การยึดจับอย่างจำเพาะของเอชแอลเอและแอนติเจนิกเพปไทด์ที่สัมพันธ์กับโรคเบห์เชตและโรค
ผิวหนังแข็ง

นางสาวศิริลักษณ์ กองแก้ว

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีชีวภาพ
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2560

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

SPECIFIC BINDING OF HLAs AND ANTIGENIC PEPTIDES ASSOCIATED WITH BEHÇET'S DISEASE AND SYSTEMIC SCLEROSIS

Miss Sirilak Kongkaew

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy Program in Biotechnology

Faculty of Science

Chulalongkorn University

Academic Year 2017

Thesis Title	SPECIFIC BINDING OF HLAs AND ANTIGENIC PEPTIDES ASSOCIATED WITH BEHÇET'S DISEASE AND SYSTEMIC SCLEROSIS
By	Miss Sirilak Kongkaew
Field of Study	Biotechnology
Thesis Advisor	Professor Dr. Supot Hannongbua
Thesis Co-Advisor	Assistant Professor Dr. Thanyada Rungrotmongkol

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.... Dean of the Faculty of Science
(Associate Professor Dr. Polkit Sangvanich)

THESIS COMMITTEE

.... Chairman
(Associate Professor Dr. Vudhichai Parasuk)

.... Thesis Advisor
(Professor Dr. Supot Hannongbua)

.... Thesis Co-Advisor
(Assistant Professor Dr. Thanyada Rungrotmongkol)

.... Examiner
(Professor Dr. Pornthep Sompornpisut)

.... Examiner
(Assistant Professor Dr. Kuakarun Krusong)

.... External Examiner
(Associate Professor Dr. Chonlaphat Sukasem)

ศิริลักษณ์ กองแก้ว : การยึดจับอย่างจำเพาะของเอชแอลเอและแอนติเจนิกเพปไทด์ที่สัมพันธ์กับโรคเบห์เชตและโรคผิวหนังแข็ง (SPECIFIC BINDING OF HLAs AND ANTIGENIC PEPTIDES ASSOCIATED WITH BEHÇET'S DISEASE AND SYSTEMIC SCLEROSIS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. ดร. สุพจน์ หารหนองบัว, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร. ธัญญดา รุ่งโรจน์มงคล, 100 หน้า.

โมเลกุลฮิวแมนลิวโคไซต์แอนติเจน (เอชแอลเอ) เป็นองค์ประกอบสำคัญสำหรับการทำงานในระบบภูมิคุ้มกัน มีบทบาทสำคัญในการนำเสนอแอนติเจนิกเพปไทด์ไปยังทีเซลล์รีเซพเตอร์ และกระตุ้นภูมิคุ้มกันร่างกาย ความสัมพันธ์ระหว่างยีนเอชแอลเอกับโรคติดเชื้อตลอดจนโรคภูมิคุ้มกันทำลายตนเองได้รับการระบุถึงปัจจัยเสี่ยงทางพันธุกรรม เช่น โรคเบห์เชตและโรคผิวหนังแข็งเป็นภาวะภูมิต้านทานตนเอง ส่งผลให้เกิดการอักเสบเรื้อรังในหลายระบบร่างกายและเกิดพังผืดบริเวณอวัยวะภายในตามลำดับ จึงได้ทำการตรวจสอบการเชื่อมโยงระหว่างยีนเอชแอลเอสัมพันธ์กับเพปไทด์ของตนเองสำหรับโรคเบห์เชตและโรคผิวหนังแข็งด้วยการจำลองแบบพลวัตเชิงโมเลกุลในโครงสร้างเชิงซ้อนเอชแอลเอพี สำหรับโรคเบห์เชต เพปไทด์ของตนเองคือเอ็มไอซีเอ-ทีเอ็มถูกจำลองร่วมกับยีนที่เกี่ยวข้องกับโรคเบห์เชตได้แก่ เอชแอลเอ-เอ*26:01 และเอชแอลเอ-บี*51:01 เปรียบเทียบกับยีนที่ไม่เกี่ยวข้องกับโรคเบห์เชต การจำลองโรคผิวหนังแข็งทำการศึกษาโครงสร้างเชิงซ้อนของเพปไทด์โทโปไอโซเมอเรส 1 (ทีโอพี1) กับยีนเอชแอลเอ-ดีอาร์ซึ่งแบ่งออกเป็นกลุ่มได้แก่ ยีนที่สัมพันธ์กับโรค (เอชแอลเอ-ดีอาร์บี1*08:02 เอชแอลเอ-ดีอาร์บี1*11:01 และเอชแอลเอ-ดีอาร์บี1*11:04) ยีนต้องสงสัย (เอชแอลเอ-ดีอาร์บี5*01:02) และยีนที่ไม่สัมพันธ์กับโรค (เอชแอลเอ-ดีอาร์บี1*01:01) อันตรกิริยาเฉพาะในแต่ละระบบถูกเปรียบเทียบต่อระบบอื่นๆ ในแง่ของพฤติกรรมเชิงพลวัตและพลังงานยึดจับ ผลการวิจัยแสดงให้เห็นว่าโครงสร้างเชิงซ้อนเอชแอลเอพีของระบบต้องสงสัยและระบบที่สัมพันธ์กับโรคมีเสถียรภาพและการยึดจับมากขึ้นเมื่อเปรียบเทียบกับระบบที่ไม่สัมพันธ์กับโรค นอกจากนี้ทีเซลล์รีเซพเตอร์ยังยึดจับกับโครงสร้างเชิงซ้อนเอชแอลเอและเพปไทด์ของตนเองตามเกณฑ์ขั้นต่ำของทีเซลล์รีเซพเตอร์ เช่นเดียวกับกรณีการติดเชื้อโรค การค้นพบนี้อาจสนับสนุนกลไกการเกิดโรคเบห์เชต โรคผิวหนังแข็งและภูมิต้านทานตนเอง นำไปสู่ความเข้าใจมากขึ้นและแนวทางสำหรับการรักษาโรคภูมิคุ้มกันทำลายตนเองต่างๆ

สาขาวิชา เทคโนโลยีชีวภาพ	ลายมือชื่อนิสิต ....
ปีการศึกษา 2560	ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
	ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

5772887223 : MAJOR BIOTECHNOLOGY

KEYWORDS: MOLECULAR DYNAMICS SIMULATIONS / HUMAN LEUKOCYTE ANTIGEN / SYSTEMIC SCLEROSIS / AUTOIMMUNE DISEASE / T-CELL RECEPTOR / BEHÇET'S DISEASE

SIRILAK KONGKAEW: SPECIFIC BINDING OF HLAs AND ANTIGENIC PEPTIDES ASSOCIATED WITH BEHÇET'S DISEASE AND SYSTEMIC SCLEROSIS. ADVISOR: PROF. DR. SUPOT HANNONGBUA, CO-ADVISOR: ASST. PROF. DR. THANYADA RUNGROTMONGKOL, 100 pp.

Human leukocyte antigen (HLA) molecules are essential components for immune functions. There play a major role to present antigenic-peptide for T-cell receptor (TCR) recognition and immune activation. Association between HLA genes and infectious as well as autoimmune diseases are identified genetic risk factors. For instance, Behçet's disease (BD) and systemic sclerosis (SSc) are autoimmunity result in chronic of multisystemic inflammation and visceral fibrosis, respectively. To investigate the link of HLAs correlated with self-peptides for BD and SSc, molecular dynamics (MD) simulations were applied on the HLAp complexes. For BD, a self-peptide named MICA-TM was modeled for two BD-association such as HLA-A*26:01 and HLA-B*51:01 in comparisons to non-association with BD. SSc simulation were studied on the complexes of topoisomerase 1 (Top1) peptide with various HLA-DR subtypes divided into association (HLA-DRB1*08:02, HLA-DRB1*11:01 and HLA-DRB1*11:04), suspect (HLA-DRB5*01:02) and non-association (HLA-DRB1*01:01) with SSc. The unique interaction for each system was compared to the others in terms of dynamical behaviors and binding free energies. Our results showed that HLAp of diseased suspect and association exhibited high protein stability and increased binding efficiency, in contrast to non-association. Moreover, T-cell receptor bound HLA with self-peptide by minimal criteria of TCR engagement as well as infectious cases. This finding might support mechanism of BD, SSc and autoimmunity TCR-recognition leading to a more understanding and guideline on the treatment for autoimmune diseases.

Field of Study: Biotechnology   Student's Signature _____
Academic Year: 2017   Advisor's Signature _____
                                          Co-Advisor's Signature _____

ACKNOWLEDGEMENTS

CONTENTS

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

LIST OF TABLES

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATA	anti-Top1 antibody
ATASSc	association of systemic sclerosis with anti-Topoisomerase 1 antibody
BD	Behçet's disease
CDR	complementarity-determining region
CG	conjugated gradient
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
K_D	equilibrium dissociation constant
MD	molecular dynamics
MHC	major histocompatibility complex
MICA	major histocompatibility complex class I chain related gene A
MICA-TM	major histocompatibility complex class I chain related gene A transmembrane
MM/GBSA	molecular mechanic/generalized Born surface area
MM/PBSA	molecular mechanic/Poisson-Boltzmann surface area
p	peptide
PBC	periodic boundary condition
PDB	Protein Data Bank
PME	Particle mesh Ewald
QM/MM-GBSA	quantum mechanics/molecular mechanics generalized Born surface area
RISM	reference interaction site model
RMSD	root mean square displacement
SASA	solvent accessible surface area
SSc	systemic sclerosis
TCR	T-cell receptor
Top1	topoisomerase 1

CHAPTER I
INTRODUCTION

1.1 Research rationality

Autoimmune diseases are found in patients every age, gender and nationality. Until now, autoimmune diseases are classified over 100 types by specific symptoms, but any diseases are cleared. Clinical and epidemiologic evidence as well as data from experimental animals demonstrate that a tendency to develop autoimmune disease is inherited. Moreover, immunity and genetic markers carry on patients that are interested to record relationship. The time consuming, expensive cost and understanding of autoimmune mechanism are limitations into clinical assay. Molecular dynamics simulation is used to solve the problems. This kind of study may set the basis for more targeted and rational therapeutic approaches.

1.2 Immune system

Human life is harmfully invaded by pathogens like microbe, bacteria or virus leading to diseases. Yet we are still survive without severe illness causing immune protection. The immune systems are mainly classified two classes consisting of i) the innate immune system and ii) the adaptive immune system. This research will describe in part of the adaptive immune system.

1.2.1 *Human leukocyte antigen*

Human leukocyte antigen (HLA) proteins decode from a complex genes on chromosome 6 (Figure 1) [1]. It is also familiar with major histocompatibility complex (MHC) gene which is called HLA for human. Classification of HLA has three classes upon functional features, however, only class I and II are importantly included immune responsibility [2].
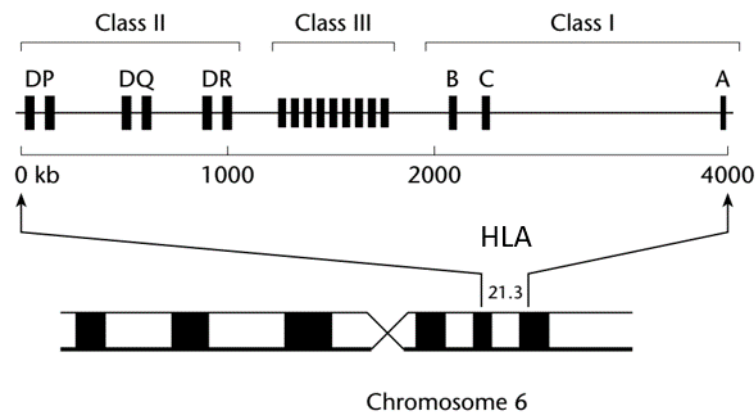
Figure 1. Human leukocyte antigen (HLA) genes locate on chromosome 6 [1].
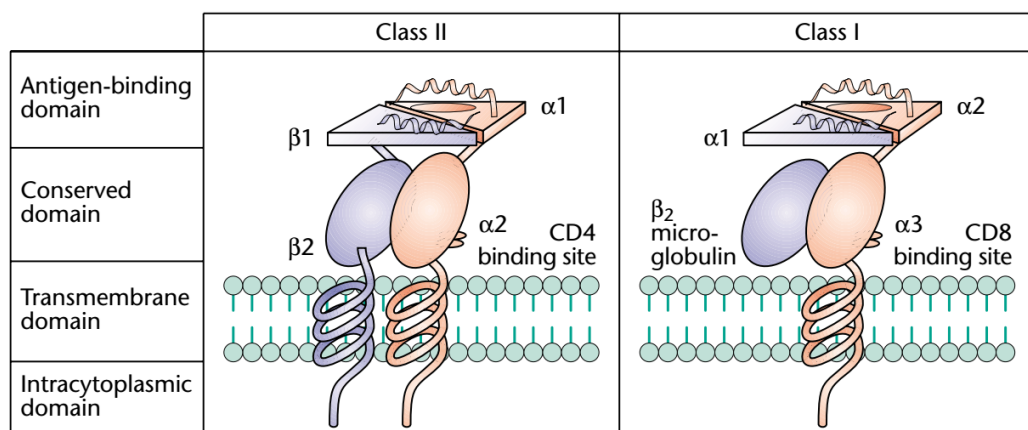
Figure 2. The elements of HLA class I and II are illustrated simple cartoon [2].

a. HLA class I molecules

HLA class I consists of three loci: HLA-A, HLA-B and HLA-C which are an antigenic presenter for CD8+ T-cell [3]. They specifically bind antigenic peptide in length of 8 to13 amino acids. Molecular structure is assembled eight anti-parallel of β-sheets between α1 and α2 helixes for antigenic binding groove and α3-helix for microglobulin (β_2m) supporting (Figure 2). HLA class I molecules are found on cell surface of dendritic cells. A viral prevention by HLA class I, viral genome is decoded and then digested into short peptide 8-13 mers by proteasomes [4]. Viral peptides are transported to the endoplasmic reticulum (ER), where located HLA class I molecules.

Viral peptides are loaded on HLA binding groove and presented to CD8+ T-cell on cell surface.

b. HLA class II molecules

HLA class II consists of three loci: HLA-DP, HLA-DQ and HLA-DR which are an antigen presenter for CD4+ T-cell [5]. They particularly mounted antigenic peptide in length of 15 to 25 amino acids. Basic structure is constructed heterodimeric α and β chains with non-covalent interaction to form an antigenic binding cleft containing eight β-sheets in between two anti-parallel α-helices (Figure 2). For HLA-DRs, the DRA and DRB genes are decoded for α and β chains, respectively, which have high polymorphic domain of α1 and β1 to produce a binding groove. HLA class II molecules are expressed on antigen presenting cells such as dendritic cells, endothelial cells, B lymphocytes and macrophages.

