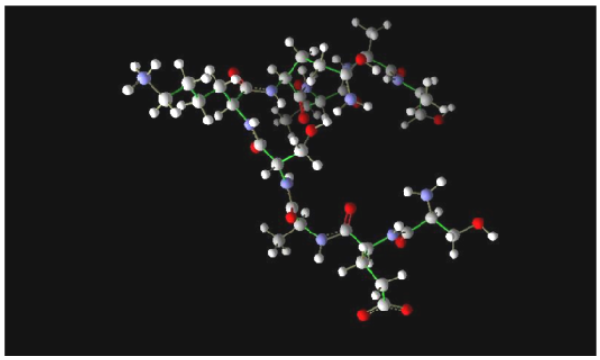


Fig. 4-6 Bio-mimicry reference peptide structure of vasopressin 9-mer [1YF4] CYF-QNC-PRG with selected bmPDA designed solution number 279 in ball-stick format.

4.3.2 Antibody vaccine peptide design: TMB-355 bio-mimicry epitope

TMB-355 is a monoclonal antibody for treatment of HIV. Structure of light chain of TMB-355 is shown in **Figure 4-7** and **4-8**. Part of Bio-mimicry prediction results of Fab region of TMB-355 is listed in **Table 4-2**. Actual reference peptide structure of TMB-355 light chain (QYY-SYR-TFG-GGT) is shown in **Figure 4-9**. The bmPDA-predicted peptide structure of solution no. 2765 (YIGSGKKTAGAG) and no. 957 (QIGSGKKASG) are shown in **Figures 4-10** and **4-11**.

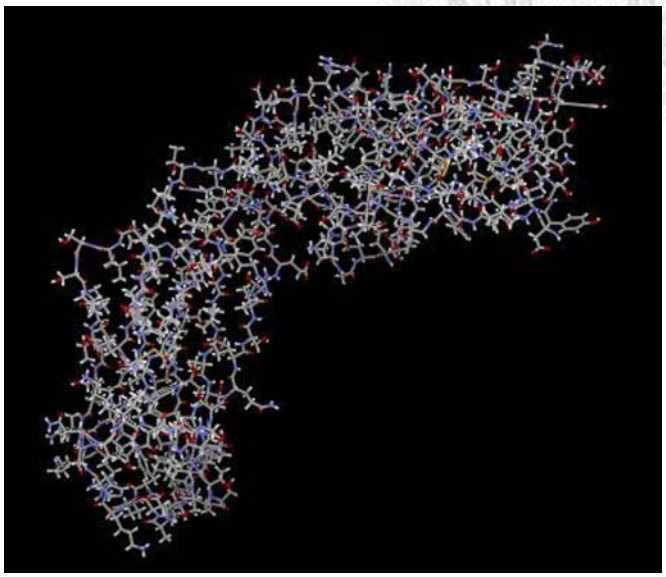


Fig. 4-7 Actual reference peptide structure of TMB-355 light-chain in bar.

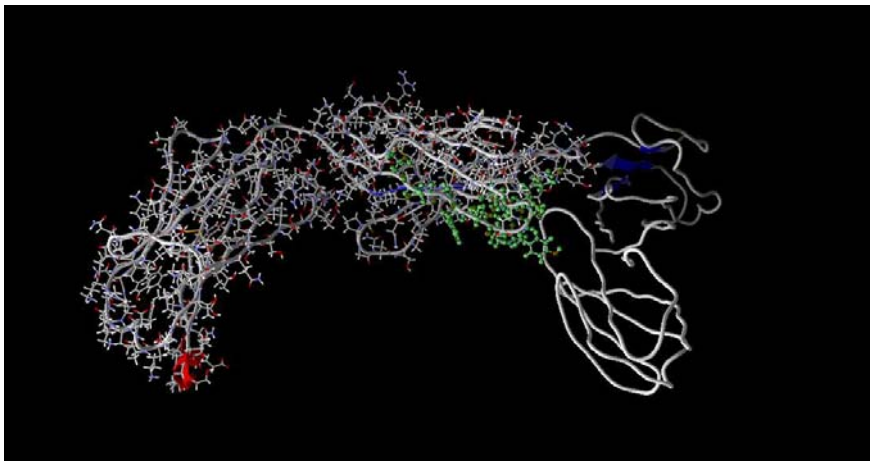


Fig. 4-8 Actual reference peptide structure of TMB-355 light-chain with green dots indicating the target structure at hypervariable region.

Table 4-2 Bio-mimicry peptide sequences of bmPDA designed structure solutions of TMB-355 light chain Fab region [QYY-SYR-TFG-GGT] sorted in backbone distance order along with appropriate peptide values of Morris class and G-factor.

Solution no.	Structure Sequence	Backbone Distance	Morris class	G-factors
3053	YSERQLTTFGDK	2.7311	3 1 4	-0.05
3692	ESEKLKYKVLAS	2.7845	2 2 4	0.05
3384	DSERQLTTFGDK	2.8166	3 1 1	-0.01
3794	HPAAGVADGSRR	3.1078	3 2 3	-0.02
3680	CNYTDKKPVLRS	3.3344	3 4 1	-0.22
3681	CNYTDKKPVLRT	3.3522	3 4 1	-0.20
3712	QNGTVLEGPTTG	3.4233	2 2 4	-0.19
2765	YIGSGKKTAGAG	3.4457	2 1 1	0.12
2511	VNTVLNGGIRKI	3.4905	4 1 1	-0.10

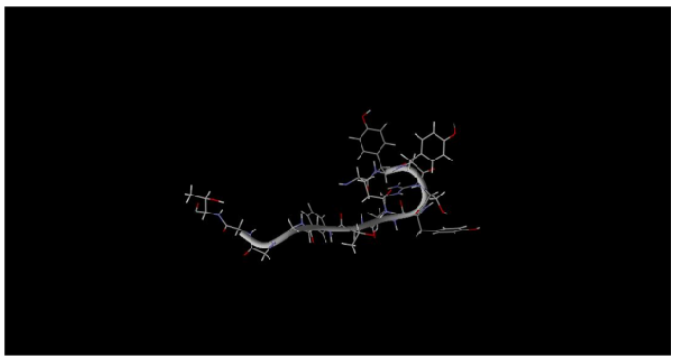


Fig. 4-9 Actual reference peptide structure of the target structure at TMB-355 light chain hypervariable region as shown in Figure 4-8 in green dots.

Fig. 4-10 Bio-mimicry target peptide structure of TMB-355 light chain Fab region with bmPDA designed structure of solution number 2765 in ball-stick format aligned with overall Fab structure in bar.

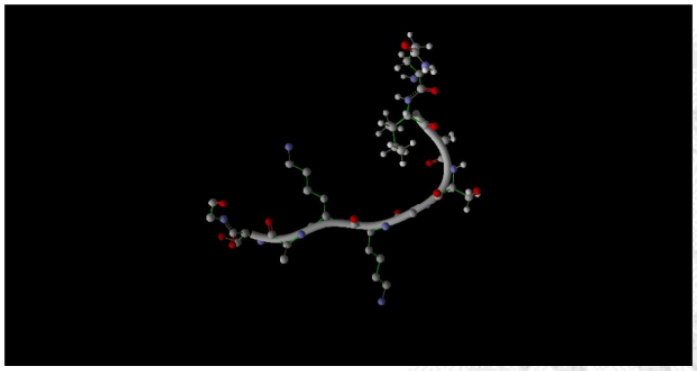


Fig. 4-11 Bio-mimicry target peptide structure of TMB-355 light chain Fab region with bmPDA designed structure of solution number 957 in ball-stick format aligned with overall Fab structure in bar.

4.4 Discussion

Based on our approach, we can perform an inverse folding search on bioactive peptides. Toward peptide backbone alignment, accompanied with physiochemical properties of residues, we can search bio-mimicry peptides more efficiently and accurately.

Chapter 5 Vaccine Peptide Evaluating towards Epitope and/or Agretope Plastic Modeling

5.1 Introduction

Nasopharyngeal carcinoma (NPC) is a squamous cell carcinoma that occurs on the epithelium of the nasopharynx.[2] It is a common malignancy in southern China, Hong Kong, and south-east Asia countries including Taiwan, Singapore, Malaysia, Indonesia, and Vietnam.[3] Genetic susceptibility, environment factors, and Epstein-Barr virus (EBV) are thought to play roles in the development of NPC.

The treatment nowadays is based on radiotherapy and concurrent chemoradiotherapy. In spite of improvement of treatment outcomes, local regional failure and distant metastasis still occur in many patients.[3] Moreover, acute side effects and long-term sequelae including secondary malignancy are often accompanied with radiation and chemotherapy. Therefore, novel approaches aiming to improve outcome and reduce the need for conventional cytotoxic therapies are under developed.

Eradication of local regional microscopic and micrometastatic disease with associated minimal toxicity to surrounding normal cells is one of the goals of adjuvant cancer therapy.[11] Immunotherapy is therefore an attractive option. However, the major obstacle of immunotherapy to cancer is absence of suitable molecularly characterized tumor antigens.[44] Before the human tumor-associated antigens (TAA) were identified, immunotherapists were forced to use undefined tumor antigens derived from tumor cell lines, tissues or their corresponding lysates.[11] With the identification of a large series of TAAs and advancement of molecular genetics, antigen-specific immunotherapy became possible in these years. The main advantage of antigen-specific immunotherapy is the capability to evaluate and monitor immune responses to targeted antigens and correlate these findings with clinical responses.[11]

EBV is a member of the herpesvirus family.[4] It has a double-stranded DNA genome of 184-kb pairs in length, encoding nearly 100 proteins.[5] It was the first virus to be associated

to human cancer. EBV attack B-lymphocyte as primary target, resulting in lifelong infection.[5] In spite of being a latent infection in B cells, inhibition by a population of EBV-specific cytotoxic T lymphocytes (CTLs) was observed.[7] Both in vitro and in vivo, these CTLs have been shown to have potent antiviral activity.

There are many viral antigens being expressed during EBV latent infection. Unlike other EBV-associated diseases, NPC expresses only some less immunogenic viral antigens, including EBNA1, LMP1, and LMP2.[8] According to literatures, EBNA antigens are poorly processed by antigen-presenting cells, so LMP1 and LMP2 are better potential targets.[9] However, according to assays about these sequences, LMP1/LMP2 is not only highly oncogenic but also seems to be poorly immunogenic in murine models.[10] Vaccination or immunotherapy based on full-length LMP1 is therefore not recommended, and polyepitope vaccine based on multiple immunogenic epitope is a preferred strategy. Fortunately, based upon the virus isolations from different geographic regions of the world, sequence analysis revealed that most of these epitopes are highly conserved and are efficiently recognized by individuals of diverse ethnic origin.[9] By using polyepitope vaccine comprising HLA class I-restricted CTL epitopes from LMP1 and LMP2, some studies makes promising progress in controlling tumor growth in animal model.[10, 45]However, the epitopes being used in previous studies were restricted in a relatively narrow spectrum of HLA class I alleles. Although HLA A2 is one of the most common HLA class I alleles, other HLA class I alleles prevalent in NPC endemic regions of the world (HLA A11, A24, B27, and B57) should also be included in LMP-based polyepitope vaccine design.[10]

In this study, we reviewed the epitopes being used in previous studies and compared with the prediction result of bioinformatic approach, and try to make additional suggestion about other potential epitopes which can be helpful for NPC vaccine design.

All nucleated cells present a selection of the peptides contained in their proteins on the cell surface in complex with MHC-I. Cytotoxic T lymphocytes (CTL) can then differentiate between healthy cells and infected cells. However, there are only a small fraction of the peptides in a pathogen proteome being able to induce a CTL response. This is primarily due to the selection process in the antigen-processing steps preceding the CTL response. There is only 1 out of 2000 potential peptides will be immunodominant for each MHC-I allele.[46]

Generation of peptides from their precursor polypeptides is necessary for the induction of a CTL response. Proteasome is the major cytosolic protease associated with the generation of antigenic peptides. After proteasomal cleavage the peptides may be trimmed at the N-terminal end by other peptidases in the cytosol. The next step is transporting of the

peptides from the cytosol to the interior of the ER. Binding of the peptides to TAP can facilitate the transportation.

Further N-terminal trimming of the peptides then proceeds inside the ER, and binding of some of the peptides to MHC-I is also done. The MHC-I:peptide complex is then transported to the surface of the cell, where it may be recognized by CTL. Binding to MHC-I is the most restrictive step in antigen presentation. Estimation for the selectivity showed that about only 1 out of 200 peptides will bind a given MHC-I allele with sufficient strength to elicit a CTL response. However, the proteasomal cleavage and the TAP transport efficiency play some roles.[46]

The most predictable part of T cell epitope generation is peptide-MHC binding. MHC-I and MHC-II genes are highly polymorphic, and the most of their variable part are located in binding pockets that restrict peptide interactions to those with particular amino acids at characteristic positions.[39]

In the past, vaccine development depends on biochemical and immunological experiment, such as phage display library, overlapping peptides, ELISA, NMR, immunofluorescence, radioimmunoassay, Western blotting, immunohistochemistry, X-ray crystallography studies of antibody/antigen structure and attenuation of the wild type pathogens by random mutations and serial passages, which is very expensive, time-consuming, with low immunogenicity and reversible.[1] Under the help of epitope prediction approach, we can narrow the spectrum of target proteins, and reduce the cost of wet experiments.

There are four approaches being applied to predict epitopes: sequence-based methods, structure based methods, hybrid methods and consensus methods.[1] The majority of epitope prediction methods are currently data-driven sequence-based, and they are more reliable than structure- based methods. On the other hand, there are several advantages for structured-based methods. First, only a smaller dataset is necessary for training. Second, it can predict peptides for alleles that have not been extensively studied. Third, discontinuous epitopes is only possible to predict by structure-based method. Last, even sequence-based approaches depend on structure information to make reliable predictions. However, the development of structure-based approach is still greatly limited due to high computational cost, development complexity and scarcity of 3D protein structures.[1]

5.2 Method

There are many viral antigens being expressed during EBV latent infection. Unlike other EBV-associated diseases, NPC expresses only some less immunogenic viral antigens, including EBNA1, LMP1, and LMP2. According to literatures, EBNA antigens are poorly processed and presented by antigen-presenting cells, so LMP1 and LMP2 are better potential targets. Cytotoxic T lymphocytes have been thought to play the key role in the generation of antitumor therapeutic effects, so we focus on the antigenecity of CTL.

5.2.1 Collections of protein sequences and structures

Protein sequence of EBV LMP1 and LMP2 were collected from National Centre for Biotechnology Institute (NCBI) database. Prevalent HLA class I locus of the southern Han Chinese population are A2, A11, A24, A33, B13, B15, B38, B40, B46, B58, C1, C3, C7, and C8. Structure files of MHC class I molecules were collected from PDB.

5.2.2 Modeling of predicted epitopes

The models of each predicted epitope were done by our peptide modeling method described in chapter 2. Instability and Ramachandran Plot Analysis were done by Pro-Check program.

5.2.3 NetCTL

NetCTL is a server who integrates predictions of proteasomal cleavage, transporter associated with antigen processing (TAP) transport efficiency, and MHC-I binding affinity into a MHC-I pathway likelihood score. All MHC-I molecules predictions were done by the NetCTL with known protein sequence and predictions for 8-, 9-,10-, and 11-mer peptides were also achieved. Optimization was done to achieve high specificity in order to maintain a low false positive rate.

The web-based NetCTL tool models human CTL epitopes in any given protein by integrating predictions of proteasomal cleavage, TAP transport efficiency, and MHC1-binding affinity. In addition, the web-based SYFPEITHI tool evaluates human CTL epitopes by the scoring system on given epitope peptide at individual amino acids respectively assigns arbitrary positive value within {+1, +15} range based on the preference scale from minimal to optimal

despite that the position-specific dislike of amino acids may be assigned with arbitrary negative value within {-1, -15} range. The arbitrary values are allocated on frequency of respective amino acid in natural ligands, T-cell epitopes, or binding peptides.

With bmPDA tool of TPSS and QSAR-SVR in addition to NetCTL server on integrated antigenecity score, the respective epitope prediction efficiency of NLMP1 and LMP1 is tested with statistical approach of receiver operating characteristic (ROC) curve for determining the discrimination power with respective efficiency based on derived values of area under curve (AUC), sensitivity, specificity, and criterion along with pairwise difference of prediction performance based on derived significance level in P value.

5.2.4 Residue Preference of Epitope

Based on the concept of each MHC molecule potentially presents a distinct set of antigenic peptides to the immune system, effort had been made to predict binding motif of different MHC alleles. Preferred residue of HLA-A*02:01 (from IEDB) on specific positions is shown in **Figure 5-1**. With hypothesized immune evasion at weak viral agretope anchors, our bmPDA tool selects putative NLMP1 nona-peptide structures with weak BAff on docking A*02:07 pit while compared with A*02:01 in order for evaluating BAff improvement on Ama candidates towards making in vitro DNA vaccine.

The anchor plastics strategy for Ama candidate is based on inferred preference as of A*02:01 case at respective positions {2nd; 9th} of nona-peptide with strong anchor {L/M; I/L/V} rather than with tolerated anchor {I/Q/V; A/M} in addition to an extra preference at position {3rd} with {D/P} as with A*02:07 case only.

Preferred Position	F Y	Position	FM W		W Y		FW		C-terminal
	1	2	3	4	5	6	7	8	9
Deleterious	DEP	Anchor	EKR			KR	GKR		Anchor
		Preferred LM							Preferred ILV
		Tolerated IQV							Tolerated AM

Fig. 5-1 Reference HLA-A*02:01 binding nona-peptides with anchor residue preferences in respective agretopes from experiment verified IEDB data in part showing high consistency with bmPDA identified LMP1 agretope candidates exemplified with VMSD in order for serving as strong agretope substitution into weak agretope segment while attached with strong immunogenic epitope segment.

5.2.5 Molecular Docking

The interactions between the predicted peptide models and MHC-I are evaluated by Molegro Virtual Docker (MVD). [47] Center of MHC-I groove is identified as potential docking site to achieve more accurate docking result. In MVD, a heuristic search algorithm named MolDock is used for accurate molecular docking. For further docking accuracy improvement, a re-ranking scoring function is then introduced. They are useful for evaluation between different poses of the same ligand. However, the MolDock and rerank score in MVD are not expressed in chemically relevant units. For comparison between multiple ligands, we use a regression model for binding affinity estimation. Followed by 'Tabu clustering', more favorable binding result is achieved in an efficient way.[48]

Optimized NLMP1np {X66863} and LMP1np {V01555} structures towards docking A*02:07 {3OXS} and A*02:01 {1BD2} pits are evaluated with regression model of Molegro Virtual Docker (MVD) on docking free energy as of binding affinity (BAff) with Tabu clustering score in order to avoid sub-optima and local optima along with NetCTL and SYFPEITHI web-servers on putative CMI antigenecity with appropriate serve scores.

5.3 Result

5.3.1 Putative omega-shape nona-peptide structure predicted with TPSP library and TPSS database

The optimal omega-shape conformation of NLMP1np structure is mined from massive combinatorial candidates in which are modeled with TPSP candidates through computation process with both bmPDA modules of Q4-GA and QSAR-SVR towards the similar structure of positive reference templates extracted from PDB dataset. Epitope prediction results from NetCTL for LMP1 were plotted in **Figure 5-2**. Epitope prediction results from SYFPEITHI for LMP1 were listed in **Figure 5-3**. Better epitopes of LMP1 is listed in **Table 5-1**. The predictions were compatible with experiment data from IEDB (**Figure 5-4**).

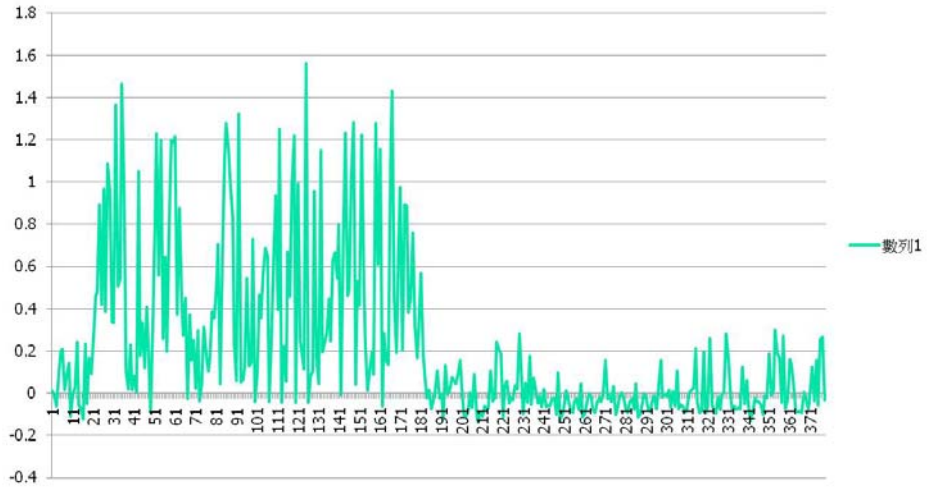


Fig. 5-2 Overall LMP1 epitope antigenecity of bmPDA predicted peptide structures analyzed by NetCTL antigenecity score server towards HLA A*0201.

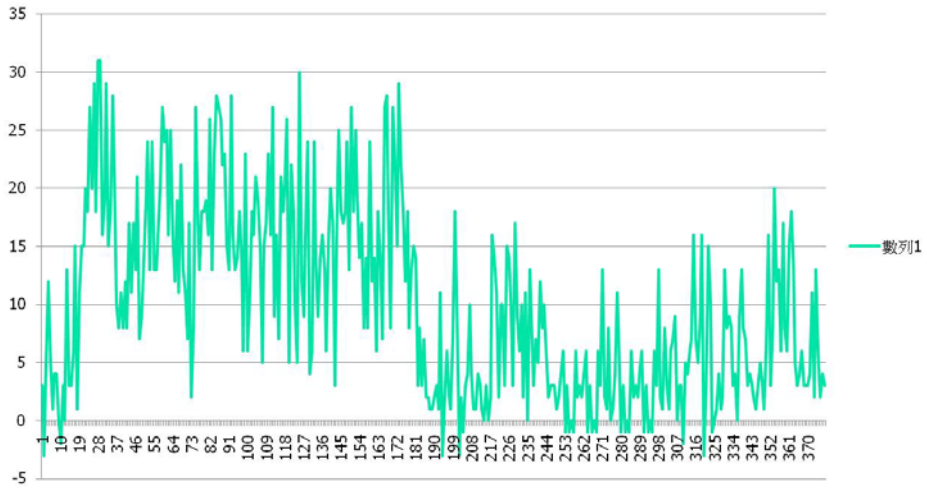


Fig. 5-3 Overall LMP1 epitope antigenecity of bmPDA predicted peptide structures analyzed by SYFPEITHI antigenecity score server towards HLA A*0201.

Table 5-1 Top ranking epitope segments of LMP1 selected by NetCTL and SYFPEITHI towards HLA A*0201 with MVD binding affinity attached.

number	sequence	MVD binding affinity	NetCTL	SYFPEITHI
125	YLLEMLWRL	-13.3300	1.5637	30
35	ALLFWLYIV	-13.4630	1.4636	28
167	LLVDLLWLL	-13.8151	1.4324	28
32	LLLALLFWL	-12.5441	1.3659	29
92	LLLIALWNL	-16.1708	1.3255	28
148	FLDLILLII	-18.5444	1.2840	24
86	LLLMITLLL	-3.28964	1.2779	27
112	FIFGCLLVL	-13.6547	1.2517	27
144	FLAFFLDLI	-16.0035	1.2339	25

Export all results: (compact | full)

ID	Reference	Epitope	Host	Immunization	Assay Antigen	Antigen Epitope Relation	MHC Restriction	Assay Description
1377922	R Khanna; Eur J Immunol 1998	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Homo sapiens	Immune reactivity to Human herpesvirus 4 (Taxonomic Parent) without evidence for disease followed by restimulation in vitro	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Epitope	HLA-A2	51 chromium release cytotoxicity **Positive**
1377923	R Khanna; Eur J Immunol 1998	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Homo sapiens	Immune reactivity to Human herpesvirus 4 (Taxonomic Parent) without evidence for disease followed by restimulation in vitro	Latent membrane protein 1 Latent membrane protein 1 Human herpesvirus 4 (strain B95-8)	Source Antigen	HLA-A2	51 chromium release cytotoxicity **Positive**
1377924	R Khanna; Eur J Immunol 1998	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Homo sapiens	Immune reactivity to Human herpesvirus 4 (Taxonomic Parent) without evidence for disease followed by restimulation in vitro	Human herpesvirus 4 (strain B95-8) Human herpesvirus 4 (strain B95-8)	Source Organism	HLA-A2	51 chromium release cytotoxicity **Positive**
1377925	R Khanna; Eur J Immunol 1998	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Homo sapiens	Immune reactivity to Human herpesvirus 4 (Taxonomic Parent) without evidence for disease followed by restimulation in vitro	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Epitope	HLA-A*02:01	51 chromium release cytotoxicity **Positive**
1377926	R Khanna; Eur J Immunol 1998	YLLEMLWRL Latent membrane protein 1 (125-133) Human herpesvirus 4 (strain B95-8)	Homo sapiens	Immune reactivity to Human herpesvirus 4 (Taxonomic Parent) without evidence for disease followed by restimulation in vitro	Human herpesvirus 4 (strain B95-8) Human herpesvirus 4 (strain B95-8)	Source Organism	HLA-A*02:01	51 chromium release cytotoxicity **Positive**

Fig. 5-4 Reference HLA-1 antigenic peptides of LMP1 selected from experiment verified IEDB data in part showing high consistency with our bmPDA designed LMP1 vaccine peptide candidates.