Table 1. HLA allele numbers (https://www.ebi.ac.uk/ipd/imgt/hla/stats.html)

HLA class I					
Gene	A		B		C
alleles	3,997		4,859		3,605
Proteins	2,792		3,518		2,497
nulls	186		147		131
HLA class II					
Gene	DRA	DRB	DQA	DQB	DPA	DPB
alleles	7	2,395	92	1,152	61	948
Proteins	2	1,751	35	779	28	658
nulls	0	66	3	31	0	22

The HLA nomenclature is represented a unique number to specify each HLA allele (Table 1). For instance, HLA-A*02:101:01:02N consequently refers to gene name (HLA), specific locus of gene (A), allele group (02), specific protein (101), coding region with DNA substitution (01), non-coding region (02), and null allele (N).
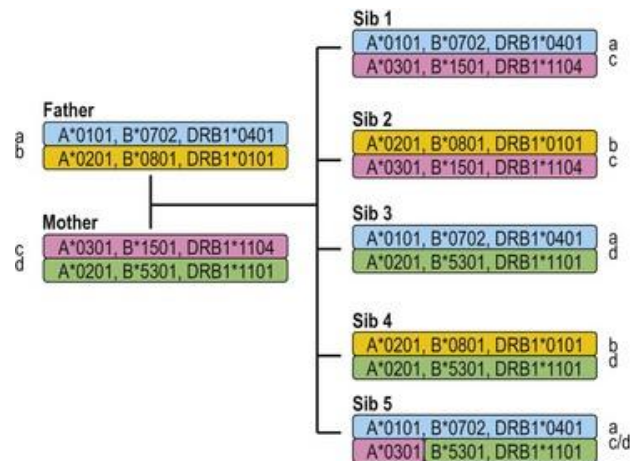
Figure 3. Illustration of HLA alleles in parent and children with at least five inherits [1].

The HLA genotypes of children are randomly inherited from parental haplotypes. For example, Figure 3 shows possibility HLA genes within 5 sibling. Four of them are common chromosomal crossover, whereas the fifth sibling is recombinant with HLA-A*0301.

1.2.2 T-cell receptor

T-cell receptor (TCR) molecule recognizing antigenic peptide laid on human leukocyte antigen presenter (HLAp) is the first step in the T cell-mediated adaptive immune responses [6]. The T cell-mediated immune response against an antigenic peptide consists of two processes, which are *i)* this peptide is detected by HLA and *ii)* HLAp complex is recognized by TCR. A T-cell recognizes more than a million different antigenic peptides through combination of heterodimeric TCRα and TCRβ chains for signal transmitting to trigger immunity [7-10]. To comprehend more about the antigenic recognition, the TCR structure is necessarily denoted as a prerequisite. Within the TCR, triple pair loops of hypervariable complementarity-determining regions (CDRs) are antigen-binding sites which recognize and bind to HLAp complex (Figure 4A) [11]. A canonical diagonal orientation of TCR generally interacts with HLA binding groove and with the antigenic peptide by germ line-encoded as CDR1 and CDR2 loops, and somatically rearranged as CDR3 loops (Figure 4B and C), respectively [12]. The CD8+ T-cell specifically detects antigenic peptides within

typically 8 to 13 amino acids length by HLA class I binding [4, 13]. The heavy chain of HLA class I is folded into a binding groove comprising of eight antiparallel β-sheets as a base between two flanking α-helices. As shown in Figure 4, the distal end is α3-subdomain, non-covalently binding with β2-microgluobulin (β2m) [14]. Typically, the peptide is tightly bound at the close ends of binding groove causing a long peptide to protrude out of the groove, and this protruding part is recognized by TCR. Researches have reported that TCR/pHLA class I is hierarchically interacting with different antigenic motifs in the order pathogen < cancer < self, according to equilibrium dissociation constant (K_D) values [15, 16].

The immune system protects a healthy body against pathogens including viruses, bacteria or microbes. In autoimmune cases, abnormal states of autoimmunity attacks own organelles and source of self-antigenic peptide. Unfortunately, the exact cause of autoimmune disease is not well understood. However, many theories suggest that autoimmunity is triggered by environmental factors, chemical irritants, drugs and biological invaders [17]. "How can a T-cell receptor response self-peptide as an antigenic peptide?" is a key question for autoimmune study [18]. Other researches have studied the effect of a TCR interacting with different peptides with HLA complex [10, 19-21]. In this study, we modeled T cell-mediated autoimmune response from risk factors to autoimmunity, which was represented in Behçet's disease (BD). BD is a kind of autoimmune disease which affects multiple parts of the body [22]. In previous reports, genetic markers with BD-association are revealed as HLA-B*51:01 protein and MICA peptide (especially, a sequence of AAAAAIFVI) cohesion [23-26]. HLA-B*51:01 is wildly known a strong BD's association as well as a resistance to HIV infection [27]. In this work, the self-autoantigen with TCR interaction was studied by molecular dynamics simulation to inspect physical properties and binding interactions to viral infected system which is the known as TCR-recognition and is resolved as a crystal structure. We aimed to answer the question of how T-cell recognizes self-peptide as an antigenic peptide. Moreover, our results might be a useful guideline for the treatment of autoimmune diseases.
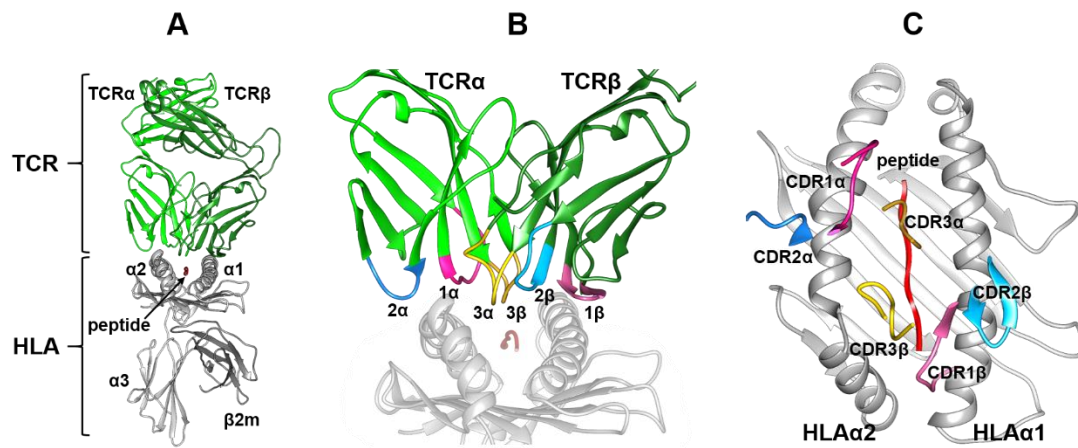
Figure 4. A) TCR/HLAp elements are depicted a ribbon model. B) An α/β T cell receptor (TCR) is colored the positions of the six variable complementarity-determining region (CDR) loops by pink, blue and yellow for CDR1, CDR2 and CDR3. C) The positions of TCR restricting on HLAp interface.

1.3 Autoimmune diseases

Immune system have an important to protect healthy body against pathogens includes virus, bacteria or microbes. In autoimmune cases, immune mis-attacks own organelles and source of self-peptide which can triggers autoimmunity. Unfortunately, the exact cause of autoimmune disease is not well understand. However, many theories suggested that multiple factors involving environmental and chemical irritants, drugs, hormones and biological invaders were autoimmune activation [27]. Nowadays, autoimmune diseases are classified over 100 types by specific symptoms. Many people with autoimmune diseases have characteristic types of self-marker genes of HLA (Table 9). If you have genetic type associate with autoimmune disease, you just higher risk to be but won't get till something trigger.

Table 2. HLA allele associated with autoimmune disease

Autoimmune disease	HLA allele	Organelles target / symptoms
Ankylosing Spondylitis	HLA-B27	arthritis affecting the spine and pelvis
Behçet's Disease	HLA-B51	a systemic inflammatory vasculitis, oral ulcers, genital ulcers, skin lesions and ocular lesions
Coeliac disease	HLA-DQ2, HLA-DQ8	an inflammatory disorder of the small intestine
Drug hypersensitivity	HLA-B*57:01	a combination of fever, skin rash and internal organ involvement, generally occurring more than 1 week after exposure to a drug
Type 1 diabetes	HLA-DR17, HLA-DQ2 HLA-DR4, HLA-DQ8	the body does not produce enough insulin
Multiple sclerosis	HLA-DR*15:01	myelin sheath degradation
Rheumatoid Arthritis	HLA-DRB1*01:01, HLA-DRB1*01:02, HLA-DRB1*04:01, HLA-DRB1*04:04, HLA-DRB1*04:05, HLA-DRB1*04:08	a chronic inflammatory arthritis that affects multiple synovial lined joints
Systemic Lupus Erythematosus	HLA-DRB1*03:01, HLA-DQB1*02:01, HLA-DRB1*15:01, HLA-DQB1*06:02	a systemic inflammatory including skin, joint, haematologic, neurologic, renal and other organ involvement

Autoimmune disease	HLA allele	Organelles target / symptoms
Systemic sclerosis	HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*11:04	i) non-elasticity of connective tissue, ii) disorder of multi-internal organs and iii) arteriosclerosis

1.4 Literature reviews

1.4.1 Behçet's disease

      Behçet's disease (BD) has caused recurrent inflammation of symptom complex in genital and oral ulceration, inflammatory lesion of central nervous systems, gastrointestinal tract and eyes leading to blindness [22]. This disease is usually reported in adults aged between 18 to 40 year olds and presents a higher mortality risk in males than females [28]. A high epidemiology of BD is commonly found in Japan, Turkey, China and the Middle East and the Mediterranean countries [29]. The systemic inflammatory disease is characterized by recurrent exacerbations and spontaneous remissions with varying healing times among patients, and the disease course usually becomes several years or more. On the ocular disease, recurrent and often bilateral episodes of acute exacerbations of intraocular inflammation (ocular attacks) in the anterior chamber (iridocyclitis) normally occur with or without posterior involvement retinal vasculitis, hemorrhages, exudates, retinal vein occlusion and optic neuritis). Recurrent ocular attacks of posterior involvement lead to poor visual prognosis [30]. Intestinal and neurological diseases are serious complications of BD and can severely deteriorate the patients' psychosomatic status. Because BD can induce severe damage in the eyes and other organs, it is important to clarify the pathogenesis of this disease and to develop a novel treatment based on the pathogenesis. The exact etiology of this disease has not been clarified yet, but it is probably mediated by a combination of excessive immune reactions, genetics factors [31], immune reactions against infectious agents [32], heat shock proteins [33], oxidative stress [34] and environmental factors. Indeed, several inflammatory cytokines involved in acute inflammatory reactions, such as interleukin (IL)-6, tumor necrosis factor-α and IL-8 might play a role in the pathogenesis of this disease [35]. In addition, genetic factors, including the human leukocyte antigen (HLA)-B51 alleles and especially the HLA-B*51:01 allele is strongly

associated with BD across different susceptible ethnic groups [36, 37]. Recently, a meta-analysis of 78 studies involving 4,800 cases of BD and 16,289 controls reported the pooled odds ratio (OR) of HLA B51/B5 carriers to develop BD compared to non-carriers (OR = 5.78, 95% CI = 5.00–6.67) [38]. The notion of an environmental trigger of BD in patients with genetic susceptibilities has long been advocated. Several infectious agents have been investigated, especially bacteria (*Streptococcus*, *Mycoplasma* and *Helicobacter pylori*) and viruses (Herpes simplex virus 1 and 2, Hepatitis virus and Parvovirus B19) [39]. The autoimmunity against bacterial and human heat-shock proteins (HSP) was also speculated as a trigger for BD, because autoantibodies against HSP65 and HSP60 have been reported in BD patients [40].On the other hand, the genetic association of the major histocompatibility complex (MHC) class I chain related to gene A (MICA) with BD has been reported in various ethnic groups [41, 42], and a strong association of the polymorphism in the transmembrane region of MICA (MICA009) with BD was observed [43, 44]. The self nonapeptide of AAAAAIFVI located on the MICA transmembrane region (MICA-TM), induced autoreactive CD8+ cytotoxic T lymphocytes in BD patients with HLA-B51 [26]. Accordingly, the MICA-TM peptide is one of the candidate bechetogenic epitopes [26, 43], where the self-antigen AAAAAIFVI could function as the antigen that triggers the T-cell receptor and develops to an autoimmune reaction [26, 45, 46]. The MICA is usually expressed in fibroblasts, monocytes, epithelial cells and endothelial cells in a stress-dependent manner, and plays a key role in initiating the antibody-dependent rejection of organ-transplants [47-49]. The HLA-B*51:01 and MICA genes are reported as the predominant risk BD whereas HLA-B*35 is absented [25, 50]. Moreover, BD patients with thrombophlebitis increased HLA-B51 (OR = 4.235, 95% CI = 1.099–16.324) and decreased HLA-B35 frequency with (OR = 0.194, 95% CI = 0.041–0.923) [51].

The association of BD with the HLA-B*51:01 allele in various ethnic groups in Japan and Korea has been frequently observed [29], while the HLA-A*26:01 allele was detected in BD patients in Greece, Japan and Taiwan [52-54]. Some possible candidate genes for BD have been studied in various HLA-A and-B alleles [55]. Recently, in Japanese uveitis patients, BD was associated with the HLA-A*26:01 allele at a 37.5% phenotype frequency more than the controls, and so HLA-A*26:01 is a possible marker

as a susceptible allele for ophthalmic BD in Japanese ethnics [56]. By meta-analysis and positive pathergy tests of the Japanese data, the BD clinical manifestations of uveitis, and skin lesions and arthritis were found to be significantly associated with the HLA-A*26:01 (OR = 1.89, 95% CI = 1.41-2.53), and HLA-A*02:07 (OR = 1.96, 95% CI = 1.34-2.85) alleles, respectively. The frequency of HLA-A*11:01 (OR = 0.78, 95% CI = 0.60-1.03) tended to be lower in BD patients [57].

Currently there is no treatment (e.g. vaccine) against BD based upon the HLA recognition antigens despite the apparent restricted association with specific alleles, since the individual risk factors with HLA are unknown.

1.4.2 *Systemic sclerosis disease*

Systemic sclerosis (SSc) is an autoimmune multisystem disease, clinically characterized by scleroderma, visceral organ fibrosis including lung, kidney, and gastrointestinal tract, microvascular injury, and immune activation with specific autoantibodies [58, 59]. The prevalence of SSc is more often in female than in male at a 4:1 ratio, and the highest occurrence appears in the age range around sixty [60-62]. The majority causes of death in SSc patients are fibrosis and pulmonary hypertension [63]. Unfortunately, the etiology of SSc is not well understood, but the possible induction by environmental agents, hormones, genetic factors and irregular immunity is assumed [64, 65]. SSc is commonly classified into two clinical subtypes with symptomatic signs; systemic sclerosis with limited cutaneous sclerosis (lcSSc) and systemic sclerosis with diffuse cutaneous sclerosis (dcSSc) [66]. In dcSSc, diffuse scleroderma and the failures in heart, gastrointestinal tract, lung, renal and other internal organs are observed, and the pulmonary hypertension is more common in lcSSc [62, 66]. The positive autoantibody testing of anti-topoisomerase I and anti-centromere antibodies are commonly accepted to clinically diagnose for dcSSc and lcSSc subtypes [67-69], respectively. The anti-Top1 antibody (ATA) directly resisted topoisomerase I (Top1) activity involving an increased collagen transcription [70]. The ATA was frequently observed in dcSSc-patients' sera who had drastic pulmonary fibrosis and cardiac arrest related to accumulation of collagen deposition in tissues leading to death [71]. Nevertheless, the ATA is more frequent in dcSSc, but is uncommon in lcSSc. Top1 peptides were identified as antigenic determinants associated with SSc positive

for ATA (SSc with ATA, ATASSc) using B-cell epitope mapping of ATA response. Interestingly, the twenty-mer sequencing RIANFKIEPPGLFRGRGNHP (349–368) from total 63 Top1 fragments exhibited ATASSc-association by 71% sensitivity and 98% specific binding with ATA [72]. According to the crystalized structure, this sequence is located at an exposed area, where it could be easily recognized by ATA [72, 73].

Although the etiology of SSc is not clarified yet, the contribution of several genetic factors including human leukocyte antigen (HLA)-DR to the pathogenesis of SSc were reported [74-77]. HLA-DR genes correlated with ATA are significantly remarked as the risk factors of ATASSc and dcSSc [75, 78-81]. For example, HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*11:04 and HLA-DRB1*15:02 associated with ATASSc was observed in Mexican admixed [62], Caucasian [82], African-America/Italian-Spanish [57, 83], and Chinese/Thai/Japanese [75, 84, 85] patients, respectively. In addition, the strong linkage disequilibrium of HLA-DRB1*15:02 is widely known as HLA-DRB5*01:02 in Thai patients with ATASSc [84]. Because of the very strong linkage disequilibrium, it is difficult to define HLA-DRB5*01:02 as the real susceptible gene. Interestingly, HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*11:04 alleles and HLA-DRB5*01:02 allele have the similar amino acids sequence at the hypervariable region of the HLA-DR β chain (residues 67-71, FLEDR) [86]. The identical hypervariable motif on HLA-DRB5*01:02 is possibly suggested as susceptibility gene of ATASSc, however the present study specifies a suspect ATASSc. On the other hand, HLA-DRB1*01:01 does not relate with the ATA, but it is instead specific to the ACA associated with SSc among Caucasian and Japanese ethnics [82, 85]. Recently, HLA-DR alleles and Top1 were reported to be susceptible to ATASSc and led to the description of binding interactions in ATASSc pathogenesis [65].
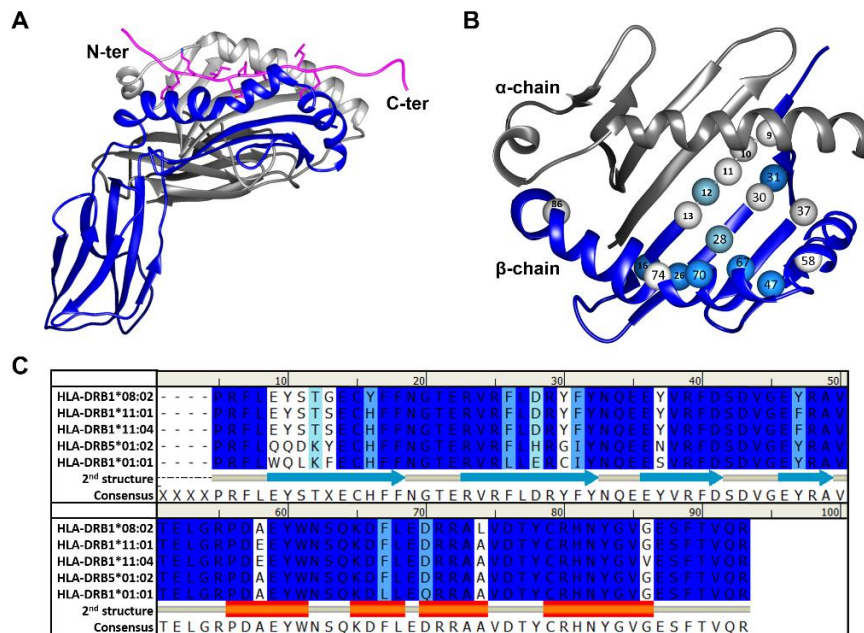
Figure 5. Basic structure of class II HLA. (A) HLA-DR is constructed of alpha (gray) and beta (blue) chains to restrict antigenic peptide (magenta ribbon). (B) Polymorphic residues of HLA-DR β-chain are marked as spheres on the top view of binding cleft, and (C) also represents sequence alignment whereas conserve residues of alpha-chain is not show.

This study had challengingly attempted to identify and characterize the relation of self-antigenic (RIANFKIEPPGLFRGRGNHP) peptide from Top1 protein in selective binding with HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*11:04, HLA-DRB5*01:02 or HLA-DRB1*01:01 molecules which could individually contribute to dcSSc disease risk at different level. Based on the HLA function and genetic markers, the interaction of HLA-DR/Top1 was evidenced by molecular dynamics (MD) simulations. Indeed, molecular understanding of the binding recognition between the motif of self-peptide and HLA-DRB alleles is useful for the avoidance of infectious agents carrying epitope such as self-antigen (Top1) that causes dcSSc disease. Furthermore, this will be used for drug design against dcSSc by HLA/self-peptide inhibition.

1.5 Objectives

i)   To predict and compare the structural stability of autoimmune associated and non-associated HLA in complex with self-antigenic peptide by binding free energy calculations

ii) To reveal the selective binding of HLA/self-peptide complexes associated to both autoimmune diseases (Behçet's disease and systemic sclerosis)

iii) To solve the question, how can a T-cell receptor response self-peptide as an antigenic peptide?

CHAPTER II
THEORIES
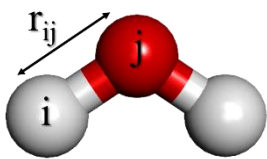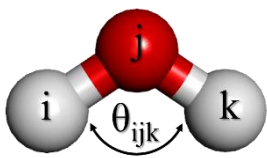
2.1 Molecular dynamics (MD) simulation
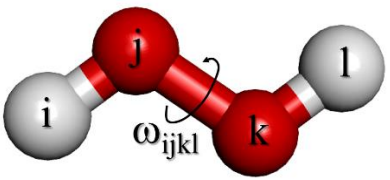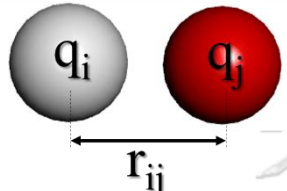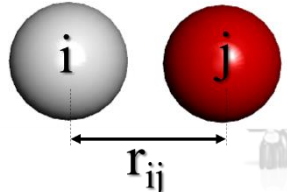
The study of molecular structure is an importance in biological understanding. To observe molecular interactions by experimental processes is difficult, expensive, and takes a long time. MD simulation is used to assist for experimental limitations. MD techniques have been widely study function of biological macromolecules using specific potential energy functions, E_{pot} [87]. Basically, every macromolecule is constructed with atom interactions, which have distinct energy and forces leading to calculate for MD with potential energy functions or force-field. An empirical force-field is approximated as a series of charged points (atoms) linked by springs (bonds). A collection of force-field in molecular mechanics includes simple terms: bonded and non-bonded, as show details in Table 3.

$$E_{pot} = E_{bonded} + E_{non\text{-}bonded} \qquad (1)$$

$$E_{pot} = (E_{bond} + E_{angle} + E_{dihedral}) + (E_{ele} + E_{vdW}) \qquad (2)$$

Table 3. Force-field in molecular mechanics

Interaction	Described function
Valence terms:	
Bond stretching	$E_{bond} = \sum_{bond} K_r (r - r_0)^2$ r_0, r = ideal and actual bond length K_r = force constant for bond stretching
Bond bending	$E_{angle} = \sum_{angle} K_\theta (\theta - \theta_0)^2$ θ_0, θ = ideal and actual bond angle K_θ = force constant for bond bending

Dihedral	$$E_{dihedral} = \sum_{dihedral} \frac{V_n}{2}\left(1 + \cos[n\omega - \delta]\right)$$
	V_n = rotation barrier height n = rotate periodic ω = torsion angle δ = conformation of molecule in phase
Non-valence terms:	
Electrostatic 	$$E_{ele} = \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$ q_i, q_j = atomic charges of i and j atoms r_{ij} = distant between i and j atoms ε_0 = dielectric constant
van der Waals 	$$E_{vdw} = \sum_{i<j}^{atoms} 4\varepsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right)$$ ε = the well depth σ = particle distance at potential energy is zero According to the definition of London's dispersion, atomic electrons overlap and repulse which is the r^{12} term. The r^{6} is attractive long-range term.

Classical MD simulation is developed based on the Newton's second law equation of motion, $F_i = m_i a_i$. Where Fi is external force acting on the i atom, m_i and a_i are mass and acceleration of the i atom, respectively. Basically, the acceleration can be evaluated by the first-order differential of velocity respect to time or second-order differential of the position respect to time. Moreover, the external force can also obtained from the negative gradient of potential energy (E_{pot}) as show in eq.1 and 2.

Each simulated output into trajectory is snapped atomic next position along update-time (*t+dt*). Nowadays, MD simulation is achieve on popular programs such as AMBER, CHARMM, ACEMD, GROMACS, or NAMD.

2.1.1 cutoff distance

All nonbonded pair interactions acting on the i atom over the containing millions atoms in simulation system takes a very computational expense. To overcome this limitation, the non-bonded terms in van der Waals and electrostatic are considerably truncated in more compacted region called cutoff radius, to reduce the time-consuming of MD simulation. Normally, the cutoff limit is set around 10-12 Å centered on the focusing atom. An applied cutoff may have not accuracy of the calculated forces, however, other rest atoms is assumed to less effect cause of too long length.

2.1.2 Periodic Boundary Condition (PBC)

A realistic environment of molecular biology has a huge number of particles surrounding especially for water molecules. Periodic boundary condition is an implementation to mimic the infinite bulk surrounding protein target without surface effects. Shapes of boundary are called simulation box, allowed for any three-dimensional geometry that can properly repeat every site. The particles in the simulation box move in unison with their images. Various models has been used in simulation, for example cubic, octahedral, orthorhombic, rectangular boxes, *etc*.

2.1.3 Water models

An appearance of water molecules in protein-protein interaction is supplementary or prohibitive activity. To get more accuracy simulated results and to compare with biological experiment, water molecules are necessary included in computational simulations. Water 3D-models are developed from realistic environment such as geometry, density, hydrogen bond forming, atomic charges and Lenard-Jones (LJ) parameters. Many types of models such as 3, 4, 5 and 6-site are classified by number of interacting site. The parameters of model geometries are O-H distance, H-

O-H angle involving LJ and atomic charges. Three-site models have 3 point charges which implement for high computation leading to widely use. For example, TIP3P, SPC and SPC/E are member of 3-site water models. Particularly, TIP3P has better performance in calculating specific heats.
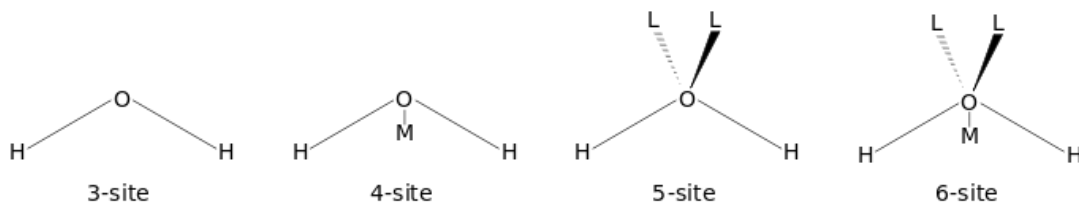
Figure 6. Types of water models for molecular dynamics simulation

(https://en.wikipedia.org/wiki/Water_model)

2.1.4 Integration algorithm

Once the potential energy was determined, the position and velocity of atoms based on equation of motion can be predicted at time $t + \Delta t$ using integration algorithm. There are numerous algorithms such as Verlet algorithm, leapfrog algorithm, Beeman's algorithm etc. All integration algorithms approximate the positions, velocities and accelerations by Taylor series expansion given in eq. 3-5.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + ... \tag{3}$$

$$v(t + \Delta t) = v(t) + a(t)\Delta t + \frac{1}{2}b(t)\Delta t^2 + ... \tag{4}$$

$$a(t + \Delta t) = a(t) + b(t)\Delta t + ... \tag{5}$$

Integration algorithm which is commonly used for molecular modeling is the Verlet algorithm, written as following eq. 6 and 7.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{6}$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \tag{7}$$

with combinations of these two equations, one gets

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 . \tag{8}$$

This technique determines the new position at time $t + \Delta t$ or $t - \Delta t$ from the current position and acceleration, see in eq. 8.
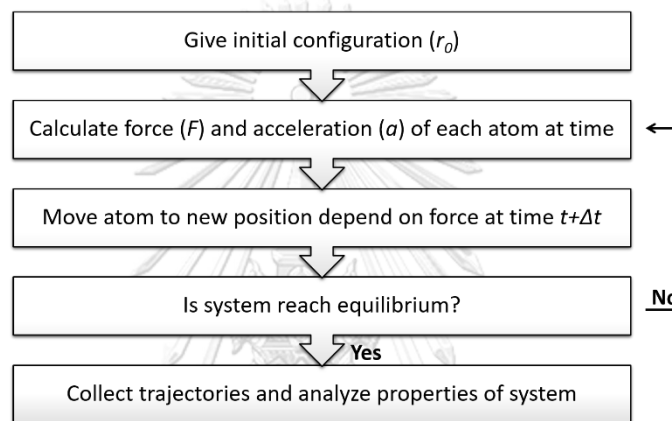
Figure 7. Basic process of molecular dynamics simulation

1. The initial configuration of protein are obtained from protein data bank in PDB format, that included atomic coordinates (x, y, z). The starting velocity (v_0) is generated using random number generator.

2. The potential energies of molecular mechanic are applied with Newton's second law of motion to generate force (F) and acceleration (a).

3. Atoms are moved to the new positions at the time $t + \Delta t$

4. Systemic stability is monitored by considering the energy, temperature and root mean square deviation (RMSD).

5. Trajectories are harvested for analysis.

2.2 Implicit solvation

To reduce cost of computation associated with all-atom simulations, implicit solvent models or continuum solvation are instead of individual "explicit" solvent molecules. Implicit solvation method is frequently used to roughly calculate free energy interaction between solute and solvent. The free energy of implicit solvation can divided into three contributions as i) the cavity formation of solvent rearrangement due to the protein, ii) hydrophobicity effect of protein surface, and iii) the electrostatic interactions of polarized solvent (eq. 9). The charge distributions of solute-solvent interaction are estimated with both individual solute and solvent. The reduction from explicit to implicit solvent still has limitation as a treatment of charge effects and parameterization. However, this method is directly operated in solvent accessible surface area (SASA) method.

$$\Delta G_{solvation} \; = \; \Delta G_{cavity} \; + \; \Delta G_{disp} \; + \; \Delta G_{elec} \tag{9}$$

2.2.1 Solvent accessible surface area (SASA)

Basic idea of a cavity model is imaged as sphere, as same as, surface area of protein molecule is computationally defined as the vdW mask of a solute molecule. Solvent accession of protein surface is created by solvent sphere rolling on surface (Figure 8). The solvation of solute bases on cavity surface and dispersing interaction, whereas the dielectric of solvate constants. Total SASA creation can generated with eq. 10.
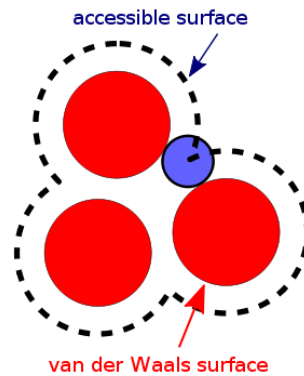
Figure 8. Solvent-accessible surface area of the red particles (black dash) and their van der Waals surface of each atomic radius (red).