5.3.2 HLA1 vaccine peptide design: LMP1 agretope prediction

Agretope docking result from our method for LMP1 is plotted in **Figure 5-5**. Better agretopes of LMP1 is listed in **Table 5-2**. According to NetCTL, SYFPEITHI, and the docking result of our method, the preferred epitope for vaccine design were listed in **Table 5-3**.

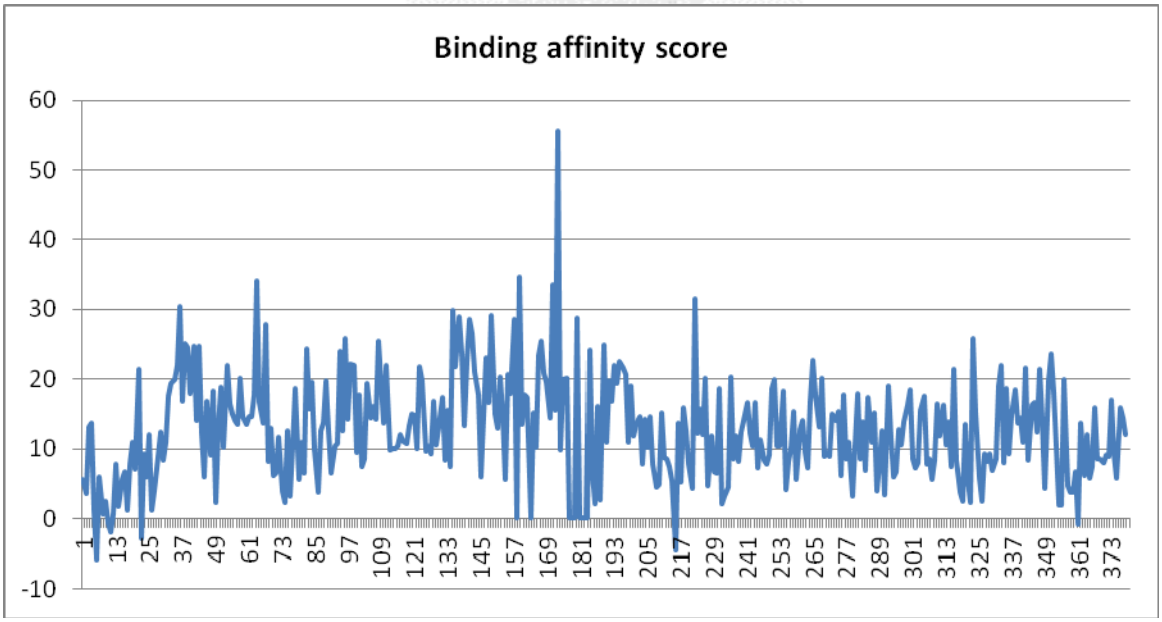


Fig. 5-5 LMP1 agretope docking scores of bmPDA predicted peptide structures onto HLA A*0201 pocket sub-zones evaluated by Molegro Virtual Docker software.

Table 5-2 Top ranking agretopes to HLA A*0201 of LMP1 selected by MVD binding affinity.

number	sequence	MVD binding affinity	NetCTL	SYFPEITHI
173	WLLLFLAIL	-55.6000		
159	YLQQNWWTL	-34.6364		
64	IIILIIFIF	-34.1266		
171	LLWLLLFLA	-33.6373		
222	GRHHLLVSG	-31.5068		
36	LLFWLYIVM	-30.4040		
135	ATIWQLLAF	-29.8328		
149	LDLILLIIA	-29.0841		
137	IWQLLAFFL	-28.9083		
180	ILIWMYYHG	-28.6749		

Table 5-3 The bmPDA-designed priority LMP1 vaccine peptide candidates towards HLA A*0201 sorted by [agretope – epitope – agretope] docking scores of MVD binding affinity in comparison among antigenicity scores of NetCTL and SYFPEITHI.

position	sequence	MVD Binding Affinity	NetCTL	SYFPEITHI
173	WLLLFLAIL	-55.59999686	0.9859	29
159	YLQQNWWTL	-34.6364165	1.2840	24
64	IIILIIFIF	-34.12661984	0.4515	16
171	LLWLLLFLA	-33.63727296	1.0874	22
222	GRHHLLVSG	-31.50681184	0.0081	10
36	LLFWLYIVM	-30.4039776	1.0005	18
135	ATIWQLLAF	-29.83278843	0.3298	14
149	LDLILLIIA	-29.08412344	0.0862	13
137	IWQLLAFFL	-28.90825584	0.2581	13
180	ILIWMYYHG	-28.67487832	0.2667	15

5.3.3 Putative LMP1 nona-peptide structures of assorted anchors onto docking A*02:01 pit

Based on the existing experimental data, many methods had been developed for predicting MHC binding motif. Information about preferred residues on specific position was also integrated into IEDB. According to peptide MHC binding motif data from IEDB, substitution of residue on position 2 and C-terminal anchors of potential HLA-I: A*02:01-binding epitopes was done. After structure prediction by bmPDA and docking by MVD, binding affinity result summary is shown in **Table 5-4**.

Table 5-4 The bmPDA-designed LMP1 priority vaccine peptide candidates with improved [agretope – epitope – agretope] HLA-I A*02:01 docking scores in comparison with improved NetCTL and SYFPEITHI antigenecity scores in which original bmPDA-predicted peptide structures in weak agretope group I/Q/V and A/M are modified towards new bmPDA-predicted peptide structures with

substituted strong agretope group L/M and I/L/V.

	Sequence	NetCTL	SYFPEITHI	MVD binding affinity
	YLLEMLWRL	1.5754	30	-10.2532
	YMLEMLWRL	1.5857	28	-18.4164
	YLLEMLWRI	1.5376	28	-13.6425
*	YLLEMLWRV	1.5570	30	**-33.7690**
	YMLEMLWRI	1.5349	26	-19.1742
	YMLEMLWRV	1.5595	28	-13.0074
	ALLFWLYIV	1.4500	28	-20.5003
	AMLFWLYIV	1.5000	26	-30.8523
	ALLFWLYII	1.3600	26	-25.0526
*	ALLFWLYIL	1.4000	28	**-38.8726**
	AMLFWLYII	1.4000	24	-17.3758
	AMLFWLYIL	1.4500	26	-28.7211
	LLVDLLWLL	1.4700	28	-11.1361
	LMVDLLWLL	1.4936	26	-9.32915
*	LLVDLLWLI	1.4374	26	**-13.2684**
	LLVDLLWLV	1.4664	28	-5.6147
	LMVDLLWLI	1.4464	24	-11.1991
	LMVDLLWLV	1.4837	26	-10.3239
	LLLALLFWL	1.4195	29	-13.1378
	LMLALLFWL	1.4403	27	-22.6202
	LLLALLFWI	1.3509	27	-13.7874
	LLLALLFWV	1.4105	29	-12.4064
	LMLALLFWI	1.3649	25	-17.6992
*	LMLALLFWV	1.4302	27	**-35.1014**
	LLLIALWNL	1.3461	28	-16.6802
	LMLIALWNL	1.3932	26	-20.0396
	LLLIALWNI	1.2997	26	-10.9213
	LLLIALWNV	1.3755	28	-13.8491
	LMLIALWNI	1.3230	24	-18.0534
*	LMLIALWNV	1.4087	26	**-26.6008**
*	**FLDLILLII**	1.3504	24	**-24.3467**
	FMDLILLII	1.3486	22	-14.3563
	FLDLILLIL	1.3256	26	-8.5435
	FLDLILLIV	1.4062	26	-17.9722
	FMDLILLIL	1.3934	24	-7.9018
	FMDLILLIV	1.4400	24	-16.6111
	LLLMITLLL	1.3029	27	-13.1837
*	LMLMITLLL	1.3564	25	**-20.2986**
	LLLMITLLI	1.2659	25	-9.8575
	LLLMITLLV	1.3428	27	-12.3941
	LMLMITLLI	1.3035	23	-15.3174
	LMLMITLLV	1.3924	25	-10.4206
	FIFGCLLVL	1.3194	27	-13.8768
	FLFGCLLVL	1.4435	29	-26.7132
	FMFGCLLVL	1.4676	27	-26.5304
	FIFGCLLVI	1.2739	25	-11.8059
*	FIFGCLLVV	1.3658	27	**-27.2390**
	FLFGCLLVI	1.4128	27	-13.2616
	FLFGCLLVV	1.4587	29	-21.5721
	FMFGCLLVI	1.4262	25	-15.5724
	FMFGCLLVV	1.4768	27	-25.2802

After residue substitution, we collected more peptides with relative better NetCTL and SYFPEITHI antigenecity scores on HLA-I A*0201. Binding affinity was then predicted by MVD. List of epitopes with good epitope score (NetCTL and SYFPEITHI antigenecity scores)

and agretope score (MVD binding affinity) is shown in **Table 5-5**. Initially, the bmPDA tool designs omega-shape nona-peptide structures only upon the top 8 ranking candidates including 1/125 (1.570), 2/167 (1.470), 3/035 (1.450), 4/032 (1.420), 5/148 (1.350), 6/092 (1.346), 7/112 (1.319), and 8/086 (1.303) of original LMP1np structures which are conveniently selected on the NetCTL and/or SYFPEITHI antigenicity scores after performing nona-peptide scanning throughout full length LMP1 amino acid sequence.

The appropriate omega-shape conformation of nona-peptide structures for docking A*02:01 pit including top 8 LMP1np candidates are selected from massive combinatorial structures which are modeled with TPSP candidates through computation process with both bmPDA Q4-GA and QSAR-SVR modules according to the positive reference template structures extracted from PDB dataset. For docking onto A*02:01 pit structure, the applied LMP1np structures based on the above LMP1np order of NetCTL score ranking may show rather inconsistent ranking order along with greater difference in simulated MDV BAff values as of 8/125 (-10.253), 7/167 (-11.136), 2/035 (-20.500), 6/032 (-13.138), 1/148 (-24.347), 3/092 (-16.680), 4/112 (-13.877), and 5/086 (-13.184) with BAff ranking indicated.

Table 5-5 LMP1 candidate epitopes with good epitope score and agretope score

Modified Epitopes
YLLEMLWRV
ALLFWLYIL
LLVDLLWLI
LMLALLFWV
LMLIALWNV
FLDLILLII
LMLMITLLL
FIFGCLLVV
FMAFFLDLI
LMVLYSFAI

5.4 Conclusion and Future Works

Knowledge of the 3D structure of epitopes is essential in structural immunoinformatics. With the structure of peptides binding to MHC molecules, further elucidation about immune reactions such as epitope-MHC molecular interactions can be done. There are many methods developed for protein structure prediction, however, there are relatively few methods for short sequence peptide structure prediction. In the immunoinformatic field, epitope is primarily the target of concern, which is consisted of about 8-12 amino acids. There is little previous effort for structural prediction of peptides binding to MHC Class I molecules.[49]

The bmPDA structure prediction result of epitope revealed our method was good at predicting structure of short sequence of peptide such as epitope. This is compatible with the previous experience that protein threading method is more accurate in short sequence of protein than homology modeling.

Based on the structure prediction result, we can actually simulate the binding between peptide and MHC molecules, instead of other indirect method. For binding affinity evaluation between peptide and MHC class I molecule, docking is done with MVD regression model based on tabu clustering in order to avoid GA sub-optima and GA local optima.[48]

The intended vaccine peptide of epitope and agretope may be delivered in the format of "*in silico* DNA vaccine" which is constructed with expression DNA sequence deduced from the intended vaccine peptide sequence and as well with upstream control sequence of LMP1/2 promoter sequence. The developed "*in silico* DNA vaccine" with intended specific expression in EBV latent infection lymphocytes may be verified with NPC cell line of EBV-latent infected B lymphocytes for immunogenic induction in order to demonstrate the potential ability in shifting cell-mediated immunity (CMI) pathway towards MHC-I Tc cell of CTL while away from MHC-II Th cell.