(https://en.wikipedia.org/wiki/Accessible_surface_area)

$$\Delta G_{SASA} = \Delta G_{cavity} + \Delta G_{dispersion} \approx \gamma\,SASA + \beta \tag{10}$$

Non-electrostatic contribution is presented as ΔG_{SASA}, which is solvent accessible surface area, to consider proportionate cavity and solute–solvent van der Waals terms with surface tension constant (γ) to 0.0072 kcal/mol·Å^2 using a solvate probe radius of 1.4 Å and offset to correct (option) the value of the non-polar contribution to the solvation free energy term (β). Whilst, the solvation in electrostatic part is approached by Poisson-Boltzmann equation and generalized Born model.

2.2.2 Poisson-Boltzmann (PB) model

To compute electrostatic potential $\phi(r)$ for continuum solvent models, the Poisson equation is conducted as a spatial charge distribution $\rho(r)$ in homogeneous medium with dielectric constant (ε). The ∇ is the Laplacian operator for three-dimensional Cartesian coordinates.

$$\nabla\varepsilon(r)\nabla\varphi(r) = -4\pi\,\rho(r) \tag{11}$$

In case of heterogeneous solvent, dielectric constant is determined by position of mixtures, $\varepsilon(r)$. The Poisson equation is used for nonionic conditions, however, the

distribution of ionic mobile is executed by the Boltzmann distribution as called Poisson-Boltzmann (PB) equation:

$$\nabla \bullet \left(\varepsilon(r)\nabla\phi(r)\right) - \kappa^2\left(\frac{k_B T}{q}\right)sinh\left[\frac{q\,\phi(r)}{k_B T}\right] = -4\pi\,\rho(r)\,, \tag{12}$$

$$\kappa^2(r) = \frac{8\pi\,q^2 I}{k_B T}\,, \tag{13}$$

where κ^2 is the Debye–Hückel parameter dependent on the ionic strength of solution (I); k_B Boltzmann's constant; T is the absolute temperature; and q is electrolyte charge. The electrostatic free energy is given in eq. 14, where charge density of solute multiplies electrostatic potential in solution (ϕ_{sol}) and for the same charge distribution in vacuum (ϕ_{gas}).

$$\Delta G_{ele} = \frac{1}{2}\int \rho(r)\,(\,\phi_{sol} - \phi_{gas}\,)dr \tag{14}$$

2.2.3 Generalized Born (GB) model

The Generalized Born (GB) method is developed using PB method to solve the electrostatic contribution in term of the solvation free energy. A simple calculation is the burial particle into molecule by approximation of the Born radius of the particle, elsewhere distance of the partial atomic charges (r_{ij}), Born radii (α), and the smoothing function (f_{GB}). The Debye-Hückel screening parameter (κ) is applied as PB equation.

$$\Delta G_{GB} = -\frac{1}{2}\sum_{ij}^{atoms}\frac{q_i q_j}{f_{GB}\left(r_{ij},\alpha_i,\alpha_j\right)}\left(1 - \frac{e^{-\kappa f_{GB_{ij}}}}{\varepsilon_w}\right) \tag{15}$$

$$f_{GB} = \left[r_{ij}^2 + \alpha_i\alpha_j\exp\left(\frac{-r_{ij}^2}{4\alpha_i\alpha_j}\right)\right]^{\frac{1}{2}} \tag{16}$$

The GB model is usually benchmarked or even parameterized against the PB model.

2.3 QM/MMGBSA

Likewise, binding free energy with quantum mechanics (QM) method is computed within the QM/MM-GBSA approach[88, 89]. QM/MM approach in a couple with GB model is examined to calculate enthalpy and solvation, respectively.
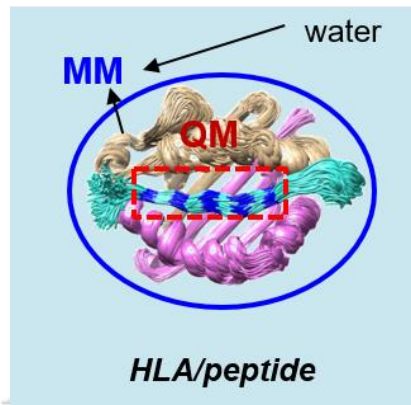
Figure 9. An example of QM and MM regions identification

Two regions of QM and MM are partitioned, then, energy calculation on these two regions can be employed by using the two alternative schemes as additive and subtractive approaches. The total energy of the system is obtained from the summation of each energetic term together.

$$E_{TOTAL} = E_{QM} + E_{MM} + E_{QM-MM} \tag{17}$$

From eq.17, E_{QM} is energy of the QM region (QM treatment), E_{MM} is the energy of the MM region (force field treatment) and E_{QM-MM} is the interaction between QM and MM region. In addition, the E_{QM-MM} term includes an electrostatic interaction ($E_{ele,QM-MM}$), van der Waals interaction ($E_{vdW,QM-MM}$) and bonded interaction ($E_{bonded,QM-MM}$).

$$E_{QM-MM} = E_{ele,QM-MM} + E_{vdW,QM-MM} \tag{18}$$

$$E_{ele,QM-MM} = -\sum_{i \in QM} \sum_{j \in MM} \frac{Q_j}{r_{ij}} + \sum_{A \in QM} \sum_{j \in MM} \frac{Z_A Q_j}{R_{Aj}} \tag{19}$$

$$E_{vdW,QM-MM} = \sum_{i \in QM} \sum_{j \in MM} 4\varepsilon_{ij} \left\{ \left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^{6} \right\} \tag{20}$$

In subtractive scheme, the system is divided layers and subtracts double counted energies of the smaller with different energetic calculated method as follows:

$$E_{TOTAL} = E_{QM}^{QM\ region} + E_{MM}^{QM+MM\ region} - E_{MM}^{QM\ region} \tag{21}$$

The chosen of either additive or subtractive schemes is not critical whilst the QM region determination should be concerned due to the energy interference between QM and MM regions especially in the case of through bond cutting model.

SCC-DFTB Approach

QM calculation based on self-consistent charge density functional tight binding (SCC-DFTB) method[90] is specifically applied on the hot-spot residues. The semi-empirical implementation of SCC-DFTB is a high level and accurate potential method that can be applied for the calculation of the binding free energy. In SCC-DFTB approximation, a second order expansion of the DFT Kohn-Sham energy with respect to charge density fluctuations expresses the total energy.

$$E_{TOTAL}^{DFTB} = \sum_{i}^{occ} n_i \langle \Psi_i | \hat{H}_0 | \Psi_i \rangle + \frac{1}{2} \sum_{\alpha\beta}^{N} \gamma_{\alpha\beta} \Delta_{q\alpha} \Delta_{q\beta} + E_{rep}$$

$$= E_0[n_0] + E_2[n_0, \delta_n] + E_{rep} \tag{22}$$

The eigenstates Ψ_i of the effective Kohn-Sham Hamiltonian $\hat{H}_0$ derived under the approximation that the initial electronic density of the many-atom system, where n_0 is neutral atomic charge densities. E_2 is exchange correlation functional with respect to charge density fluctuations δ_n, approximated as atomic point like charges (Δ_q) together with an analytical interpolating function γRβ. This term is important electronegativity differences of atoms. E_{rep} accounts for the double-counting terms and the ionion core interaction of pairwise repulsive potentials.

2.4 Alanine scan analysis

Experimental alanine scanning mutagenesis is expensive and time-consuming. Theoretical methods are needed. Brief description of alanine scan mutagenesis method represents to binding free energy difference that is shown in Figure 10.
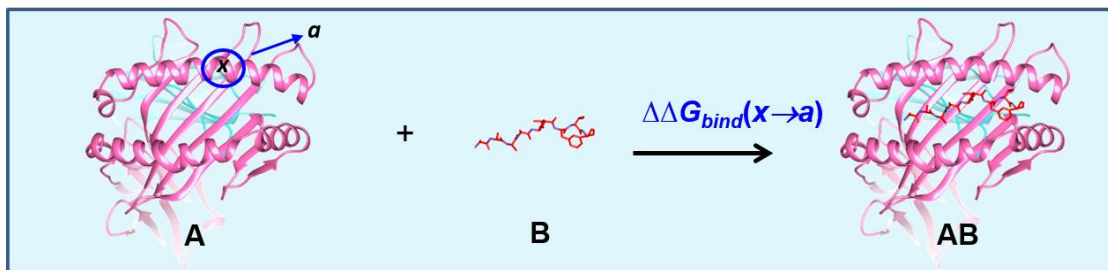
Figure 10. Definition of binding free energy difference in alanine scanning mutagenesis process, whereas A) HLA receptor with wild-type (x) and mutate (a), B) peptide and C) complex of HLA ($x \rightarrow a$) and peptide.

Residue-Specific Free Energy Contribution. Free energy contributions of individual residues to protein−protein binding are typically calculated by binding free energy difference ($\Delta\Delta G_{bind}^{x \rightarrow a}$) upon alanine mutation, which is defined as

$$\Delta\Delta G_{bind}^{x \rightarrow a} = \Delta G_{bind}^{x} - \Delta G_{bind}^{a}, \tag{23}$$

where

$$\Delta\Delta G_{bind}^{x} = \Delta G(A(x) - B), \tag{24}$$

$$\Delta\Delta G_{bind}^{x} = \Delta G_{gas}(A(x) - B) + \Delta G_{sol}(A(x) - B), \tag{25}$$

similar as

$$\Delta\Delta G_{bind}^{a} = \Delta G(A(a) - B), \tag{26}$$

$$\Delta\Delta G_{bind}^{a} = \Delta G_{gas}(A(a) - B) + \Delta G_{sol}(A(a) - B). \tag{27}$$

The individual proteins of the protein−protein complex are represented by "A" and "B", respectively. The superscript "x" specifies residue of protein A at the protein−protein interface, while "a" specifies alanine mutation which replaces to x residue. $\Delta G_{bind}(x)$ and $\Delta G_{bind}(a)$ are the binding free energy of the wild type and mutant

type of protein A. $\Delta G(A(x)-B)$ and $\Delta G(A(a)-B)$ represent, respectively, the binding free energy of the wide type and mutation type complex. Further explanation of eq.28 can write in below:

$$\Delta\Delta G_{bind}^{x\to a} = \Delta G_{gas}(A(a)-B) + \Delta G_{sol}(A(a)-B),\tag{28}$$

$$\Delta\Delta G_{bind}^{x\to a} = \Delta G_{gas}(A(a)-B) + \Delta G_{sol}(A(a)-B) - \left[\Delta G_{gas}(A(x)-B) + \Delta G_{sol}(A(x)-B)\right],\tag{29}$$

$$\Delta\Delta G_{bind}^{x\to a} = \Delta\Delta G_{gas}^{x\to a} + \Delta\Delta G_{sol}^{x\to a}.\tag{30}$$

The above equation, we can make the following simplification,

$$\Delta\Delta G_{gas}^{x\to a} = \Delta G_{gas}(A(a)-B) - \Delta G_{gas}(A(x)-B),\tag{31}$$

$$\Delta\Delta G_{gas}^{x\to a} \approx \Delta G_{gas}(a-B) - \Delta G_{gas}(x-B).\tag{32}$$

The difference proteins in gas phase interaction energies between the wild type $A(x)-B$ and the mutant $A(a)-B$ complexes are equal to residue $x-B$ and residue $a-B$. Environment changing to solvent then the free energy of solvation part in eq.30 is followed by:

$$\Delta G_{sol}(A(x)-B) = G_{sol}(A(x)\wedge B) - G_{sol}(A(x)) - G_{sol}(B)\tag{33}$$

where $G_{sol}(A(x)\wedge B)$ is the solvation energy of the A(x)−B complex and as similar to

$$\Delta G_{sol}(A(a)-B) = G_{sol}(A(a)\wedge B) - G_{sol}(A(a)) - G_{sol}(B).\tag{34}$$

That are

$$\Delta\Delta G_{sol}^{x\to a} = \Delta G_{sol}(A(a)-B) - \Delta G_{sol}(A(x)-B)\tag{35}$$

$$\Delta\Delta G_{sol}^{x\to a} = G_{sol}(A(a)\wedge B) - G_{sol}(A(x)\wedge B) - \left[G_{sol}(A(a)) - G_{sol}(A(x))\right].\tag{36}$$

Influence of the Internal Dielectric Constant Value.

The internal dielectric constant (ε_{int}) is an electric field of solute without explicit treating. Based on mutation of amino acid residues, side chain re-orientates and results to dipolar reorganization arising from conformational changing. Twenty amino acids is classified into basic four groups upon side chain properties.

i) polar and neutral: aspargine, glutamine, cysteine, tyrosine, serine, and threonine

ii) nonpolar and neutral: valine, alanine, leucine, isoleucine, phenylalanine, proline, glycine, methionine, and tryptophan

iii) basic and charged: lysine, arginine, and histidine

iv) acidic and charged: aspartic acid and glutamic acid

In case of histidine, it can be uncharged or charged depend on pH.
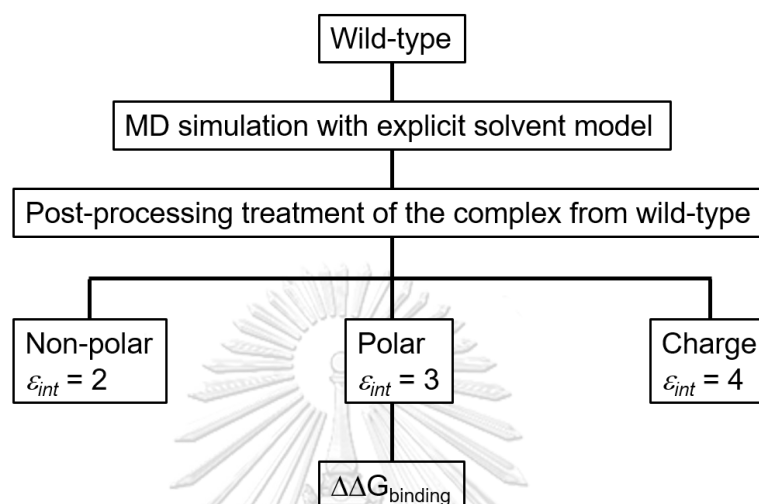
Figure 11. Schematic representation for computational alanine screening mutagenesis [91]

Various ε_{int} explains different degree of relaxation, when different types of amino acids are mutated for alanine; the stronger the interactions these amino acids establish, the more extensive the relaxation should be, and the greater the internal dielectric constant value must be to mimic these effects.

2.5 3D-RISM model

Reference interaction site model (RISM) concerns the theory of molecular liquids to handle molecular systems in solution. RISM analysis of the free energy diagram reports the solvation structure and thermodynamics in the statistical-mechanical ensemble. A solving the RISM integral equations for correlation functions is yielded solvation thermodynamics in a single form. Solvent distributions surround solute molecule under 3D maps is provided 3D-RISM theory. An important component of 3D-RISM, the closure relations are approximated by Kovalenko and Hirata (KH) [92]. The 3D-RISM-KH theory represents features of solvation such as electrostatic and non-polar with hydrogen bonding, hydrophobicity, steric effects and electrochemical.

This method has been coupled with ab-initio quantum theory in a self-consistent description of electronic structure for optimizing the geometry of solution.

The first principles foundation of 3D-RISM integral equation was derived from the molecular Ornstein - Zernike (MOZ) equation. The total correlation function $h(r_{12}, \Omega_1, \Omega_2)$ can be decomposed into direct correlation function $c(r_{12}, \Omega_1, \Omega_2)$ and indirect correlation function (eq.37)

$$h(r_{12}, \Omega_1, \Omega_2) = c(r_{12}, \Omega_1, \Omega_2) + \rho \int c(r_{13}, \Omega_1, \Omega_3) h(r_{23}, \Omega_3, \Omega_2) dr_3 d\Omega_3 \qquad (37)$$

where r_{12} is the distance, Ω_1 and Ω_2 are the angular orientation of particles 1 and 2. The definition of total correlation function is $h(r_{12}, \Omega_1, \Omega_2) = g(r_{12}, \Omega_1, \Omega_2) - 1$. The MOZ equation can be written in terms of an infinite series in which the direct correlation function expanded in the power of ρ (eq.38)

$$h(r_{12}, \Omega_1, \Omega_2) = c(r_{12}, \Omega_1, \Omega_2) + \rho \int c(r_{13}, \Omega_1, \Omega_3) c(r_{32}, \Omega_3, \Omega_2) dr_3 d\Omega_3 +$$
$$\rho^2 \int c(r_{13}, \Omega_1, \Omega_3) c(r_{34}, \Omega_3, \Omega_4) c(r_{42}, \Omega_4, \Omega_2) dr_3 d\Omega_3 dr_4 d\Omega_4 + ... \qquad (38)$$

In case of homogenous and isotropic, the molecule does not depend on its orientation, the equation can be written in real space, r (eq.39).

$$h(r) = c(r) + \rho \int c(|r - r'|) h(r') dr' \qquad (39)$$

Using Fourier transform, this equation is simplified in reciprocal space (k space) (eq.40).

$$\hat{h}(k) = \frac{\hat{c}(k)}{1 - \rho \hat{c}(k)} \qquad (40)$$

However, the solution of MOZ equation cannot achieve since containing two unknown functions. Hence, the closure relations have been proposed to related two functions such as Hypernetted-chain (HNC), Percuc-Yevick (PY), Kovalenko and

Hirata (KH) and Partial series expansion of order-*n* (PSE-*n*) etc. The general closure relation to the MOZ equation is (eq.41)

$$g(r_{12}, \Omega_1, \Omega_2) = \exp\left[-\frac{u(r)}{k_B T} + h(r_{12}, \Omega_1, \Omega_2) - c(r_{12}, \Omega_1, \Omega_2) + b(r_{12}, \Omega_1, \Omega_2) \right] \quad (41)$$

where $u(r)$ is interaction potential between two particles in a system, k_B is Boltzmann constant at temperature *T* and $b(r_{12}, \Omega_1, \Omega_2)$ is bridge function.

To apply with polyatomic molecules, the reference interaction-site model (RISM) has been proposed by Chandler and Andersen through averaging the angular coordinates of molecules and fixing the distance between a pair of interaction sites.[82] The six-coordinates of solvent are generalized to one dimension called as 1D-RISM. The distribution function of solvent around solute molecule can be calculated by three-dimensional RISM (3D-RISM) by taking statistical average the angular coordinate of solvent. Similar to MOZ equation, the 1D-RISM and 3D-RISM equations need the closure relation to solve until self-consistency.

The procedure of 3D-RISM calculation, the site correlation functions between solvent molecules based on specified atomic interaction potential of water model was calculated as 1D-RISM. Then, the correlation functions, or susceptibilities, are utilized for solvent-solute interaction using 3D-RISM equations coupled with closure relations. The general idea of the solvation structure is the probability density of the solvent contribution around the solute molecule in 3D space. This density distribution can be defined as $\rho_\gamma g_\gamma(r)$, a product of average density of the solvent bulk (ρ_γ) times the 3D distribution function at the *r* position of solvent molecules around the solute molecule ($g_\gamma(r)$).

$$g_\gamma(r) = h_\gamma(r) + 1 \quad \text{for} \quad d(r) > 0 \quad (42)$$

where $g_\gamma(r)$ is the pair correlation function, $h_\gamma(r)$ is the total correlation function of solvent site γ in 3D space, and $d(r) = -\dfrac{u(r)}{k_B T} + h(r) - c(r)$. The free energy of solvation can be compute from the distribution function in eq.43.

$$\Delta\mu_{KH} = \rho^v k_B T \sum_\gamma \int dr \left[\frac{1}{2} h_\gamma^{uv}(r)^2 \Theta\left(-h_\gamma^{uv}(r)\right) - c_\gamma^{uv}(r) - \frac{1}{2} h_\gamma^{uv}(r) c_\gamma^{uv}(r) \right] \tag{43}$$

where u and v denote for solute and solvent molecule, γ labels atom types of model and Θ is the Heaviside step function.

2.6 Pairwise residue contact analysis

The distance between all possible amino acid residue pairs of 3D protein structure is measured and reported in 2D matrix graph. For pair residues *a* and *b* is given 1-*ab* element within distant cutoff. The contact analysis has variety distant definition such as C_α-C_α atoms with 6-12 Å of cutoff, C_β-C_β atoms with 6-12 Å of cutoff and distance between side-chains center of gravity. The major advantages of pairwise residue contact analysis are i) binding site prediction and ii) protein folding analysis by secondary structure.
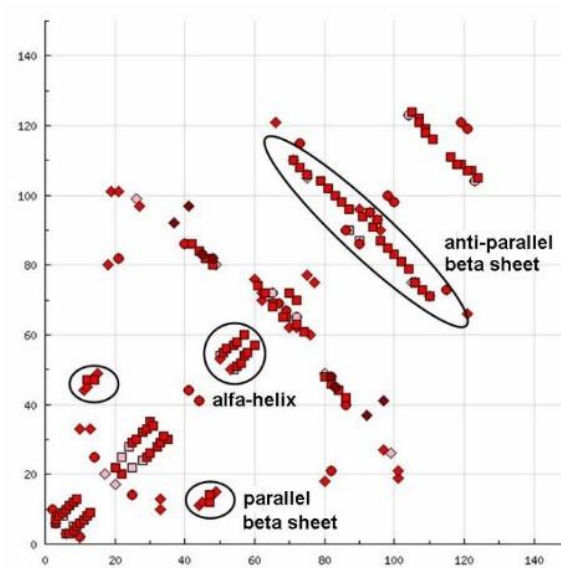
Figure 12. Secondary structure elements are swapped parallel and anti-parallel sheets Three classes of hydrogen bonds are distinguished by color-coding; short (distance smaller than 2.5 Å between donor and acceptor), intermediate (between 2.5 Å and 3.2 Å) and long hydrogen bonds (greater than 3.2 Å).

(https://en.wikipedia.org/wiki/Protein_contact_map)

Secondary structure elements in HB plot

Base on protein folding by hydrogen bonding interaction, the plot of pairwise residues is displayed in Figure 12. The secondary structure patterns are characterized by:

1) Helices are presented as adjacent to the diagonal.
2) Parallel beta sheets are shown as parallel to the diagonal
3) Antiparallel beta sheets are identified by cross-diagonal.
4) Loops are defied as breaks in the diagonal between the cross-diagonal beta-sheet motifs.

2.7 B-factor

Based on X-ray scattering, protein crystal are scattered and evaluated atomic positions. The static and dynamic components of crystal are uncertain positions. Cause of static disorder may be temperature factor also called B-factor. In general, B-factor value has confidence of < 30 Å^2, whereas the value has disorder of > 60 Å^2.

In computational calculation, the B-factor can imply the atomic flexibility, because it related to the mean square displacement of atoms $\langle u^2 \rangle$:

$$B = 8\pi^2 \langle u^2 \rangle . \tag{44}$$

$$\langle u^2 \rangle = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((a_{ix} - b_{ix})^2 + (a_{iy} - b_{iy})^2 + (a_{iz} - b_{iz})^2\right)} \tag{45}$$

where, n is atoms in protein. The mean square displacement of atoms is calculated by three coordinates (x, y, z). B-factor values are measured during refinement process to improve between structural model and experimental diffraction. B-factor for each atom is recorded in PDB file, as show in Figure 13.

					coordinate			B-factor		
					x	y	z			
ATOM	1	N	GLY A	1	16.224	0.563	-41.764	1.00	39.15	N
ATOM	2	CA	GLY A	1	14.870	0.580	-42.306	1.00	38.78	C
ATOM	3	C	GLY A	1	13.839	1.114	-41.332	1.00	41.73	C
ATOM	4	O	GLY A	1	13.160	2.105	-41.626	1.00	41.62	O
ATOM	5	N	SER A	2	13.709	0.448	-40.170	1.00	36.96	N
ATOM	6	CA	SER A	2	12.773	0.833	-39.117	1.00	36.21	C
ATOM	7	C	SER A	2	13.472	1.790	-38.156	1.00	38.91	C
ATOM	8	O	SER A	2	14.493	1.433	-37.557	1.00	38.75	O
ATOM	9	CB	SER A	2	12.241	-0.401	-38.393	1.00	39.72	C
ATOM	10	OG	SER A	2	11.237	-0.060	-37.453	1.00	48.85	O
ATOM	11	N	HIS A	3	12.945	3.023	-38.049	1.00	34.23	N
ATOM	12	CA	HIS A	3	13.518	4.075	-37.209	1.00	33.54	C

Figure 13. A PDB file shows atomic coordination and B-factor values.

2.8 Root-mean-square deviation (RMSD)

Root-mean-square deviation (RMSD) is a measure of the different coordinates between MD structures at time respected to initial structure. The RMSD of atoms is calculated by three coordinates (x, y, z).

$$\langle u^2 \rangle = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((a_{ix} - b_{ix})^2 + (a_{iy} - b_{iy})^2 + (a_{iz} - b_{iz})^2\right)} \tag{46}$$

whereas $\langle u^2 \rangle$ is mean square displacement of atoms, n is atoms of protein, a and b are coordinate at points of starting and MD production.

CHAPTER III
RESEARCH METHODOLOGY

3.1 Part I: Behçet's disease (BD)

3.1.1 Structure preparation

The starting structures of three of the HLA class I alleles were taken from the Protein Data Bank, being entries 1E27 [15] for B*51:01 with the HIV-1 epitope at 2.20 Å, 1X7Q [93] for A*11:01 with the SARS nucleocapsid at 1.45 Å, and 1A9E [94] for B*35:01 with the EVB peptide at 2.5 Å. That for A*26:01 was built from homology modeling using the above A*11:01/SARS nucleocapsid structure as a template and the amino acid sequence of A*26:01 from GenBank (accession no. AAA03720) [95], which has 94.3 and 96.6% amino acid identity and similarity to A*11:01, respectively. The HLA/MICA-TM complexes were constructed by changing the original peptide within the X-ray structure to be consistent with the nine amino acids of the antigenic MICA-TM peptide (AAAAAIFVI) using the align sequence profiles module implemented in the Discovery studio 2.5 (Accelrys, Inc.).

Table 4. Sequence alignments among four HLAs are reported as the percentages of identity and similarity.

% iden. \ % sim.	B*51:01	B*35:01	A*26:01	A*11:01
B*51:01	-	94.9	84.1	85.5
B*35:01	96.4	-	84.5	86.3
A*26:01	87.7	88.4	-	92.4
A*11:01	89.5	89.9	94.9	-

3.1.2 Molecular dynamics (MD) simulation

The four studied HLA/MICA-TM complexes were investigated by molecular dynamics (MD) simulations using the AMBER 10 software package [96] with the ff03 force field. The missing atoms were added using the LEaP module [87]. The protonation state of all possible charged residues (arginine, lysine, histidine,

aspartate and glutamate) in HLA-allele complexes was assigned at pH 7.0 by PROPKA server [97]. The total charges with negative value of the HLA/MICA-TM complexes were randomly neutralized by Na^+ counterions (2, 1, 1 and 2 ions for the B*51:01, B*35:01, A*26:01 and A*11:01 systems, respectively). Afterwards, the individual complex was solvated by TIP3P [98] water molecules leading to approximately 65,000 atoms in total. The dimensions of the simulation box used for all the systems were 86 $\times$ 90 $\times$ 88 $Å^3$. The periodic boundary condition with the NPT ensemble and a simulation time step of 2 fs was used. All energy minimizations and MD simulations were performed using the SANDER module of AMBER 10. All bonds and angles concerned to hydrogen atoms were constrained by algorithm of SHAKE [99]. The long-range electrostatic interactions were treated by particle mesh Ewald method and the non-bonded interactions with a cutoff distance of 12 Å were considered [100]. All MD simulations were run with a 12 Å residue-based cutoff for non-bonded interactions and the particle mesh Ewald method was applied for an adequate treatment of long-range electrostatic interactions [101]. Each system was subjected to the four stages of the restrained MD simulations at 298 K with force constants of 10, 7.5, 5 and 2.5 kcal·mol^{-1}·$Å^2$ for 500 ps in each stage accordingly. These subsequent steps could allow the peptide to adapt its geometry and orientation from the initial model to fit better within the peptide binding groove. Then, the constraints were completely removed and fully unrestrained MD simulations were performed until 50 ns. The convergences of energies, temperature, and global root mean-square displacement (RMSD) were used to verify the stability of the systems. The MD trajectories were collected every 0.2 ps from the production phase for further analysis.

3.2 Part II: Systemic sclerosis disease (SSc)

3.2.1 HLA-DRs/Top1 complex preparation

The 3D structures of the HLA-DRs/Top1 in this study are not available. Amino acid sequences of HLA-DRB1*08:02 [102], HLA-DRB1*11:01 [103], HLA-DRB1*11:04 [104] and HLA-DRB5*01:02 [105] proteins were obtained from the National Center for Biotechnology Information (NCBI) database

(http://www.ncbi.nlm.nih.gov/), seeing in Table 5. With over 85% identity, the derived crystal structure from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/), HLA-DR3 carrying CLIP 87-101 peptide (pdb code: 1A6A [5]) was used as the template for HLA-DRB1*08:02, HLA-DRB1*11:01 and HLA-DRB1*11:04, whereas HLA-DRB5*01:01 binding with myelin 86-105 peptide (pdb code: 1FV1 [106]) was chosen to construct HLA-DRB5*01:02. Few amino acids in HLA-DR templates were virtually mutated to four HLA-DR/Top1 complexes in this work using the Discovery studio 2.5 (Accelrys, Inc.) executed by the align sequence module. Binding between residues of HLA-DR and Top1 spanning RIANFKIEPPGLFRGRGNHP (349–368) sequence is determined by conservation of key anchor sites [107]. The PDB entry 1AQD [108] for HLA-DRB1*01:01 carrying an endogenous peptide with 15 amino acids in length was established by changing the original peptide to Top1. HLA-DRs in their free form were obtained by removing the antigenic peptide from the binding groove. Both groups of HLA-DR structures, with and without Top1 peptide, were modeled for MD simulations.

Table 5. HLA-DR sequences are constructed from identical protein templates and used for MD calculation.

HLA	X-ray structure (PDB code)	X-ray template (PDB code)	Sequence (Accession no.)	% iden. from BLAST search
DRB1*08:02	-	1A6A	Q30134	85
DRB1*11:01	-	1A6A	CAM84026	86
DRB1*11:04	-	1A6A	CAJ01187	89
DRB5*01:02	-	1FV1	BAO73173	97
DRB1*01:01	1AQD	-	-	-

Table 6. Sequence alignments among five HLA-DRs are reported as the percentages of identity and similarity.