The BAff evaluation of predicted nona-peptide agretope structure towards docking HLA1 pit structure exploits MVD regression model with tabu clustering parameter in order to avoid sub-optima and local optima with GA method. In that, the HLA1 BAff of predicted activity with tabu clustering parameter may show good correlation with experiment activity and as well may show better reproducibility of computation result when compared with default rerank score of MVD regression model. The BAff computation time with nona-peptide structure on docking HLA1 pit structure is respectively 150 minutes or 40 minutes in average with default parameter or tabu clustering parameter in MVD regression model.

In addition to antigenecity priority on LMP1np epitope candidates in original amino acid sequence, the LMP1 agretope anchor plastics candidates based on bmPDA structure modeling may comply with inferred position-specific preference on A*02:01 pit binding nona-peptide at positions {2nd; 9th} with {L/M; I/L/V} towards improving MVD BAff vale while in disregard of tolerated {I/Q/V; A/M}. Notably, putative LMP1np structures of bmPDA prediction match with compatible biological experiment verified IEDB entries of A*02:01 epitopes ID 1377922~1377926 from Herpesvirus 4 Strain B95-8 with LMP1np-125(~133) published by Khanna R et al. in 1988. Despite of top ranking antigenecity scores with NetCTL and SYFPEITHI severs with LMP1np-125 {L; L} within preferred {L/M; I/L/V}

anchor group, the MDV BAff value of LMP1np-125 structural docking appears at the last ranking position within the indicated group in original amino acid sequence whereas the LMP1np-125 {L; V}.

With the priority LMP1 epitope candidates for HLA1 predicted with the antigenicity scores of NetCTL and/or SYFPEITHI servers, our bmPDA designed nona-peptide structures may further contribute on the need of structural evaluation in order to move on the diligence in practical feasibility towards application direction of mining Ama candidates of in vitro DNA vaccine for in vitro cell activation and Ace candidates of in vivo twin adhesive for in vivo subject therapy based on MDV BAff values. Despite of accords in good antigenicity scores level of NetCTL and SYFPEITHI, NetCTL score rankings within Ama group of {L/M; I/L/V} seem to be more consistent with MVD BAff value rankings of our bmPDA designed LMP1np:Ama structures while at noticeable inconsistency with SYFPEITHI score rankings. Moreover, high NetCTL antigenicity scores of top 8 candidates with the distribution range from 1.575 to 1.303) may fail to offer adequate {L/M; I/L/V} intra-group resolution as in contrast to the distribution range of low score antigenicity cases; whereas BAff value range from -24.347 to -10.253 of top 8 candidates may offer appropriate power for differentiation among LMP1np:Ama candidates as to be a supplemental indicator for analyzing HLA1-binding nona-peptides.

Instead of external delivery in vaccine peptide towards regular AMI induction, the intended vaccine peptide of HLA1 binder epitope and agretope for CMI induction shall in future take internal delivery strategy in the format of DNA vaccine which is constructed with coding sequence DNA insert deduced from intended HLA1 binder peptide sequence and as well with upstream control sequence of appropriate promoters. The intended CMI vaccine peptide thus in the format of in vitro DNA vaccine may serve good application for in vitro cell activation towards in vivo adoptive cell transfer with mixed lymphocyte reaction (MLR) plate separated with dialysis membrane from host cell and antigen presenting cell (APC), T cytotoxic (Tc) cell, and T helper (Th) cell. The in vitro DNA vaccine is hypothesized with practical feasibility while bypassing immune suppression within in vivo tumor microenvironment and while avoiding adverse clinical cytotoxicity on innocent bystander cells due to in vivo non-specific delivery of DNA vaccines.

Chapter 6 NPC-CMI Peptide Evaluating towards HLA1 and Agretope Complex Enhancements

6.1 Introduction

Nasopharyngeal carcinoma (NPC) at nasopharynx epithelium of squamous cell origin is common in south-eastern Asia countries including Taiwan, Indonesia, Singapore, Malaysia, and Vietnam in addition to Hong Kong and southern China.[2] Genetic susceptibility, environmental factors, and Epstein-Barr virus (EBV) infections comprises the important interplays towards NPC oncogenesis. [4] The EBV of herpesvirus family with 184-kB double-stranded DNA genome for encoding nearly 100 viral proteins often attacks the primary target of B lymphocytes and often results in lifetime latent infection.[7] Presence of EBV genome is demonstrated virtually in most NPC cells through oncogenesis process of EBV latent infections.

In background review, the EBV-NPC oncogenesis process may equip both proliferation advantage and immune evasion in order to overcome efficient anti-EBV immune clearance mechanisms of antibody-mediated immunity (AMI) with antibody-dependent cell-mediated cytotoxicity (ADCC) as well as cell-mediated immunity (CMI) with cytotoxic T lymphocyte (CTL)-initiative cytotoxic apoptosis during either latent and/or regular EBV infection phases. [7] Specifically, the basic proliferation advantage is likely from encoding EBV latent infection membrane protein 1 (LMP1) with growth factor receptor-like mutants and as well the critical immune evasion is likely from mutating EBV genome for poor immunogenicity responses at AMI-antigen epitopes and CMI-antigen epitopes/agretopes within LMP1/LMP2 and/or EBNA of EBV-encoded proteins.[8] Highly likely, the EBV-NPC immune evasion on ultimatum CMI-agretope mutant maybe the most crucial strategy for oncogenic negative selection against which the host immune system cannot counter-act efficiently despite that the other EBV-NPC immune evasions on AMI-epitopes and CMI-epitopes in oncogenic cells maybe eventually removed with affinity maturations of B cell receptor (BCR) and T cell receptor (TCR) through gene hyper mutations during the entire long-term process of EBV-NPC oncogenesis.[8]

The EBV-NPC immune evasion of CMI-agretope mutant maybe well exemplified with the interplay case of ethnic prevalence difference on class I human leukocyte antigen (HLA1) spectrum with additional LMP1 mutant assorts. The HLA A*0207 (common in Taiwan population) shows higher chances of EBV-NPC than the A*0201 (common in Caucasian population) even further with additional synergistic B*4601/B*14 and extended haplotype HLA A*3303- B*5801/2- DRB1*0301- DQB1*0201/2- DPB1*0401. [11,12] The EBV-NPC biopsies in Taiwan population select LMP1 variants of immune evasion NLMP1 which shares high amino acid sequence homology prominently with prototype B95.8-LMP1 and CAO-LMP1 in China population. Importantly, the NLMP1 over-expression in Balb/c class I major histocompatibility complex (MHC1)-context towards regressing the experimental murine EBV-NPC of NLMP1 expression may result from regaining strong CMI agretope presentation of NLMP1 in mice MHC1 context in disregard of the original selected immune evasion of NLMP1 in human HLA1 context likely with weak agretope binding presentation.[11,12,13] The restricted HLA1 spectrum of genetic susceptibility may indicate that the overlooked anchoring CMI-agretope of omega-shape nona-peptide is required for crucial docking onto HLA1-cleft in order for adequate CMI-epitope immunogenicity presentation towards effective CMI induction.

Immunoinformatics is with remarkably high practical potential in application aspect of epitope/agretope peptide binders with AMI-antibody and CMI-HLA1/2 towards screening the putative agretope complex enhancement (Ace) molecules with increased binding affinity (BAff) of NLMP1 agretope and HLA A*0207 in order likely to specifically improve EBV-NPC CMI with low non-specific adverse. Our structure-based immunoinformatic approach aims at EBV-LMP1 immunogenicity-related omega-shape nona-peptide design.[1] The bio-mimicry peptide design algorithm tool (bmPDA-tool) implements three sections including peptide building blocks database construction, peptide backbone modeling of building block candidates, and quality evaluation of predicted nona-peptide structures.[1]

Considerably, the restricted HLA spectrum may indicate the crucially overlooked agretopes of vaccine peptide for anchoring onto both ends of antigen pit in MHC-I with which seems to be required for adequate antigen presentation on peptide epitope towards good immunogenicity. The design strategy of MHC-I vaccine peptide thus seemingly demands both optimized agretopes and immunogenic epitope to which additional peptide segments for improved APC proteasome processing are attached at both flanking sides.

The immunotherapeutic regime against EBV-NPC for instance may conveniently exploit various aspects including AMI-ADCC with vaccine peptides, CMI-CTL with DNA vaccines, and microenvironment immune suppression with in vitro cell activation towards in vivo

adoptive cell transfer. The current challenges except vaccine peptides still install obstacles including weak HLA1-binding agretope in host cell or dendritic cell, and specific delivery of DNA vaccine to host cells without damaging innocent bystander cells.[15,16,17,18] Immunoinformatics is with remarkably high practical potential in feasible application of epitope/agretope binders onto AMI-BCR and CMI-HLA/TCR towards mining putative anchor modified agretope (Ama) and agretope complex enhancer (Ace) with reinforced binding affinity (BAff) of NLMP1 agretope and A*02:07 pit in order to likely improve NPC-CMI specifically while with low adverse cytotoxic effect due to non-specificity.

In this study, we implement bio-mimicry peptide design algorithm (bmPDA) comprising peptide database construction of building blocks, peptide backbone modeling of building block candidates, and quality evaluation on predicted nona-peptide structures. Our bmPDA of structure-based immunoinformatic approach aims at designing EBV immunogenicity-related omega-shape NLMP1 nona-peptide (NLMP1np) structures. We apply in-house bmPDA-tool towards applications of predicting A*02:07-binding EBV-NLMP1np structures in order that the verification on putative epitope and agretope quality may be accomplished with outsourcing tools of NetCTL server and Molegro Virtual Docker (MVD) software. The BAff with designed omega-shape NLMP1np and LMP1np structures on docking both HLA pits of A*02:07 {PDB: 3OXS} and A*02:01 {PDB: 1BD2} may be evaluated with MDV tool towards mining putative Ama and Ace candidates among which may be identified in modified-anchor assorts and FDA-approval drugs based on stable BAff of NLMP1np agretope and A*02:07 pit in order to specifically improve NPC CMI yet likely with low adverse effect due to non-specificity.

6.2 Method

The overall workflow is shown in **Figure 6-1**.

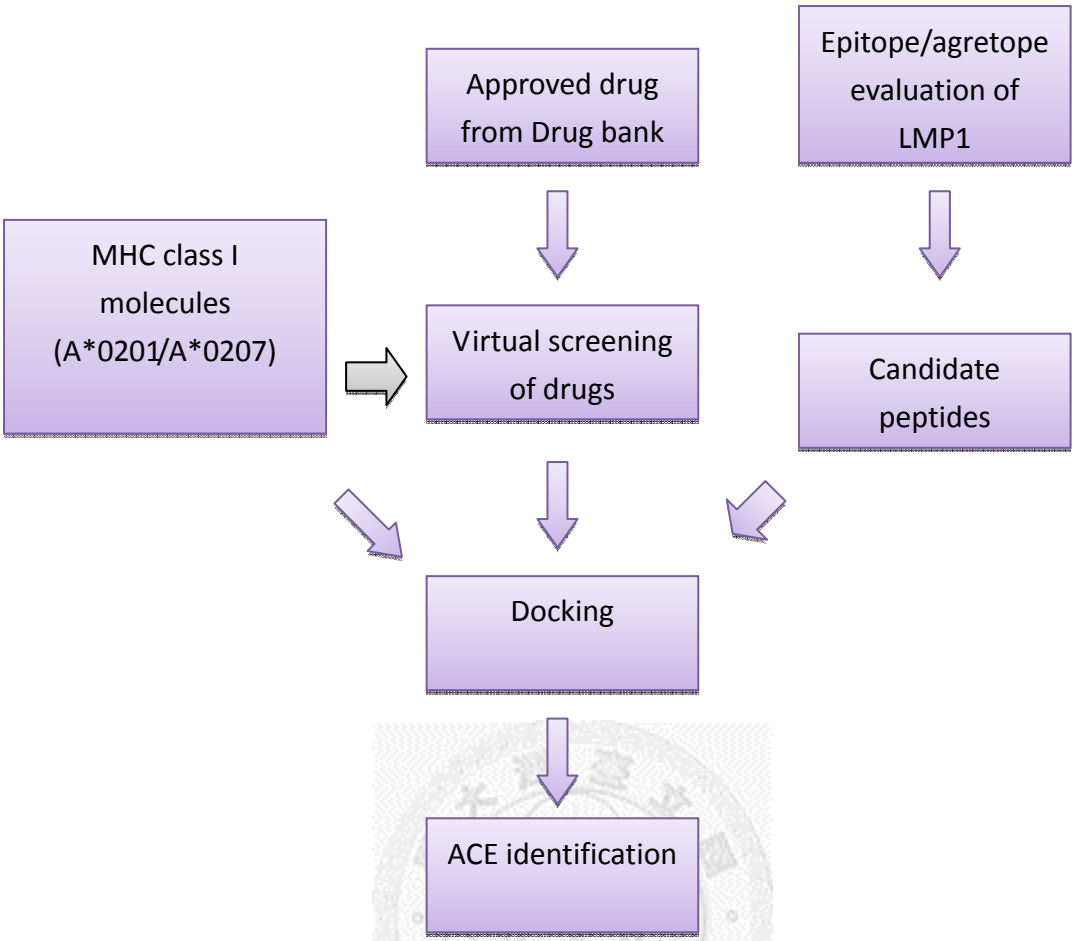


Fig. 6-1 Work flow of identification towards action/agretope complex enhancement (Ace) drugs.