% iden. \ % sim.	DRB1*08:02	DRB1*11:01	DRB1*11:04	DRB5*01:02	DRB1*01:01
DRB1*08:02	-	97.3	96.8	89.3	90.9
DRB1*11:01	98.9	-	99.5	89.3	90.9
DRB1*11:04	98.4	99.5	-	88.8	90.4
DRB5*01:02	95.7	95.7	95.2	-	92.0
DRB1*01:01	95.7	95.7	95.2	97.9	-

3.2.2 Molecular dynamics (MD) simulation

Assisted Model Building with Energy Refinement (AMBER) version 14 with LEaP module [109] was employed to add all missing atoms within the starting structures based on the ff03.r1 force field. This version contains the charge set for –COO$^-$ and –NH$_3^+$ terminal groups of proteins updated from original ff03 for more accuracy [87]. Added hydrogen atoms were minimized, while the others were frozen. H++ web-prediction of protonation (http://biophysics.cs.vt.edu/H++) was applied to the protein complexes kept at pH 7.0 which were then by randomly neutralized by 7, 7, 7, 12 and 4 sodium ions for HLA-DRB1*08:02, HLA-DRB1*11:01, HLA-DRB1*11:04, HLA-DRB5*01:02 and HLA-DRB1*01:01; respectively. HLA-DRs with and without Top1 peptide were subsequently solvated by TIP3P [110] water molecules in an orthogonal cage of $97.87 \times 109.36 \times 78.92$ Å^3. The isothermal-isobaric (NPT) ensemble with constrained number of atoms (N), pressure (P) and temperature (T) was applied in a periodic boundary. The SANDER module of AMBER 14 was used to minimize all water molecules and protein complexes, respectively. The systems were heated to 298 K for 100 ps under constrained atoms of HLA-DR/Top1 by a weak force of 60.0 kcal·mol^{-1}·Å^2. The SHAKE algorithm constrained all bonds and angles involving hydrogen atoms with a time step of 2 fs [111]. Long range electrostatic interactions within a cutoff radius of 12 Å were calculated using the particle mesh Ewald (PME) method [101], while short range non-bonded interactions were evaluated with 12 Å atom-based cutoff for Lennard-Jones potential [110]. MD simulation using the Verlet algorithm[112] was performed for all HLA-DR/Top1 complexes and the snapshots were stored every 0.2 ps during the operation of 30 ns. The systems of HLA-DR free form have been setup for MD procedure analogously to their complexed structures. The

pmemd.cuda module of AMBER was performed in all simulations. For analysis, the MD trajectories in production phase were collected for analysis of complex stability, binding free energy, hydrogen bonding and water distribution.

3.3 Part II: TCR/HLAp recognition

3.3.1 Complex preparation

An adaptive immune resistance to HIV by TCR/HLAp formation was downloaded from the PDB database under the accession ID: 4MJI [27]. In this X-ray structure, 8-mer (TAFTIPSI) HIV peptide was complexed between TCR and HLA-B*51:01 molecules. The HLA-B*51:01 in complex with 9-mer (LPPVVAKEI) HIV epitope (PDB code: 1E27) [15] was used as a template for the Behçet's disease (BD) model. For BD system, the 9-mer HIV epitope was mutated to MICA peptide sequence (AAAAAIFVI) using the Discovery studio 3.0 (Accelrys, Inc.) program. Then, a ternary complex was constructed by superimposition of HLA-HIV9p and HLA-MICA into TCR/HLA-HIV8p template. In this study, we focus on three TCR/HLAp complexes: TCR/HLA-HIV8p, TCR/HLA-HIV9p for HIV infections and TCR/HLA-MICA for BD diseases. To reduce the simulation time, the constant region of both α/β chains, which was the distal end, was removvde in all complexes.

3.3.2 Molecular dynamics (MD) simulations

The LEaP module [109] in Assisted Model Building with Energy Refinement (AMBER) version 14 was used to add all missing atoms for the starting structures with the ff03.r1 force field [87]. The minimization was applied after adding hydrogen atoms. Protein protonation was predicted by the H++ website (http://biophysics.cs.vt.edu/H++). Protein complexes were soaked in a periodic box of TIP3P water model with the size of 76×68×138 Å^3. Counter ions were added to neutralize the protein with 12 sodium ions in all system. The SANDER module of AMBER 14 was applied to minimize the systems by two procedures. First, water molecules and ions were allowed to move freely, while the proteins were kept under 500 kcal·mol^{-1}·Å^2 constraint. Second, the whole system was relaxed until convergence. The system was heated to 298 K within 0.1 ns with restraint on the protein at 50

kcal·mol^{-1}·Å^2. The NPT ensemble condition was applied on the solute atoms with a periodic boundary condition. To restrain covalent bonds and angles involving hydrogen atoms, SHAKE algorithm was employed [111]. The particle mesh Ewald (PME) method [101] and Lennard-Jones potential [110] were used to calculate the long range electrostatic interactions and short range non-bonded interactions, respectively. The Verlet algorithm [112] for time integration was applied and snapshots were collected every 0.2 ps during simulation. The fully MD simulations were performed until 200 ns. The trajectories extracted from 125-200 ns were accumulated for further analysis.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Part I: Behçet's disease

4.1.1 Stability of the HLA/MICA-TM complexes

To study the structural stability of the four MD simulations, the RMSDs for all atoms of the four HLA alleles (B*51:01, B*35:01, A*26:01 and A*11:01 alleles) complexed with the MICA-TM peptide compared with those of the starting structures were monitored along 50 ns of simulation time using the PTRAJ module implemented in AMBER10 package [113]. The RMSD fluctuations of each complex and its structural components (binding groove and β_2-microglobulin) and MICA-TM peptide, were plotted in Figure 14. The RMSD values quickly increased until ~5 ns and then reached a plateau except for the HLA-B*51:01/MICA-TM complex where the RMSD value continuously increased over the first ~15 ns. The MICA-TM in HLA-B*51:01 system had a low fluctuation around 1 Å, and suddenly at 23 ns, it jumped up to 2 Å until the end simulation. The RMSD increasing caused from the non-polar of MICA-TM peptide change orientation to hydrophobic region within pocket. Although the whole HLA/MICA-TM complex for the two BD-associated HLAs (B*51:01 and A*26:01) fluctuated a great deal, their binding groove and the incoming MICA-TM peptide demonstrated a rather low level of fluctuation. All systems seemed to reach equilibrium after 25 ns with ~1 Å of RMSD fluctuation, and so the MD trajectories from 25 to 50 ns (the production phase) were used for analysis.
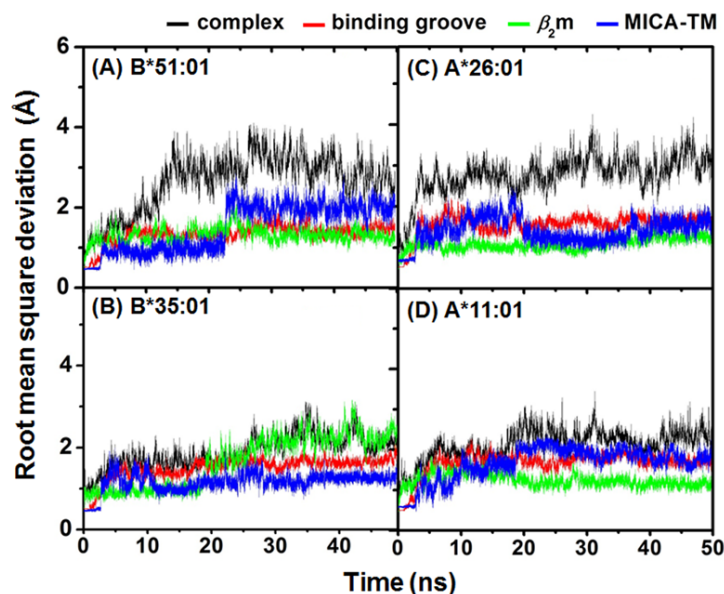
Figure 14. Stability of the HLA/MICA-TM complexes. Root-mean square displacements (RMSDs) of all atoms relative to those of the initial structure for the HLA/MICA-TM complex, peptide binding groove, β_2-microglobulin and MICA-TM peptide in the (A) B*51:01, (B) B*35:01, (C) A*26:01 and (D) A*11:01 HLA alleles bound to the MICA-TM peptide.

4.1.2 Regional flexibility

The backbone flexibility of the HLA-B*51:01, B*35:01, A*26:01 and A*11:01 complexes with MICA-TM were investigated by B-factor calculation over the last 25 ns trajectories (Figure 14). Note that high flexibility of protein was displayed in red and *vice versa* in blue. Among the four systems, the HLA-A*11:01 allele showed the highest degree of protein flexibility in the region of peptide binding groove and in particular at the S7–S9 sub-sites (defined in Figure 15D), whereas the other HLA alleles were likely to have similar degree of flexibility (Figure 16). This might be due to the strong intramolecular interactions of either the α1 or α2 domains and the intermolecular hydrogen bonds (H-bonds) between the C-terminal of the MICA-TM peptide and the binding residues of the three HLAs (B*51:01, B*35:01 and A*26:01 alleles), which is discussed in a later section.
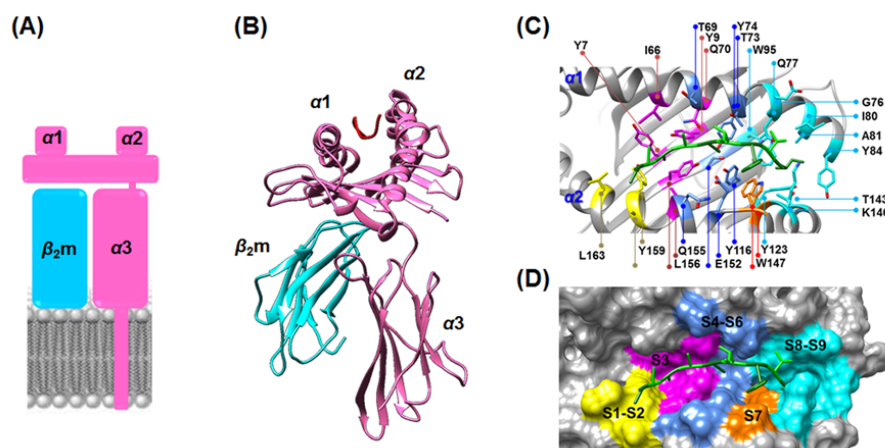
Figure 15. Structural basis of HLA class I. (A) Schematic model of HLA buried in the transmembrane. (B) HLA (pink) contains the $\alpha 1$ and $\alpha 2$ subdomains that contribute to the peptide binding groove, while $\alpha 3$ is the C-terminal domain in complex with $\beta 2$-microgluobulin (β_2m) as a noncovalently supported protein (cyan). (C) Ribbon and (D) van der Waals surface representations of the MICA-TM nonapeptide (green stick model) occupied in the peptide binding sub-sites (S1–S9, shaded by different colors) of HLA-B*51:01.
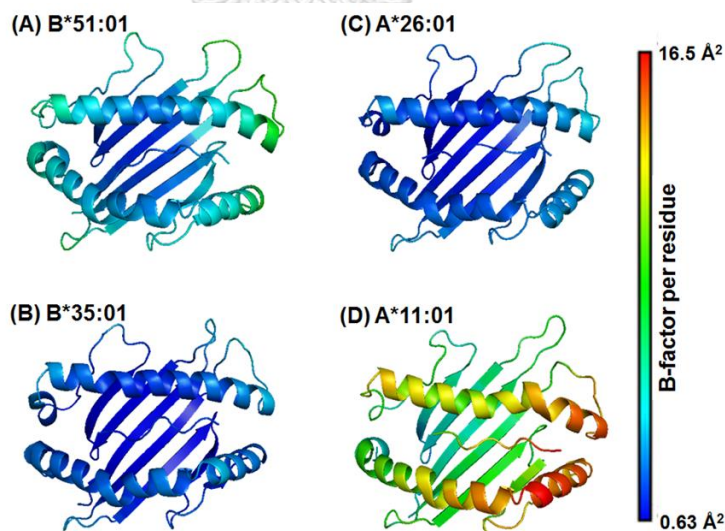
Figure 16. Structural flexibilities of the HLA alleles bound with the MICA-TM peptide. Structural flexibilities were evaluated by B-factor. The ribbon color changes from blue (rigid) to red (flexible) to represent a low to high protein flexibility. Note that for clarity only the binding groove structure and the MICA-TM peptide are shown.

4.1.3 Per-residue decomposition (DC) energy

The decomposition (DC) energies were calculated and used to scan for potentially important residues for binding [114]. In order to seek the fingerprint of the HLA/MICA-TM interactions, the interaction energy between each HLA residue and the MICA-TM peptide and *vice versa* was calculated over 100 snapshots of the production phase. The obtained DC energies of the HLA and MICA-peptide residues are plotted in Figure 17Figure 18, respectively. The HLA/MICA-TM interactions mostly occurred on the *α1* and *α2* helixes with additionally some residues of *β*-strands (Figure 17). Using the criterion of a total DC energy < -0.5 kcal/mol as an important residue, then 22, 18, 16 and 14 potentially important residues of the HLA-B*51:01, B*35:01, A*26:01 and A*11:01 alleles (Table S1 in appendix), respectively.
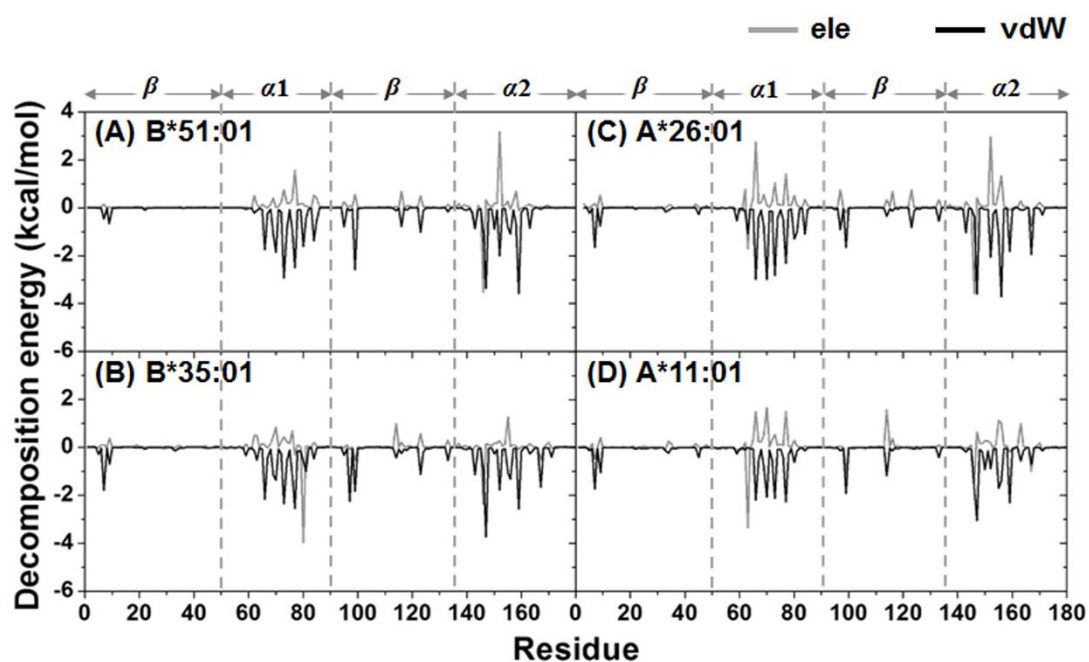
Figure 17. Decomposition energy per HLA residue fingerprint plots. The HLA contribution to the MICA-TM binding is shown in terms of the electrostatic (ele) and van der Waals (vdW) interactions.

van der Waal (vdW) interactions were found to play an important role in the complex, where the magnitude of the MICA-TM peptide binding with the BD-associated HLA alleles (B*51:01 and A*26:01) was greater than that for the corresponding matched

non-BD-associated HLA alleles (B*35:01 and A*11:01). For the nonapeptide, as expected the non-polar residues interacted with HLA through vdW interactions (Figure 18). Based on the definition of a total DC energy of < -3 kcal/mol for strong binding, there were four peptide residues that firmly bound to HLA-B*51:01 (P3, P5, P8 and P9) and A*26:01 (P2 and P7–P9), while only two and three residues were found in B*35:01 (P7 and P8) and A*11:01 (P2, P7 and P8), respectively. Note that the P9 and P2 or its adjacent residue are known as the anchor for the peptide accommodation at the binding groove of HLA class I [115, 116]. Loss of the P9 contribution may lead to an unbinding recognition of the non-BD-associated HLAs towards the MICA-TM peptide. More details of the interaction and orientation of this peptide at the binding groove of the four HLA alleles investigated are discussed in the hydrogen bond pattern section.
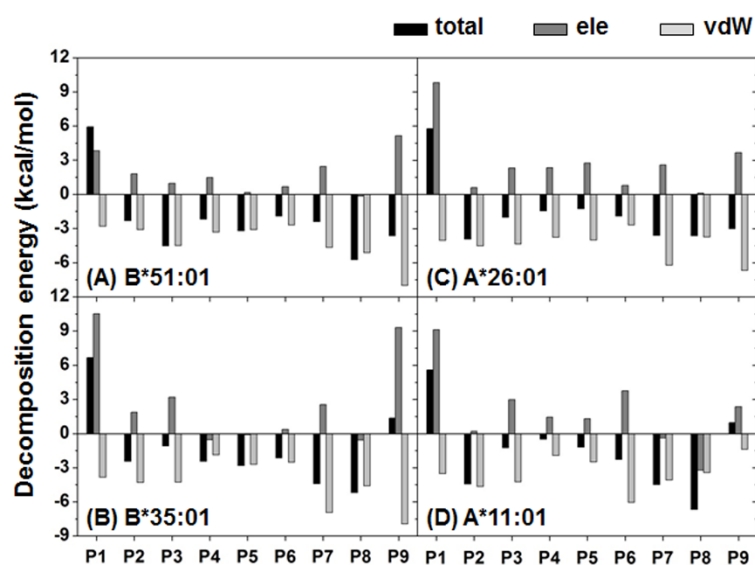
Figure 18. Averaged decomposition energy contributions in HLA binding to MICA-TM. Per-residue decomposition energies and the energy components in terms of the electrostatic (ele) and van der Waals (vdW) interactions for the P1–P9 residues of MICA-TM.

4.1.4 H-bond patterns between HLA and the MICA-TM nonapeptide

To investigate the intermolecular H-bond interactions between the HLA protein and the incoming short MICA-TM nonapeptide, the number of H-bonds was evaluated

using the two acceptance criteria of (i) a distance between the proton donor (D) and acceptor (A) atoms of D−A ≤ 3.5 Å; and (ii) an angle of D−H···A > 120°, as previously reported [114, 117]. The obtained results were compared between the two paired BD-associated and non-associated HLA alleles in each locus (A or B), and are shown in Figure 19. The peptide binding sub-sites (S1–S9) were classified by the peptide contact position in the binding groove of HLA, as seen in the HLA-B*51:01/peptide complex (Figure 15D). The last snapshot of each complex was used to represent the protein-protein H-bond formation detected from the MD simulations (see in Figure 20).
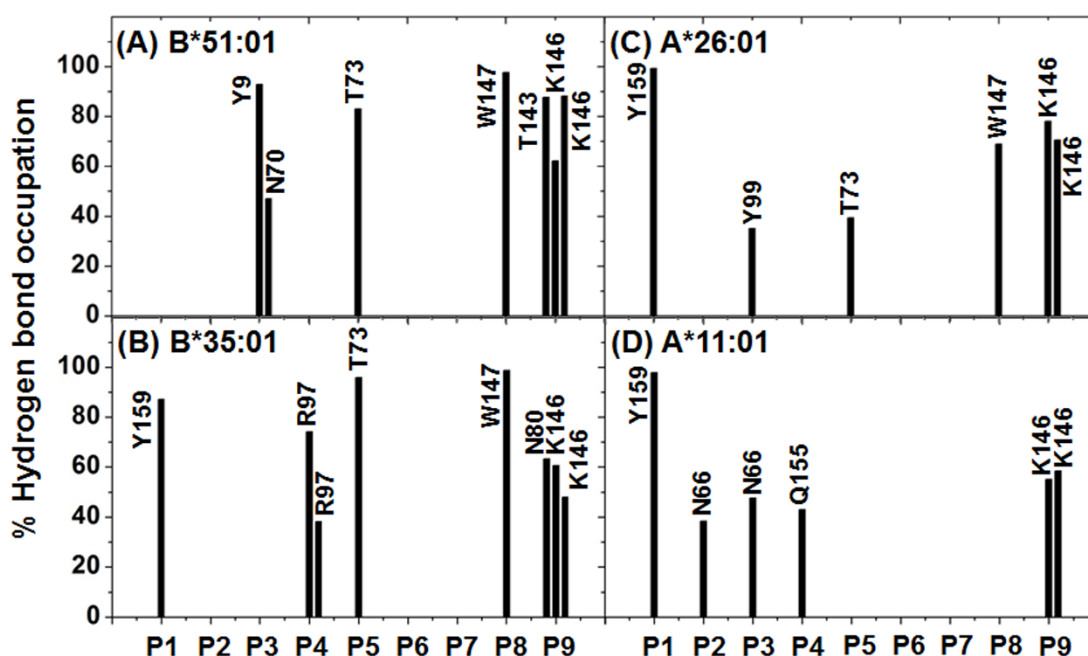
Figure 19. Hydrogen bond interactions. The percentage occupancy of H-bonds averaged over the last 25 ns of simulation time between the nine residues (P1–P9) of the MICA-TM peptide and the HLA residues for the four complexes.

As shown in Figure 15D, the peptide binding groove of HLA class I is constructed by at least five sub-pockets with a large groove volume (S1 and S2 yellow; S3 magenta; S4–S6 dark blue; S7 orange; S8 and S9 light blue), and it is able to support a variety of incoming short peptides in the binding step. Moreover, the binding groove of the HLA protein provides a hydrophobic cavity to support the nine spanning residues

(AAAAAIFVI) of the MICA-TM peptide. Due to the nonpolar nature of this peptide, H-bond formation with the HLA protein was expected through the peptide backbones. In Fig. 5A and B, both HLA-B alleles similarly stabilized the P9 residue of the MICA-TM peptide by the formation of three H-bonds which two of them are the salt bridge interactions between the C-terminal carboxylate group and the ammonium group of K146. The C-terminal residue (P9) could bind significantly stronger with the BD-associated HLA allele than the non-associated allele (~60–90% H-bond occupation compared to only ~50–70% in the non-associated allele), which could be due to the different residue 80 on the α1 helix (I80 in HLA-B*51:01 and N80 in HLA-B*35:01). Through the strong H-bond (70% occupancy) with the amide group of N80 (Figure 19B), this polar residue at the S9 sub-site of the peptide induced the C-terminal to change its orientation and move closer to the α1 helix of the non-associated BD HLA-B*35:01 (Figure 20) with the consequence of weakened H-bonds with the α2 helix from the loss of interaction with T143 and a decreased interaction with K146. With respect to the N-terminus of the MICA-TM peptide, the Y159 residue on the α2 helix of the HLA-B*35:01 allele strongly interacted with the P1 residue (~90%), whilst this end of the peptide had no interaction with any HLA-B*51:01 residues but instead the P3 residue was stabilized by two α1-helix residues, Y9 and N70, with ~90 and ~50% H-bond occupations, respectively (Figure 19A and Figure 20 in supporting information). The rearrangement of the H-bonding network at the N-terminal was previously observed in the octamer and nonamer peptides binding to HLA-B*51:01 [15]. In addition to the protein-protein interactions at the two peptide ends, a strong H-bond was formed between the backbone of the P5 Ala residue and the side chain of the T73 residue on α1 helix in the P4–P6 pocket in both HLA-B alleles (> 80%, Fig. 5A and B). This is congruent with the previous observation that the P5 residue of HIV epitopes (Ile, Val and Pro) was found as an unexpected anchor deeply pointing into the sub-pocket of the binding groove of HLA-B*51:01 [15]. In addition, these two HLA-B alleles had a strong H-bond between the P8 backbone of the MICA-TM nonapeptide and the W147 indole ring on the α2 helix.
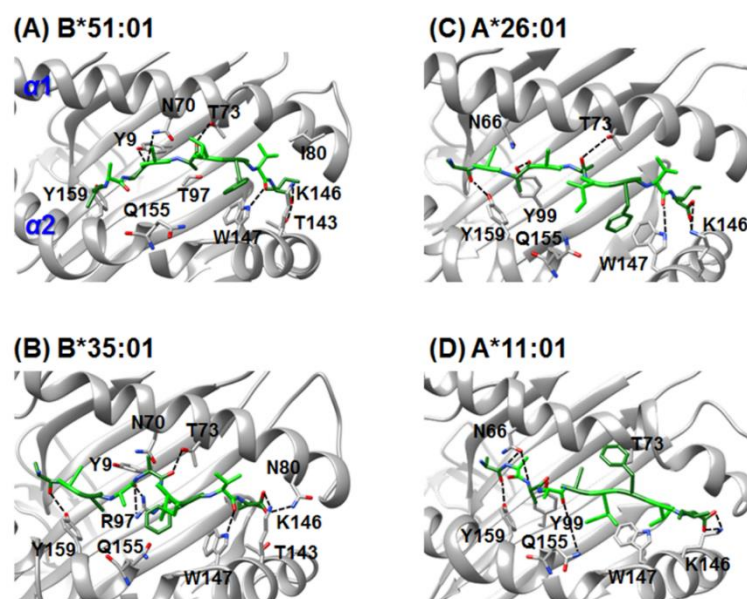
Figure 20. Hydrogen bond interactions (dashed line) in the HLA/MICA-TM complexes. The MICA-TM peptide and HLA residues at the binding groove are shown in green and white sticks.

In the case of the two HLA-A alleles (Figure 19C and D), the N- and C-terminal regions of the MICA-TM peptide showed a firmly established interaction, P1 with the Y159 phenyl ring (~100%) and P9 formed two H-bonds or salt bridge interactions with the K146 ammonium group (~70–80% in the BD-associated HLA-A*26:01, but < 60% in the non-associated one), respectively, at the α2 helix. Only HLA-A*26:01 (Figure 19C) contained a strong interaction at the nearly C-terminal end (P8 residue) with the α2 helix residue W178 (~70%) and the anchor P5 residue weakly interacted with the α1 helix residue T73 (~40%). The Y99 residue on the β-sheet of HLA-A*26:01 somewhat stabilized the P3 residue of MICA-TM (~40%), while the α1 helix residue N66 in HLA-A*11:01 supported the P2 and P3 residues (~40–50%). It is worthy to note that the high protein flexibility of HLA-A*11:01 (Figure 16D) had consequently led to a low binding of the incoming peptide, and in particular at the P5–P9 residues.

Based on the formed H-bonds, the peptide binding recognition was better distinguished in the HLA-A alleles. The lowered binding strength at the C-terminal P9 residue observed in HLA-B*35:01 was suspected to be the most important reason for the low selective binding affinity of this non-BD-associated HLA-B*35:01 allele. Thus, the relatively high protein flexibility in the non-associated HLA-A*11:01 allele led to

a decreased H-bond strength with the P9 residue and no stabilization for the P5 and P8 residues. All simulations fairly agreed with the crystal structures of the HLA/peptide complexes in which this P6 residue outwardly located off the binding groove [115].

4.1.5 HLA/peptide binding affinity

To determine the MICA-TM peptide binding strength towards the four studied HLA alleles, the molecular mechanic/Poisson-Boltzmann surface area (MM/PBSA) approach was applied on the same set of 100 snapshots taken from the production phase. The MM/PBSA approach has been extensively used to predict the overall Gibbs free energy (ΔG_{bind}) in biomolecular systems [118]. The ΔG_{bind} of the HLA alleles with the peptide bound at the binding groove was calculated from a summation of the total MM energy (ΔE_{MM}), the solvation free energy (ΔG_{sol}) and the entropic term (TΔS). The last term was estimated from the normal mode analysis [119]. Although the ΔE_{MM} components suggested that the short peptide was better stabilized by the non-BD-associated HLAs in the gas phase, the large destabilization from the solvation effect on the HLA/peptide complex led to a lower total ΔG_{bind}. Rather, within each paired HLA-I gene (HLA-A or HLA-B), the MICA-TM peptide interaction was stronger with the BD-associated allele, where B*51:01 (-56.10 kcal/mol) > B*35:01 (-48.88 kcal/mol) and A*26:01 (-42.97 kcal/mol) > A*11:01 (-31.14 kcal/mol) in **Table 7**. The obtained results from the present study help to differentiate the HLA alleles and explain a source of BD.

Table 7. The binding free energy and energy components (kcal/mol) for the four HLA/MICA-TM complexes predicted by the MM/PBSA method.

	HLA-B*51:01 [a]	HLA-B*35:01 [b]	HLA-A*26:01 [a]	HLA-A*11:01 [b]
ΔE_{vdW}	-74.7 ± 4.2	-78.1 ± 4.7	-80.0 ± 4.8	-63.7 ± 4.2
ΔE_{elec}	-168.4 ± 16.4	-199.4 ± 17.1	-127.2 ± 20.8	-208.1 ± 25.3
ΔE_{MM}	-243.1 ± 16.1	-277.6 ± 16.4	-207.3 ± 21.6	-271.7 ± 25.0
ΔG_{polar}	199.1 ± 13.9	241.1 ± 14.8	177.3 ± 20.4	252.2 ± 23.6
$\Delta G_{non\text{-}polar}$	-12.1 ± 0.4	-12.5 ± 0.3	-13.0 ± 0.3	-11.6 ± 0.4
ΔG_{sol}	187.0 ± 13.9	228.7 ± 14.8	164.3 ± 20.4	240.6 ± 23.4
-TΔS	9.7±25.8	12.7±24.1	21.5±28.6	15.4±21.4
ΔG_{bind}	-46.4	-36.2	-21.5	-15.7

HLA alleles are [a] associated or [b] not associated with Behçet's disease (BD). Data are shown as the mean ± SD, derived from independent simulations. Means within a paired row (HLA-A or HLA-B alleles that are associated with BD versus that are not) followed by a different letter are significantly different. ΔG_{bind} is the binding energy with inclusion of entropic term.

4.1.6 Conserved interaction of the BD associated HLAs

Based on MM/PBSA method, the alanine scanning mutagenesis commonly used to signify the important function of residues was carried out on the two BD associated HLA residues within the 5 Å sphere of the MICA-TM peptide. The results of the relative binding free energy ($\Delta\Delta G_{binding} = \Delta G_{wild-type} - \Delta G_{mutant}$) are given in Table 8. Note that the entropy term of the complex is not significantly changed by only one residue substitution with alanine; therefore the entropic calculation was neglected. By a mutation to alanine on the 40 residues, the HLA residue numbers 70, 73, 99, 146, 147 and 159 contributing the $\Delta\Delta G$ value < -2 kcal/mol in either HLA-B*51:01 or HLA-A*26:01 (Table 2) were considered as the important residues. These six residues could provide the conserved interaction of both BD associated HLAs for binding of the incoming MICA-TM.