6.2.1 Material preparation

LMP1 epitopes with good epitope score (NetCTL and SYFPEITHI antigenecity scores) and agretope score (MVD binding affinity) were collected as method in chapter 5. Structure files of HLA A*0201 and HLA A*0207 MHC class I molecules were collected from PDB. Structure files of FDA approved drugs were collected from DrugBank.

6.2.2 virtual screening for approved drugs

Virtual screening deals with large number of ligands against a receptor in reasonable time. The interactions between the drugs and MHC-I molecules are evaluated by Molegro Virtual Docker (MVD)[47]. Virtual screening was done for finding drugs with better binding affinity to HLA A*0201/HLA A*0207.

6.2.3 identification of action/agretope complex enhancement (Ace) drugs

After virtual screening, MHC-drug complex were docked with peptides for better epitope and agretope performance. Binding affinity was evaluated and compared with the original binding affinity without drugs. If the binding affinity increases, the drug will be viewed as action/agretope complex enhancement (Ace) molecule for improving binding affinity of predicted nona-peptide structures.

With hypothesized immune evasion at weak viral agretope anchors, our bmPDA tool selects putative NLMP1np structures with weak BAff value on docking ethnic A*02:07 and A*02:01 pit structures in order for evaluating BAff improvement on Ace candidates towards making in vivo twin adhesives. The twin adhesives strategy for Ace candidate is based on A*02:07 pit with better BAff on docking while at the presence of small chemical molecules from DrugBank of 1,435 FDA-approval drugs. With internal delivery in medication towards regular CMI induction, the intended Ace candidates for agretope anchors may in future take internal delivery strategy for CMI induction in the format of in vivo twin adhesive towards therapeutic drug for new indications.

The Ace candidates from FDA-approval drugs to in vivo twin adhesive thus in disease treatment format may serve good application for in vivo subject therapy towards likely in vivo NPC-CMI activation against latent infection phase NPC host cells with MLR in tumor microenvironment including in vivo adoptive cell transfer of in vitro activation cells and from local APC, Tc, Th, and myeloid derived suppressor cells (MDSC). The in vivo subject therapy towards likely NPC-CMI improvement with Ace candidates are speculated with practical feasibility while inducing specific cytotoxicity only upon NPC cells with in vivo specific presence of targeted weak NLMP1np anchors despite of in vivo non-specific distribution of Ace candidates.

6.3 Result and Dataset

Along with NetCTL antigenicity scores for HLA1 epitopes, our bmPDA-designed nona-peptide structures may move onto structural evaluation in feasible application practice of mining Ama candidates of in vitro DNA vaccine for in vitro cell activation and Ace candidates of in vivo twin adhesive for in vivo subject therapy based on MDV BAff values. For immune evasion likely towards tumorigenesis, the optimal NLMP1np structures of omega-shape conformation which are hypothesized to dock differently on A*02:07 and A*02:01 may have been somehow revealed with our bmPDA tool based on BAff value. The feasible applications of mining Ama candidates of in vitro DNA vaccine for in vitro cell

activation and Ace candidates of in vivo twin adhesive for in vivo subject therapy may have been somehow revealed again with preliminary results on the case of NLMP1np structures and A*02:07 pit structure.

6.3.1 Putative NLMP1 nona-peptide structures onto docking ethnic

A*02:01 versus A*02:07 pits

Better epitopes of LMP1 to HLA A*0201 is shown in **Table 6-1**. Virtual Screening Result for Drug on better epitopes of LMP1 to HLA A*0201 is shown in **Table 6-2**. Binding affinity evaluation between epitopes, HLA A*0201 MHC molecules, and adjuvant drug is shown in **Table 6-3**.

Table 6-1 The bmPDA-designed priority LMP1 nona-peptide agretopes to HLA A*0201 sorted with [agretope – epitope – agretope] docking scores in comparison among NetCTL and SYFPEITHI antigenicity scores.

position	sequence	Binding Affinity	NetCTL
125	YLLEMLWRL	-13.3300	1.5637
35	ALLFWLYIV	-13.4630	1.4636
167	LLVDLLWLL	-13.8151	1.4324
32	LLLALLFWL	-12.5441	1.3659
92	LLLIALWNL	-16.1708	1.3255
148	FLDLILLII	-18.5444	1.2840

Table 6-2 Virtual screening for Ace drug candidates on the bmPDA-designed priority LMP1 nona-peptide agretopes to HLA A*0201.

ID	DrugBank ID	Name	Binding Affinity
Drug_868	DB00868	Benzonatate	-848.723
Drug_770	DB00770	Alprostadil	-570.985
Drug_442	DB00442	Entecavir	-525.688
Drug_585	DB00585	Nizatidine	-509.538
Drug_927	DB00927	Famotidine	-502.321

Table 6-3 Binding affinity evaluation with Ace drugs towards bmPDA-designed priority LMP1 nona-peptide agretopes onto HLA A*0201.

LMP1	original Binding affinity	DB00868 Benzonatate	DB00770 Alprostadil	DB00442 Entecavir	DB00585 Nizatidine	DB00927 Famotidine
YLLEMLWRL	-13.3300	-8.6442	-7.91187	-25.7789	-12.0708	-14.1084
ALLFWLYIV	-13.4630	-18.4250	-18.1605	-52.6991	-26.3374	-17.1538
LLVDLLWLL	-13.8151	-20.1659	-3.61057	-7.4801	-18.7837	-9.9925
LLLALLFWL	-12.5441	-36.6455	-16.4207	-17.3280	-12.8342	-9.5650
LLLIALWNL	-16.1708	-6.8241	-11.333	-11.7121	-17.9643	-10.0029
FLDLILLII	-18.5444	-10.8383	-7.2318	-12.4783	-23.4822	-11.4594

Epitope Docking Result for LMP1 is shown in **Table 6-4**. Antigenecity was decreased in HLA A*0207 than HLA A*0201. Binding affinity was also decreased in HLA A*0207. Both epitope and agretope were decreased in HLA A*0207 than HLA A*0201 (**Table 6-4**). Binding affinity of NLMP1 also weakened in binding with HLA A*0207 (**Table 6-5**).

Table 6-4 The bmPDA-designed priority LMP1 nona-peptide agretopes towards HLA A*0201 and A*0207 sorted with [agretope – epitope – agretope] docking scores in comparison among NetCTL antigenicity score.

position	sequence	NetCTL score	HLA A*0201 Binding Affinity	HLA A*0207 Binding Affinity
125	YLLEMLWRL	1.5637	-13.3300	-14.4478
35	ALLFWLYIV	1.4636	-13.4630	-11.1488
167	LLVDLLWLL	1.4324	-13.8151	-12.3144
32	LLLALLFWL	1.3659	-12.5441	-15.2408
92	LLLIALWNL	1.3255	-16.1708	-12.5091
148	FLDLILLII	1.2840	-18.5444	-12.5585
86	LLLMITLLL	1.2779	-3.2896	-5.7820
112	FIFGCLLVL	1.2517	-13.6547	-22.2827
144	FLAFFLDLI	1.2339	-16.0035	-22.1579

Table 6-5 The bmPDA-designed priority NLMP1 nona-peptide agretopes towards HLA A*0201 and A*0207 sorted with [agretope – epitope – agretope] docking scores in comparison among NetCTL antigenicity score.

position	sequence	NetCTL	HLA A*0201 Binding Affinity	HLA A*0207 Binding Affinity
35	ALLFWLYIV	1.4606	-23.8611	-15.6310
166	LLVDLLWLL	1.4140	-16.7493	-5.8554
32	LLLALLFWL	1.3495	-15.0933	-18.9673
92	LLLIALWNL	1.3190	-14.6203	-9.7118
112	FIFGCLLVL	1.2954	-14.8420	-19.5422
86	LLLMITLLL	1.2798	-14.0656	-6.3505
158	YLQQNWWTL	1.2788	-26.0066	-29.4955
142	FILAFFLAI	1.2690	-19.0063	-17.4068
147	FLAIILLII	1.2620	-17.7329	-9.1896
61	MLIIIILII	1.2355	-14.9290	-19.4790

On immune evasion likely towards tumorigenesis, optimal NLMP1np structures {X66863} of omega-shape conformation may be somewhat revealed on docking unstably with A*02:07 {3OXS} pit of Asian and Taiwan population for evading CMI in contrast to A*02:01 {1BD2} pit of Caucasian population may be versified with our bmPDA tool based on BAff value. With exclusion on NLMP1np-125, the NLMP1np candidates of top ranking NetCTL antigenecity scores are similar to LMP1np case {V01555} in which both NLMP1np and LMP1np comply with inferred position-specific preference on HLA1-binding nona-peptide at positions {2nd; 9th} in {L/M; I/L/V} as of A*02:01 case in addition to an extra preference at position {3rd} in {D/P} as of A*02:07 case only.[38] The listed NLMP1np candidates with top ranking NetCTL antigenecity scores may all comply with

stated preference {L/M; I/L/V} except NLMP1np-112/142 with tolerated {I, L}. In general, the resulted BAff values of agretope complex with HLA1 {A*02:07, A*02:01} pit structures and NLMP1np agretope structures seem to suggest that N-LMP1np structures show greater loss of BAff value especially in A*02:07 docking case which may lose on CMI antigen presentation likely for immune evasion.

Based on available NetCTL score difference of A*02:07 on A*02:01, the rankings of NLMP1np structures are as of 1/035 (+0.011), 2/086 (-0.023), 3/112 (-0.024), 4/092 (-0.027), 5/166 (-0.056), 6/032 (-0.071), and 7/142 (-0.088) with position number of initial amino acid indicated. Further based on available BAff value difference of A*02:07 on A*02:01, the rankings of NLMP1np structures are as of 1/166 (+10.89), 2/147 (+8.54), 3/035 (+8.23), 4/086 (+7.72), 5/092 (+4.91), 6/142 (+1.60), 7/158 (-3.49), 8/032 (-3.87), 9/061 (-4.55), and 10/112 (-4.70). Despite of narrow NetCTL score gaps with losing antigenecity tendency among applied NLMP1np except NLMP1np-035 structures, ethnic BAff value differences of A*02:07 on A*02:01 in docking NLMP1np structures seem in general to suggest a greater tendency of increasing BAff values which may indicate loss of docking stability upon NLMP1np-166 (#2), 147 (#9), 035 (#1), 086 (#6), and 092 (#4) as in contrast to a moderate tendency of decreasing BAff values which may indicate gain of docking stability upon NLMP1np-158 (#7), 032 (#3), 061 (#10), and 112 (#5) with NetCTL rankings numbers attached.

The above results seem suggest that A*02:07 pit structure of Asian population may lose greater binding stability with NLMP1np structures especially from top ranking antigenecity candidates based on NetCTL scores. Particularly, NLMP1np-035/166 structures of top 1st/2nd NetCTL antigenecity score may lose the most binding stability with A*02:07 pit structure of Asian population which may likely take place in EBV latent infection host cells with LMP1 proliferative effect at the point of exploiting less efficient antigen presentation to Tc-TCR for immune evasions with survival advantages. The survival advantages disclosed possibly in EBV latent infection host cells towards NPC tumorigenesis of long-term selection process may zoom into negative selection track at certain stage on mutating agretope anchors in order for immune evasion onto less efficient CMI clearance while still with original epitope bulge of excellent antigenecity.

With this regard, possible remedy for reverting the indicated immune evasion upon ethnic difference of HLA1 pits and mutant agretopes of excellent epitope context may apparently require to assist ethnic pit and mutant agretope on increasing binding stability in that we propose efforts on mining Ama and Ace candidates for intended practical applications despite that we not yet put attention on the opposite case of ample

HLA1-binding blockers as with excellent agretopes yet with bad epitope. To intended practical applications, our bmPDA tool may specifically mine Ama candidates of putative NLMP1np structures with lower BAff value of better docking stability towards making in vitro DNA vaccine of intracellular expression in order for in vitro cell activation onto in vivo adoptive cell transfer while bypassing in vivo immune suppression within tumor micro-environment and while avoiding in vivo adverse cytotoxicity upon innocent bystanders due to in vivo non-specific delivery of DNA vaccines. Further, our bmPDA tool may mine Ace candidates among FDA-approval drugs with lower BAff value of better docking stability towards making in vivo twin adhesive between weak ethnic pit of A*02:07 structures and weak agretope anchor of NLMP1np structures in order for in vivo subject therapy while inducing specific cytotoxicity only upon NPC cells with in vivo specific presence of weak ethnic pit and weak agretope anchor despite of in vivo non-specific distribution of Ace drugs.

6.3.2 Mining anchor-modified agretope for NLMP1 nona-peptide on docking A*02:07 pit structure

The feasible application of mining Ama candidates for in vitro DNA vaccine of cell activation is pursued on the case of NLMP1np and A*02:07 while with assorted anchors which may comply with inferred position-specific preference at respective position {2nd; 9th} with {L/M; I/L/V} for A*02:07 and A*02:01 in addition to position {3rd} with {D/P} for only A*02:07 towards improving MVD BAff value. On docking A*02:07 pit structure, NLMP1np-035 ({L; V}; {M; V}) respectively without or with anchor modification gives MVD BAff values (-15.6310; -35.9599) in which notable difference value of 20.33 may represent the improvement level. The visualization of agretope complex of NLMP1np-035 ({L; V}; {M; V}) with A*02:07 pit structure seems to shift from slant docking towards balanced docking at good improvement of complex stability. On docking A*02:01 pit structure, same LMP1np-035 ({L; V}; {M, V}) gives respective MVD BAff values (-20.5003; -30.8523) while merely moderate 10.35 of derived difference value.

To this point, the Ama candidates of putative NLMP1np structures may subsequently be verified in the same A*02:07 genetic background with in vitro mixed lymphocyte reaction (MLR) comprising host cells transfected with constructs of NLMP1np candidates, Tc cells, and Th cells for analyzing CTL activity according to the practical application of making in vitro DNA vaccine of intracellular expression in order for in vitro cell activation onto in vivo adoptive cell transfer while bypassing in vivo immune suppression within tumor

micro-environment and while avoiding in vivo adverse cytotoxicity upon innocent bystanders due to in vivo non-specific delivery of DNA vaccines.

6.3.3 Mining agretope complex enhancer for NLMP1 nona-peptide on docking A*02:07 pit structure

HLA alleles are important in the pathogenesis of virally induced tumors. HLA A*0207 is more prevalent in Taiwan. Consistent association was found between HLA-A*0207 (common among Asian but not among Caucasians) and NPC but not between HLA-A*0201 (most common HLA-A2 allele in Caucasians) and NPC [50]. NLMP1 is an EBV strain prominent in Taiwanese population.[51] Antigenecity prediction by NetCTLpan for NLMP1 in HLA A*0201 and HLA A*0207 is shown in **Figure 6-2**. NLMP1 Epitopes with good antigenecity were selected for candidate epitopes (**Table 6-6**).

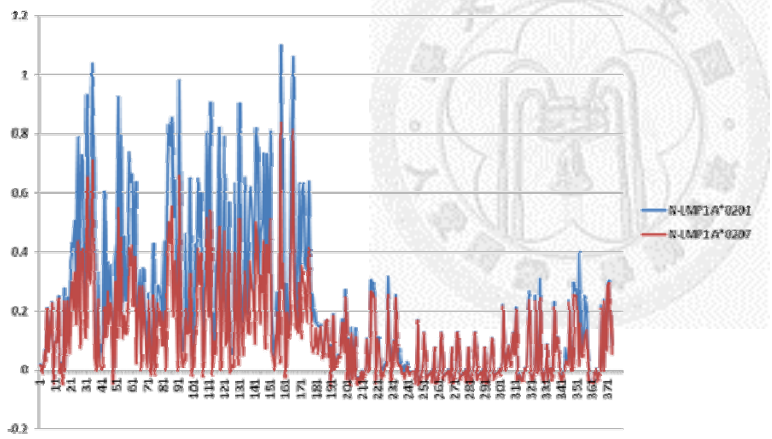


Fig. 6-2 Overall NLMP1 epitope antigenecity of bmPDA predicted peptide structures analyzed by NetCTLpan antigenecity score server towards both HLA A*0201 and A*0207.

Table 6-6 Top ranking agretopes of NLMP1 towards HLA A*0207 selected by MVD binding affinity.

position	sequence
35	ALLFWLYIV
166	LLVDLLWLL
32	LLLALLFWL
92	LLLIALWNL
112	FIFGCLLVL
86	LLLMITLLL
147	FLAIILLII

Virtual Screening Result of FDA approved drug for NLMP1 on HLA A*0207 is shown in **Table 6-7**. Docking result summary is shown in **Table 6-8**. Entecavir, Nizatidine,

Famotidine performed better, so they were viewed as Ace drugs. The feasible application of mining Ace candidates for in vivo subject therapy is pursued on the case of NLMP1 and A*02:07 while with available DrugBank (www.drugbank.ca) chemicals of 1,435 FDA-approval drugs and while upon proposed new indication of which aims at improving weakened binding of variant NLMP1 agretope and HLA1 pit for reverting in vivo immune evasion and at lifting CMI reactivity towards tumor cell removal gain which maybe likely to take place at in vivo NPC microenvironment. With the exclusion of NLMP1np-061, 142, and 158, the rest NLMP1np structures are evaluated including NLMP1np-166 (#2), 147 (#9), 035 (#1), 086 (#6), and 092 (#4) of greater loss on docking stability with top NetCTL rankings attached and NLMP1np-158 (#7), 032 (#3), and 112 (#5) of moderate gain on docking stability.

Among 1,435 FDA-approval drug structures from DrugBank, the harvest of mining Ace candidates may be exemplified with DB585 Nizatidine, DB868 Benzonatate, DB442 Entecavir, DB927 Famotidine, and DB770 Alprostadil according to docking stability gain of less BAff value while docking NLMP1np structures onto A*02:07 pit structure. Moreover, Ace candidates are in general class of Histamine-2 receptor antagonist (H2-blocker) with literature reports showing CMI enhancement effects for supplementing cancer therapy additionally exemplified with an especially interesting Ace candidate {Entecavir} for clinical indication of treating HBV chronic infection that has concluded clinical trial in Taiwan.

Table 6-7 Virtual screening for Ace drug candidates on the bmPDA-designed priority NLMP1 nona-peptide agretopes to HLA A*0207.

ID	DrugBank ID	Name	Binding Affinity
Drug_868	**DB00868**	**Benzonatate**	**-858.24**
Drug_770	**DB00770**	**Alprostadil**	**-535.50**
Drug_927	**DB00927**	**Famotidine**	**-507.71**
Drug_442	**DB00442**	**Entecavir**	**-505.93**
Drug_585	**DB00585**	**Nizatidine**	**-490.57**

Table 6-8 The bmPDA-designed priority NLMP1 nona-peptide agretopes towards HLA A*0207 sorted with [agretope – epitope – agretope] docking scores in comparison among NetCTL antigenicity score.

NLMP1	original Binding affinity	DB00868 Benzonatate	DB00770 Alprostadil	DB00927 Famotidine	DB00442 Entecavir	DB00585 Nizatidine
ALLFWLYIV	**-15.6310**	**-17.4564**	**-20.3148**	**-15.2905**	**-32.6494**	**-17.8726**
LLVDLLWLL	**-5.8553**	**-13.8805**	**-10.5714**	**-10.6011**	**-17.3471**	**-9.4336**
LLLALLFWL	**-18.9673**	**-15.3687**	**-9.9673**	**-8.3331**	**-22.0679**	**-12.6647**
LLLIALWNL	**-9.7116**	**-36.4624**	**-7.3806**	**-17.4920**	**-9.2982**	**-7.7141**
FIFGCLLVL	**-19.5422**	**-18.3895**	**-23.8061**	**-21.4551**	**-36.7102**	**-36.2241**
LLLMITLLL	**-6.3505**	**-8.2676**	**-9.5218**	**-15.7051**	**-7.1375**	**-8.7732**
FLAIILLII	**-6.8421**	**-15.6089**	**-6.6913**	**-13.5632**	**-10.6795**	**-20.8778**

To this point, Ace candidates of FDA-approval drugs may subsequently be verified in genetic background of ethnic A*02:07 on A*02:01 with in vitro mixed lymphocyte reaction (MLR) comprising host cells such as tissue cells and/or dendritic cells transfected with in vitro DNA vaccine constructs of NLMP1np candidates, Tc cells, and Th cells in order for analyzing CTL activity without or with in vitro Ace candidate treatment as of simulating practical application of making in vivo twin adhesive between A*02:07 pit structure and weak agretope anchor of NLMP1np structures in order for in vivo subject therapy while inducing specific cytotoxicity only upon NPC cells with in vivo specific presence of intracellular responsive NLMP1np structures despite of in vivo non-specific distribution of Ace drugs within host cells of none NLMP1 existence.

On docking A*02:07 pit structure, NLMP1np-035 {L; V} without or with Ace drug DB442 Entecavir gives MVD BAff values of (-15.6310 without Ace; -32.6494 with Ace) in which notable value difference of 17.02 may represent improvement level. Visualizing agretope complex of NLMP1np-035 {L; V} structure with A*02:07 pit structure seems to show slant docking towards extending anchor at good improvement of complex stability upon trial conditions of without or with Ace drug treatment.

Despite of potential structural difference between parallel Ama and Ace applications while with good improvement level of complex stability, the epitope-bulge structures between Ama and Ace cases respectively of NLMP1np-035 {M; V} on A*02:07 pit structure of balanced docking and of NLMP1np-035 {L; V} on A*02:07 pit structure of slant docking at anchor extension may indeed exist minor RMSD structure difference based on SuperPose server. Anyhow, minor difference at epitope-bulge of TCR binding may not cause major trouble upon practical applications due to that versatile TCR may exist adequate cross reactivity and as well flexible TCR may generate adequate variant conformations for improving specific binding with different epitope-bulges as of TCR affinity maturation via genetic processes of gene rearrangements and so forth.

6.4 Conclusion and Future Works

HLA alleles are important in the pathogenesis of virally induced tumors. HLA A*0207 is more prevalent in Taiwan. Consistent association was found between HLA-A*0207 (common among Chinese but not among Caucasians) and NPC but not between HLA-A*0201 (most common HLA-A2 allele in Caucasians) and NPC. Within the HLA A2 group, only HLA-A*0207 alleles (a genotype common among individuals of Chinese descent but rare

among Caucasians) is related to NPC risk, although the HLA-A*0201 allele (a genotype common among Caucasians) is not associated with NPC. T-cell epitopes of LMP1 and LMP2 are efficiently presented by HLA-A*0201. Data suggest that the HLA-A*0207 allele is less efficient than the HLA-A*0201 allele at inducing cytotoxic T-lymphocyte responses.[50]

Our research result is consistent with previous studies. In our study, both epitope score and agretope score were decreased in HLA A*0207 group than in HLA A*0201 group. Based on the observation, we can postulate immune reaction of EBV should be weaker in HLA A*0207 group. Latent infection may be more prevalent and lead to NPC eventually. That could partly explain why NPC is more prevalent in Taiwan compared with western country.

Peptide-based vaccines are safe, stable, and easy to produce in large scale. Specific immune responses can also be monitored easily and correlated with clinical responses. However, Peptide-based vaccines need to identify the immunogenic epitope of the tumor-associated antigen. Most peptide-based vaccines focused on antigenic peptides which bind the HLA-A2 molecule due to its high frequency of expression in up to 50% of the Caucasian. Compared to bacterial or viral vaccine vectors, immunogenicity of peptide vaccines are relatively poor. Researchers in this area had focused on finding adjuvant immune-enhancing agents such as chemokines, cytokines, and co-stimulatory molecules to enhance the potency of the peptide vaccine. For maximizing the immunological responses elicited from peptide-based vaccines, it is important to identify the appropriate adjuvants and route of administration[11]. In our study, we demonstrated a method for searching preexisting FDA approved drug as adjuvant agent. In adoptive immunotherapy for autologous CTLs transfer, our method may be applied in similar way.

In our study, Entecavir, Nizatidine, Famotidine were viewed as ACE (agretope complex enhancement) drugs. Entecavir is an antiviral drug used in the treatment of hepatitis B infection. Our result implied that it may have some immune regulatory effect besides antiviral function. Surprisingly, Nizatidine and Famotidine enhance agretope complex in our study. These histamine-2 antagonist are used for treatment of peptic ulcer for a long time. However, there are some studies mentioned about their immune-related function.[52-57] Other than effect on histamine receptor, these drugs may have role in cell-mediated immunity. It is very interesting for further investigation.

Chapter 7 Summary and Future Works

In this thesis, structural bioinformatics are applied to solve problems of EBV-related immunotherapy. Our system can be applied on both two important aspect of immune reaction: AMI-Epitopes and CMI-Agretopes. We applied structure-based immunoinformatics methods of in-house bmPDA tool for practical applications of immunotherapy that specifically covers the case of designing NLMP1np structure and docking HLA1 A*02:07 pit structure in order for mining intended Ama and Ace candidates. Our preliminary results of bmPDA tool that designs putative structures unknown for nona-peptides and measures binding stability of action complex may likely be a supplementing extension for NetCTL server of antigenecity score.

With the intention to explore the possibility of differential agretope binding stability as a pathway of immune evasion among ethnic populations, our results may at least shed some light on the predisposing mechanism towards NPC formation due to ethnic difference of A*02:07 versus A*02:01 in which NPC-related NLMP1np structures may lose greater agretope binding stability especially among top antigenecity NLMP1np structures of high NetCTL scores while still as of defective binders of HLA1 A*02:07 pit structure. Without even doing detailed analysis on the defective HLA1 binders of octa-peptide and deca-peptide structures and on the effective HLA1 blockers of oligo-peptide structures with strong agretope binding stability yet with rather inadequate epitope antigenecity, we still feel educational with the stated shedding of light in which we feel the need of Ama and Ace candidates to be the remedy for reverting inefficient agretope binding while with appropriate formats of practical application in future.

The NLMP1 agretope-oriented immune evasion for negative selection on HLA1 A*02:07 host cell with survival advantage for advancing oncogenesis may simply due to the innate agretope inefficiency and as well the mutant agretope inefficiency on docking A*02:07 pit structure for CMI induction after latent infection stage. On the opposite case, the NLMP1 epitope-oriented immune evasion may simply be successful merely in a short term manner due to the innate adaptive variation mechanism of BCR and TCR genes within proliferative lymphocytes towards selecting affinity maturation for efficient immune removal while comparing to the less efficiency on innate adaptive variation of HLA1 genes.

Specifically, it is likely that the ethnic difference between A*02:07 and A*02:01 pit structures on docking NLMP1np structures does render the host cell of A*02:07 genetic background with survival advantage for negative selection from efficient CMI removal as of immune evasion especially until latent infection stage with merely NLMP1 and EBNA

expression and whereas without survival advantage for negative selection from efficient AMI removal while during regular infection stage with many viral envelop proteins expressed on host cell membrane. Subsequently, it is likely that the survival advantage for negative selection favors NLMP1np agretope mutants of further less efficient binding onto A*02:07 pit structure in which may render the host cell even greater ability of immune evasion from efficient CMI removal.

Our proposed remedy of Ama and Ace candidates for abolishing the suggested immune evasion on reverting the NLMP1 agretope inefficiency on docking A*02:07 pit structure may likely be the practical applications of in vitro DNA vaccine and in vivo twin adhesive. Detailed computational study may need to be done with preliminary Ace drugs as of true positive cases on agretope complexes in order for verifying the complete exclusion on being intermediate complex enhancer (Ice) as of false negative cases along the antigen presentation pathway of class I MHC pertaining to intermediate complexes exemplified with calnexin and Tapasin and so forth. Meanwhile, the Ama and Ace candidates either in respective or cocktail format may subsequently be verified in same A*02:07 genetic background with in vitro mixed lymphocyte reaction (MLR) of Th cells, Tc cells, and host cells transfected with in vitro DNA vaccine of NLMP1np Ama candidates in order for analyzing CTL activity without or with treatment of Ace candidates.

Interestingly, the Ace candidates from available DrugBank chemicals of 1,435 FDA-approved drugs towards potential clinical indications have come out to be in general classes of Histamine-2 receptor antagonist (H2-blocker) and anti-HBV drug Entecavir with literature reports showing CMI enhancement effects for cancer therapy supplements. Notably, the clinical indication of Entecavir for improving histological inflammation on treating HBV chronic infection may be with subtle supportive linking with our initiative exploration in recent progress which is to investigate whether or not that Entecavir may enhance HBeAg agretope binding to HLA1 for improved HBeAg-CMI efficiency towards reducing chronic infection and as well may in turn assist HLA1 in abolishing immune evasion of HBeAg agretope towards reducing HBV HCC oncogenesis potential as of the common remedy similarly proposed on EBV NPC case.

We summarize the results of previous chapters in the following sections.

7.1 Summary

In this dissertation, we try to apply structural bioinformatics in NPC immunotherapy. NPC is strongly related with EBV, so we mainly focus on EBV-related immune responses.

In the chapter 1 of this dissertation, we reviewed the carcinogenesis of NPC and EBV-related immune response, and problems of immunotherapy. We try to solve them by structural immunoinformatics.

In the chapter 2 of this dissertation, we developed a block feature definition system for describing protein structures. Followed the concept of structure alphabet, we extract the structural information from protein structures and identified 22 states of the structural alphabet that represent pattern profiles of the backbone fragments based on our block feature definition. Basic structural immunoinformatic databases were then constructed, such as 5mers fragment library and TPSS. Several tools are also developed for peptide block manipulation.

In chapter 3, we proposed a method named bmPDA for peptide block assembly and developed an algorithm for peptide block modeling by genetic algorithm. Epitopes with known structure of HLA A*0201 were collected from IEDB(Immune Epitope Database) and PDB databases. Peptide structures predicted by our bmPDA method were compared with the actual structures from PDB. Alignment was done by Superpose and RMSD score was calculated. The result is good compare to the result of other short peptide structure prediction server. After prediction of peptide structure from sequence was done, we extract the structural information from predicted structures as structure alphabet and build a QSAR model for epitope prediction by using SVR. ROC analysis revealed our QSAR-SVR model is comparable with the best sequence-based epitope prediction server NetCTL.

In chapter 4, we develop a method called “bio-mimicry peptide design“. Follow the concept of inverse folding search, we develop an approach to find possible sequence combinations mimicking target structure. Evaluation on the structure similarity with target peptide, physiochemical property, and structure stability of predicted solutions were done for finding better potential candidates. We exemplify our method on two targets: vasopressin and a monoclonal antibody TMB-355.

In chapter 5, we applied the above structural immunoinformatic approaches for nasopharyngeal carcinoma (NPC) vaccine design. NPC is a common malignancy in southern China, Hong Kong, and south-east Asia countries including Taiwan, Singapore, Malaysia, Indonesia, and Vietnam. It is strongly associated with Epstein-Barr virus (EBV). Immunotherapy for NPC is currently focusing on the tumor-associated antigens called LMP1 and LMP2. However, poor antigenecity of LMP1/LMP2 limited the efficacy of EBV vaccine in NPC immunotherapy. We predicted the structure of every possible epitopes of LMP1/LMP2 from sequence, docked them with MHC-I HLA A*0201 molecule, and compare the docking result with predicted antigenecity of LMP1/LMP2 from NetCTL and

SYFPEITHI prediction servers. Epitopes with better performance of antigenecity were collected as candidates for polyepitope regimen. According to the preference observation on known epitopes, residues on specific position of the candidate epitopes were modified to become epitopes with even better antigenecity. Agretope performance was evaluated by binding affinity prediction from docking with MHC receptor. We collected epitopes with better performance on epitope and agretope to be candidates of polyepitope regimen on NPC immunotherapy.

In chapter 6, we collected approved drugs from DrugBank. Virtual screening was done by docking with MHC receptor. Drugs with better binding affinity with MHC receptor were collected as possible candidate for adjuvant immunotherapy. Epitopes with better performance of antigenecity were collected by the same procedure in chapter 5. Epitope structure prediction was done by modeling method in chapter 3. MHC receptor and candidate drugs were docked with candidate epitopes. Drugs which could enhance the binding affinity between epitope and MHC receptor were identified. We suggest drugs with ACE (action site enhancement) to be adjuvant immunotherapy for NPC.

7.2 Future Works

Our peptide modeling method is based on genetic algorithm and bmPDA block assembly method. The performance is good locally but not so good in large molecules. Application on the structure prediction of epitopes is perfect, because they are consisted of about 8-12 amino acids. But if we want apply our method on structure prediction of large proteins, the impact of protein folding still need to be considered.

Despite the good accuracy of our QSAR-SVR epitope prediction method, the time-consuming peptide modeling process was the rate-determine step in our method. The block picking process during genetic algorithm may be speedup by block clustering.

Docking is a time-consuming process. Compare to other docking software, the Molegro Virtual Docker is more accurate and fast. However, if there is a faster and accurate docking software, the process can still be speed-up.

There are several directions for future research:

1. More peptide Segment sizes: Our epitope modeling method is limited to predict 9mers peptides till now. It can be applied to prediction on arbitrary length of protein. However, accuracy is still our concern.

2. Structure clustering on FluV neutralization epitopes and predictive immune evasions on FluV neutralization epitope
3. Extend the viral oncogenesis model of HLA1 immune evasion at agretope on miscellaneous viruses such as HBeAg

BIBLIOGRAPHY

1. Yang X, Yu X: **An introduction to epitope prediction methods and software**. *Reviews in medical virology* 2009, **19**(2):77-96.

2. Wei WI, Sham JST: **Nasopharyngeal carcinoma**. *The Lancet* 2005, **365**(9476):2041-2054.

3. Guigay J: **Advances in nasopharyngeal carcinoma**. *Current opinion in oncology* 2008, **20**(3):264-269.

4. Thompson MP, Kurzrock R: **Epstein-Barr virus and cancer**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2004, **10**(3):803-821.

5. Jenson HB: **Epstein-Barr virus**. *Pediatrics in review / American Academy of Pediatrics* 2011, **32**(9):375-383; quiz 384.

6. Masmoudi A, Toumi N, Khanfir A, Kallel-Slimi L, Daoud J, Karray H, Frikha M: **Epstein-Barr virus-targeted immunotherapy for nasopharyngeal carcinoma**. *Cancer treatment reviews* 2007, **33**(6):499-505.

7. Khanna R, Burrows SR, Silins SL, Moss DJ, Poulsen LM, Burrows JM: **Cytotoxic T-lymphocyte clones specific for an immunodominant epitope display discerning antagonistic response to naturally occurring Epstein-Barr virus variants**. *Journal of virology* 1996, **70**(10):7306-7311.

8. Chou J, Lin YC, Kim J, You L, Xu Z, He B, Jablons DM: **Nasopharyngeal carcinoma--review of the molecular mechanisms of tumorigenesis**. *Head & neck* 2008, **30**(7):946-963.

9. Duraiswamy J, Burrows JM, Bharadwaj M, Burrows SR, Cooper L, Pimtanothai N, Khanna R: **Ex Vivo Analysis of T-Cell Responses to Epstein-Barr Virus-Encoded Oncogene Latent Membrane Protein 1 Reveals Highly Conserved Epitope Sequences in Virus Isolates from Diverse Geographic Regions**. *Journal of virology* 2003, **77**(13):7401-7410.

10. Duraiswamy J, Sherritt M, Thomson S, Tellam J, Cooper L, Connolly G, Bharadwaj M, Khanna R: **Therapeutic LMP1 polyepitope vaccine for EBV-associated Hodgkin disease and nasopharyngeal carcinoma**. *Blood* 2003, **101**(8):3150-3156.

11. Wu AA, Niparko KJ, Pai SI: **Immunotherapy for head and neck cancer**. *Journal of biomedical science* 2008, **15**(3):275-289.

12. Ponomarenko JV, Bourne PE: **Antibody-protein interactions: benchmark datasets and prediction tools evaluation**. *BMC structural biology* 2007, **7**:64.

13. Yano A, Onozuka A, Asahi-Ozaki Y, Imai S, Hanada N, Miwa Y, Nisizawa T: **An ingenious design for peptide vaccines**. *Vaccine* 2005, **23**(17-18):2322-2326.

14. Yang JM, Tung CH: **Protein structure database search and evolutionary classification**. *Nucleic acids research* 2006, **34**(13):3646-3659.

15. Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Cadet F, Bornot A, Tyagi M, Valadie H *et al*: **A short survey on protein blocks**. *Biophysical reviews* 2010, **2**(3):137-147.

16. Pandini A, Fornili A, Kleinjung J: **Structural alphabets derived from attractors in conformational space**. *BMC bioinformatics* 2010, **11**:97.

17. Daga PR, Patel RY, Doerksen RJ: **Template-based protein modeling: recent methodological advances**. *Current topics in medicinal chemistry* 2010, **10**(1):84-94.

18. Kolodny R, Levitt M: **Protein decoy assembly using short fragments under geometric constraints**. *Biopolymers* 2003, **68**(3):278-285.

19. de Brevern AG, Etchebest C, Hazout S: **Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks**. *Proteins* 2000, **41**(3):271-287.

20. Fourrier L, Benros C, de Brevern AG: **Use of a structural alphabet for analysis of short loops connecting repetitive structures**. *BMC bioinformatics* 2004, **5**:58.

21. Etchebest C, Benros C, Hazout S, de Brevern AG: **A structural alphabet for local protein structures: improved prediction methods**. *Proteins* 2005, **59**(4):810-827.

22. Friedberg I, Harder T, Kolodny R, Sitbon E, Li Z, Godzik A: **Using an alignment of fragment strings for comparing protein structures**. *Bioinformatics* 2007, **23**(2):e219-224.

23. Schenk G, Margraf T, Torda AE: **Protein sequence and structure alignments within one framework**. *Algorithms for molecular biology : AMB* 2008, **3**:4.

24. Guyon F, Camproux AC, Hochez J, Tuffery P: **SA-Search: a web tool for protein structure mining based on a Structural Alphabet**. *Nucleic acids research* 2004, **32**(Web Server issue):W545-548.

25. Tung CH, Huang JW, Yang JM: **Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for rapid search of protein structure database**. *Genome biology* 2007, **8**(3):R31.

26. Le Q, Pollastri G, Koehl P: **Structural alphabets for protein structure classification: a comparison study**. *Journal of molecular biology* 2009, **387**(2):431-450.

27. Tung CH, Yang JM: **fastSCOP: a fast web server for recognizing protein structural domains and SCOP superfamilies**. *Nucleic acids research* 2007, **35**(Web Server issue):W438-443.

28. Tyagi M, de Brevern AG, Srinivasan N, Offmann B: **Protein structure mining using a structural alphabet**. *Proteins* 2008, **71**(2):920-937.

29. Pandini A, Bonati L, Fraternali F, Kleinjung J: **MinSet: a general approach to derive maximally representative database subsets by using fragment dictionaries and its application to the SCOP database**. *Bioinformatics* 2007, **23**(4):515-516.

30. Maupetit J, Gautier R, Tuffery P: **SABBAC: online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace**. *Nucleic acids research* 2006, **34**(Web Server issue):W147-151.

31. Deschavanne P, Tuffery P: **Enhanced protein fold recognition using a structural alphabet**. *Proteins* 2009, **76**(1):129-137.

32. Tuffery P, Derreumaux P: **Dependency between consecutive local conformations helps assemble protein structures from secondary structures using Go potential and greedy algorithm**. *Proteins* 2005, **61**(4):732-740.

33. Maupetit J, Derreumaux P, Tuffery P: **PEP-FOLD: an online resource for de novo peptide structure prediction**. *Nucleic acids research* 2009, **37**(Web Server issue):W498-503.

34. Maupetit J, Derreumaux P, Tuffery P: **A fast method for large-scale de novo peptide and miniprotein structure prediction**. *Journal of computational chemistry* 2010, **31**(4):726-738.

35. MacDonald JT, Maksimiak K, Sadowski MI, Taylor WR: **De novo backbone scaffolds for protein design**. *Proteins* 2010, **78**(5):1311-1325.

36. Cozzetto D, Tramontano A: **Advances and pitfalls in protein structure prediction**. *Current protein & peptide science* 2008, **9**(6):567-577.

37. Karney CF: **Quaternions in molecular modeling**. *Journal of molecular graphics & modelling* 2007, **25**(5):595-604.

38. Zhu Y, Tan Y, Hua Y, Wang M, Zhang G, Zhang J: **Feature selection and performance evaluation of support vector machine (SVM)-based classifier for differentiating benign and malignant pulmonary nodules by computed tomography**. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology* 2010, **23**(1):51-65.

39. Korber B, LaBute M, Yusim K: **Immunoinformatics comes of age**. *PLoS Comput Biol* 2006, **2**(6):e71.

40. Chang C-CaL, Chih-Jen: **LIBSVM: A library for support vector machines**. *ACM Transactions on Intelligent Systems and Technology* 2011, **2**(3):27:21--27:27.

41. Maiti R, Van Domselaar GH, Zhang H, Wishart DS: **SuperPose: a simple server for sophisticated structural superposition**. *Nucleic acids research* 2004, **32**(Web Server issue):W590-594.

42. Vita R, Zarebski L, Greenbaum JA, Emami H, Hoof I, Salimi N, Damle R, Sette A, Peters B: **The immune epitope database 2.0**. *Nucleic acids research* 2010, **38**(Database issue):D854-862.

43. Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M: **Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction**. *BMC bioinformatics* 2007, **8**:424.

44. Le TT, Drane D, Malliaros J, Cox JC, Rothel L, Pearse M, Woodberry T, Gardner J, Suhrbier A: **Cytotoxic T cell polyepitope vaccines delivered by ISCOMs**. *Vaccine* 2001, **19**(32):4669-4675.

45. Duraiswamy J, Bharadwaj M, Tellam J, Connolly G, Cooper L, Moss D, Thomson S, Yotnda P, Khanna R: **Induction of therapeutic T-cell responses to subdominant tumor-associated viral oncogene after immunization with replication-incompetent polyepitope adenovirus vaccine**. *Cancer research* 2004, **64**(4):1483-1489.

46. Larsen MV, Lundegaard C, Lamberth K, Buus S, Brunak S, Lund O, Nielsen M: **An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions**. *European journal of immunology* 2005, **35**(8):2295-2303.

47. Thomsen R, Christensen MH: **MolDock: a new technique for high-accuracy molecular docking**. *J Med Chem* 2006, **49**(11):3315-3321.

48. Sapre N, Gupta S, Sapre N: **Assessing ligand efficiencies using template-based molecular docking and Tabu-clustering on tetrahydroimidazo-[4,5,1-jk][1,4]-benzodiazepin-2(1H)-one and-thione (TIBO) derivatives as HIV-1RT inhibitors**. *J Chem Sci* 2008, **120**(4):395-404.

49. Bui HH, Schiewe AJ, von Grafenstein H, Haworth IS: **Structural prediction of peptides binding to MHC class I molecules**. *Proteins* 2006, **63**(1):43-52.

50. Hildesheim A, Apple RJ, Chen CJ, Wang SS, Cheng YJ, Klitz W, Mack SJ, Chen IH, Hsu MM, Yang CS *et al*: **Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan**. *Journal of the National Cancer Institute* 2002, **94**(23):1780-1789.

51. Hsieh YH, Liu HP, Lo SJ, Chang YS: **Tumor regression by expression of high physiological levels of EBV latent membrane protein 1**. *Cancer biology & therapy* 2006, **5**(3):310-317.

52. Nielsen HJ: **Histamine-2 receptor antagonists as immunomodulators: new therapeutic views?** *Annals of medicine* 1996, **28**(2):107-113.

53. Blaya B, Nicolau-Galmes F, Jangi SM, Ortega-Martinez I, Alonso-Tejerina E, Burgos-Bretones J, Perez-Yarza G, Asumendi A, Boyano MD: **Histamine and histamine receptor antagonists in cancer biology**. *Inflammation & allergy drug targets* 2010, **9**(3):146-157.

54. Naredi P: **Histamine as an adjunct to immunotherapy**. *Seminars in oncology* 2002, **29**(3 Suppl 7):31-34.

55. Smith T: **Histamine type 2-receptor antagonists and cancer immunotherapy**. *Comprehensive therapy* 1990, **16**(1):8-13.

56. Hellstrand K, Brune M, Naredi P, Mellqvist UH, Hansson M, Gehlsen KR, Hermodsson S: **Histamine: a novel approach to cancer immunotherapy**. *Cancer investigation* 2000, **18**(4):347-355.

57. Moharana AK, Bhattacharya SK, Mediratta PK, Sharma KK: **Possible role of histamine receptors in the central regulation of immune responses**. *Indian journal of physiology and pharmacology* 2000, **44**(2):153-160.

Appendix A

List of Publications

Journal Papers:

范振杰;林堂烈 (2003): "以懸壅垂顎皮瓣治療打鼾 Uvulopalatal Flap for Snoring" 臺灣耳鼻喉頭頸外科雜誌 38(2): 56~60 (2003, 03-04).

Hsueh-Ting Chu, William W. L. Hsiao, **Chen-Chieh Fan**, Chaur-Chin Chen and Cheng-Yan Kao (2012): "SeqEntropy: quantitative evaluation of sequence repeats for short read genome sequencing" *(submitted)*

Conference Papers:

Yu-Ju Chuang, Hsueh-Ting Chu, **Chen-Chieh Fan** and Kao Cheng-Yan (2009): "Prediction For Multi-epitope HCA661 Liver Cancer Vaccines". *Asia Pacific Bioinformatics Conference* 2009.

Chen-Chieh Fan, Chun-Fan Chang, Hsueh-Ting Chu, and Cheng-Yan Kao (2012): "Implement Web-version Bio-mimicry Peptide Design Algorithm from Trimer Alpha-carbon Mining towards Tri-peptide Fusing". *Symposium on Cloud and Services Computing* 2012.

Chun-Fan Chang, **Chen-Chieh Fan**, and Cheng-Yan Kao (2012): "Improve NPC CMI Likely on HLA-Agretope Docking in EBV-LMP1 Antigen of Putative Nona-peptide Structure from Peptide Design Algorithm Based on Tri-alpha-carbon Mining towards Tri-peptide Fusing". *11th International Conference in Bioinformatics (submitted)*

Hsueh-Ting Chu, William W. L. Hsiao, Theresa T. H. Tsao, Ching-Mao Chang, Yen-Wenn Liu, **Chen-Chieh Fan**, Han Lin, Hen-Hong Chang, Tze-Jung Yeh, Jen-Chih Chen, Chaur-Chin Chen and Cheng-Yan Kao (2012): "Quantitative assessment of mitochondrial DNA copies from whole genome sequencing as a biomarker of aging" *11th International Conference in Bioinformatics (submitted)*

Chun-Fan Chang, Chen-Chieh Fan and Cheng-Yan Kao (2012): "Improving CMI Likely from Docking HLA-1 Pit and Putative Agretope towards Mining Anchor- modified DNA Vaccine for in vitro Cell Activation and Action Complex Enhancement Drugs for in vivo Subject Therapy". *Translational Bioinformatics Conference 2012 (submitted)*

Patent Applications:

USA provisional patent:

Chun-Fan Chang, Cheng-Yan Kao, Thai-Yen Ling, Chen-hsiung Chan, Sheng-An Lee, Yu-Lun Kuo, **Chen-chieh Fan**, Fan-Chiang Sung (2009): CMI-inhibitory and/or CMI-stimulatory Modulators on AMI-naive and/or AMI-miscellaneous Status (Application No. 61/180,127; Filing Date: May 20, 2009)

USA utility patent:

Chun-Fan Chang, Wen-Chieh Chang, Tsung-Jui Chen, **Chen-Chieh Fan**, and Cheng-Yan Kao (2012): Method of restoring composite counter-balancing waveform compartments and extracting decomposite action waves from a composite waveform measurement and/or compartment derivatives of a corresponding operation system. (Application No. 13/526,525; Filing Date: June 18, 2012)

Taiwan

高成炎;高成榮;**范振杰**;朱學亭;張春梵。2010。睡眠呼吸中止防護裝置。中華民國發明專利第 M392409 號