Table 8. Relative binding free energy upon alanine mutation ($\Delta\Delta G_{binding}$) for the HLA residues within 5 Å sphere of the MICA peptide. The residues with $\Delta\Delta G_{binding}$ of < -2 kcal/mol for both HLA classes are shown in bold text.

residue	$\Delta\Delta G_{binding}$ (kcal/mol)	
	HLA-B*51:01	HLA-A*26:01
Y7	-0.43	-2.40
Y9	-1.49	-0.06
I66/N66	-1.89	-1.90
N70/H70	-3.37	-2.57
T73	-2.42	-2.45
Y74/D74	-1.67	-0.94
N77	-1.68	-0.73
I80/T80	-1.31	-0.57
A81/L81	0	-0.93
Y84	-0.53	-1.14
Y85	-1.01	-0.11
W95/I95	-1.04	-0.16
T97/R97	-0.08	-3.42
Y99	-2.52	-2.22
Y116/D116	-0.99	-1.07
Y123	-1.65	-1.19
T143	-2.25	-1.44
K146	-3.68	-2.08
W147	-5.54	-4.64
E152	-1.76	-0.82
Q155	-0.15	-2.12

residue	$\Delta\Delta G_{binding}$ (kcal/mol)	
	HLA-B*51:01	**HLA-A*26:01**
L156/W156	-1.04	-3.44
Y159	-3.63	-2.85
W167	-0.03	-1.82

4.2 Part II: Systemic sclerosis disease

4.2.1 Dynamics behavior of HLA-DR complexed to Top1 peptide

Root-mean-square displacement (RMSD) calculations were performed to monitor the conformational stability in the overall MD simulation, as shown in Figure 21. RMSD along simulated time was evaluated from the geometrical coordinates of protein backbone (N-Cα-C-O) with respect to those of the initial structure. For HLA-DRs without Top1, RMSDs for the binding cleft and the whole protein were plotted in Fig. 2A. In the first 2-ns, RMSD value rapidly reaches up to 2 Å for HLA-DR binding cleft and whole protein in every system, and consequently establishes dynamics equilibrium. For the complex form (Figure 21B), HLA-DR/Top1 complex, HLA-DR, binding cleft and nonameric core sequence of Top1 peptide are separately considered. Similar curves are observed for the complex and the HLA-DR protein (black and red in Figure 21B), indicating that the principle dynamics in the system largely depend on HLA-DR part. The systems enter the equilibrium phase after the simulation reaches 12.5-ns for HLA-DR free form and 5-ns for HLA-DR/Top1 complex. Several sections during equilibrium phase of HLA-DR and HLA-DR/Top1 systems are rechecked by the RMSD frequency. The high frequency of distribution displacement within 1 Å is properly examined to resolve the orientation for productive sampling, especially for HLA-DR binding cleft and Top1 peptide. The conformational selections within equilibrium production are investigated for binding interaction of HLA-DR/Top1 complexes and dynamical behavior of free HLA-DRs.
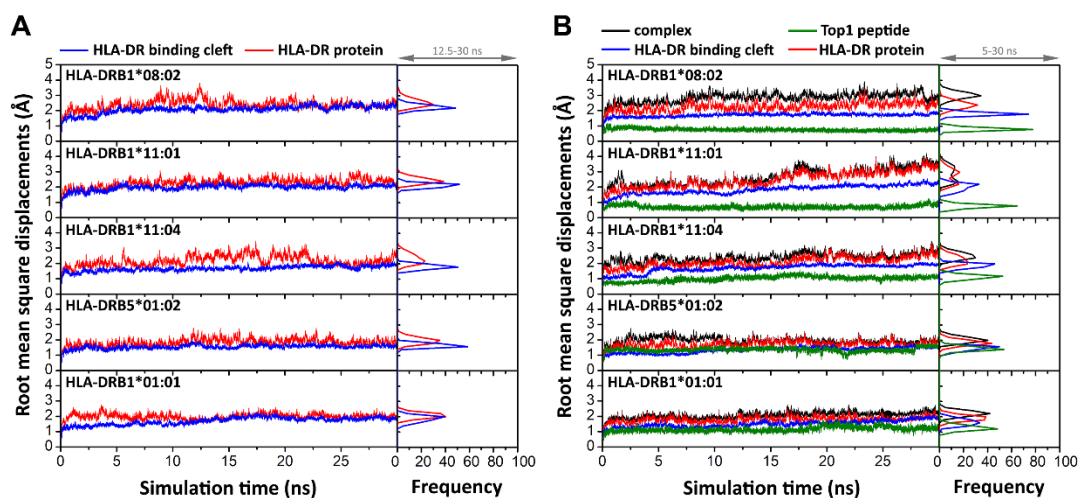
Figure 21. RMSD on time series of HLA-DRs. (A) A variety of HLA-DRs without and (B) with Top1 binding are separately plotted for RMSD of the backbone atoms for HLA-DR/Top1 complex (black), HLA-DR protein (red), binding cleft (blue) and nonameric core sequence of Top1 peptide (green). Distributions of RMSD distance during productive selection of 12.5-30 ns for HLA free form and 5-30 ns for complex form are collected as the relative frequency.

To consider flexibility of the binding cleft in cases of Top1 bound and unbound, the set of 250 snapshots is superimposed within their simulated systems. In Fig. 3A-B, the binding cleft consists of the α-chain (residues 5-80 shown in tan ribbon) and the β-chain (residues 5-93 in pink ribbon), while the cleft-distal domains are not displayed. Each chain is constructed by four antiparallel sheets and one long helix. Empty HLA-DR clefts are depicted in the first column (Figure 22A), and the next column represents individual HLA-DR of the same row binding with Top1 (Figure 22B). For the molecular set of HLA-DR types, the superimposition of empty clefts exhibits open shape and high flexibility, especially at the helix and turn parts on α-chain. The binding cleft manifests lower flexibility when binding with Top1 peptide due to an induced-fit mechanism, except for the non-associated ATASSc (HLA-DRB1*01:01) that still writhes at the turn and coil conformations on α-chain. For HLA-DR types with ATASSc association, the core length (p1-p9) of Top1 peptide orderly lies on the binding cleft, while two ends extend out of the cleft with a free motion. Unsurprisingly, Top1 peptide shows the highest flexibility in HLA-DRB1*01:01 complex (non-associated case).

To support this finding, the distance between the two centers of gravity (CG) of α-helixes are measured to represent a cleft size of HLA-DR with and without Top1 peptide bound. Overview of average distance is estimated to be in a range of 20.5-22.5 Å without peptide and becomes somewhat shorten to 19.0-22.3 Å with peptide binding. The effect of induce-fit with Top1 peptide leads to two α-helixes approaching each other only found in HLA-DRs associated and suspected ATASSc. By a comparison within identical HLA-DR molecules, the difference in the average distances of the binding cleft without and with Top1 are indicated in the parentheses as follows: HLA-DRB1*08:02 (-0.8 Å), HLA-DRB1*11:01 (-3.1 Å), HLA-DRB1*11:04 (-2.4 Å), HLA-DRB5*01:02 (-1.2 Å) and HLA-DRB1*01:01 (+1.0 Å), where the negative and positive values are referred to smaller and larger distances of binding cleft upon Top1 binding, respectively. In Fig. 3C, the profile of normal distribution of distance is quite similar for almost all HLA-DRs that are a broad peak for free protein and a narrow peak shifting to the shorter distance for the Top1-bound cases. According to the high protein flexibility, this could explain why the peak is spread out in the free form of HLA-DRs. The binding clefts become more stable by complexation with Top1, which is illustrated by a narrow peak of the distance distribution. More oscillatory configuration and the peak of distance distribution shifting to the extended distance suggest the less-fit Top1 in non-ATASSc binding cleft for HLA-DRB1*01:01.
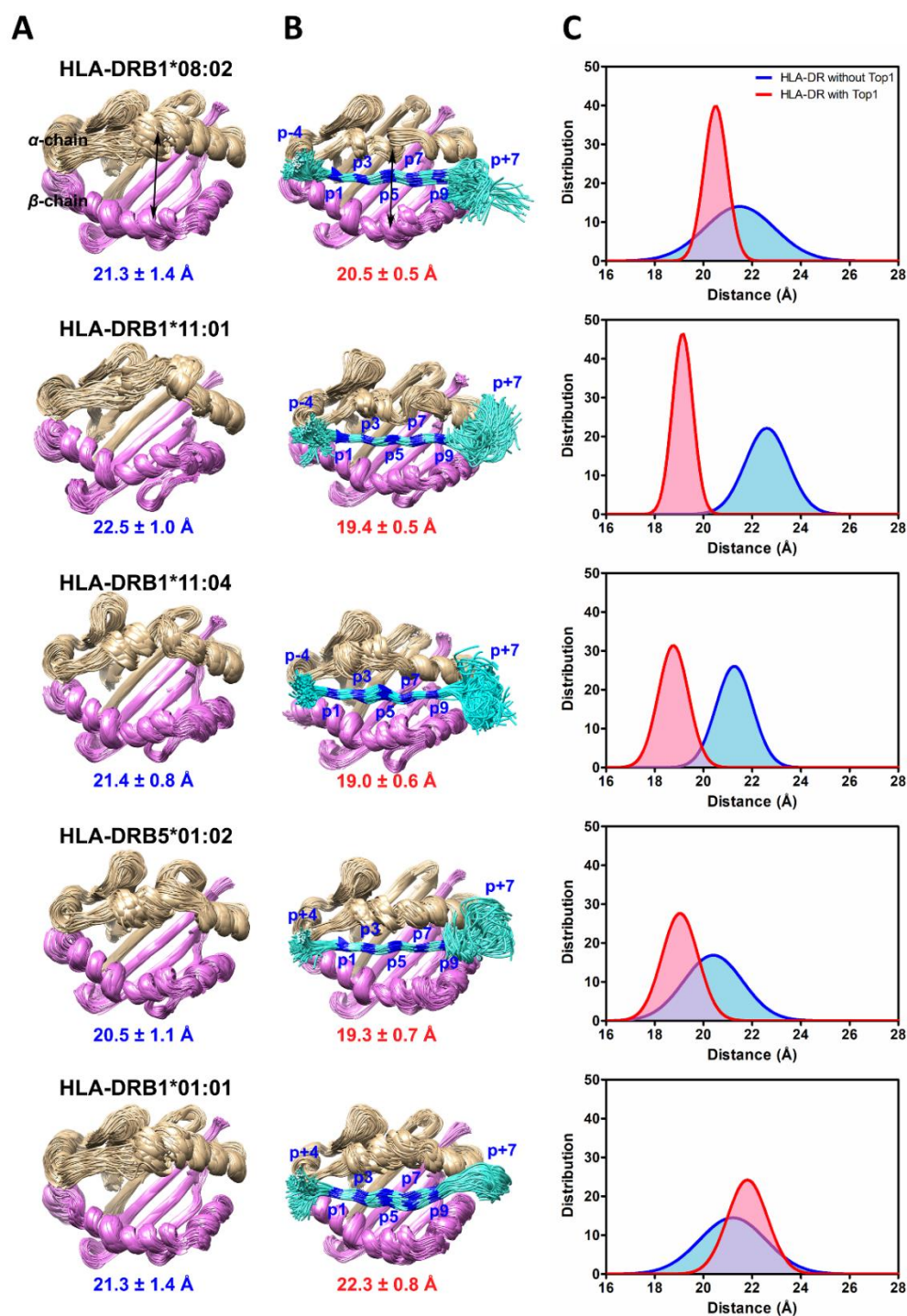
Figure 22. Comparison of HLA-DR binding clefts without and with Top1 peptide (cyan-blue multicolor) in columns A and B, respectively. Column C represents distribution of the distance between α-chain and β-chain of the binding clefts. The distance distribution of HLA-DR in free form is shown as a blue peak and that for the binding peptide is shown as a red peak.

4.2.2 Binding free energy calculations

MM-PB(GB)SA and QM/MM-GBSA binding free energy calculations were applied on the HLA-DR/Top1 complexes to quantify the antigenic binding strength in diverse antigen presenting types using 250 equilibrated frames. Binding free energy (ΔG_{bind}) based on MM-PB(GB)SA is contributed by enthalpy and entropic terms at a constant temperature (-TΔS). The enthalpy term contains molecular mechanic energy (ΔE_{MM}), and the solvation term (ΔG_{sol}). The latter one comprises of polar and nonpolar solvation energies. The energy components are enumerated in Table 9. In gas phase, the electrostatic (ΔE_{ele}) energy mainly contributes for the binding interaction between HLA-DR and Top1 peptide, which is over four times stronger than the van der Waals (ΔE_{vdW}) attraction. The most stable complex is HLA-DRB5*01:02/Top1 (the suspect system with MM energy of -928.0 kcal/mol), while three HLA-DRs with ATASSc-association systems show such ΔE_{MM} values ranging from -761 to -673 kcal/mol. The non-associated ATASSc protein, HLA-DRB1*01:01, has the lowest protein-protein interaction according to the ΔE_{MM} term of -551.8 kcal/mol. Moreover, the degree of disorder in the systems is examined by the entropic term consisting of rotational, translational and vibrational modes. The normal mode analysis at 298 K was invoked on the same set of 50 snapshots via mmpbsa_py_nabnmode program to derive the entropic term. Cooperation of enthalpy and entropy invoke from binding free energy clearly exposes that Top1 peptide had strong binding with ATASSc-suspect HLA-DRB5*01:02 (-49.0 kcal/mol), moderate binding with ATASSc-associated HLA-DRB1*08:02 (-27.0 kcal/mol), HLA-DRB1*11:01 (-35.8 kcal/mol), HLA-DRB1*11:04 (-29.7 kcal/mol) and rather weak binding with ATASSc-unassociated HLA-DRB1*01:01 (-11.6 kcal/mol), according to MM-PBSA approach. MM-GBSA binding free energies reveal similar tendency as stated in Table 1. Surprisingly for ATASSc-suspect, the HLA-DRB5*01:02/Top1 has the tightest binding interaction among HLA-DRs studied here. However, the self-antigen is only slightly bound to non-associated with ATASSc type, HLA-DRB1*01:01, as hypothesized. From the obtained results, it can be implied that not only HLA-DRB1*15:02 is the known gene expression in Thai-SSc patients but the strong linkage disequilibrium of HLA-DRB5*01:02 also is related to SSc disease. Work on the clinical data to confirm the relation between the genes linkage disequilibrium and the severity of the SSc disease are in progress.

Table 9. The binding free energy and energy components (kcal/mol) for the five HLA-DR/Top1 complexes predicted by the MM-PB(GB)SA and QM/MM-GBSA methods.

Energy Component	HLA-DRB1*08:02[a]	HLA-DRB1*11:01[a]	HLA-DRB1*11:04[a]	HLA-DRB5*01:02[b]	HLA-DRB1*01:01[c]
Gas term					
ΔE_{vdW}	-118.0 ± 8.2	-132.9 ± 7.2	-126.5 ± 6.6	-136.6 ± 7.4	-110.2 ± 8.5
ΔE_{ele}	-554.7 ± 47.4	-627.6 ± 54.1	-573.5 ± 56.3	-791.4 ± 80.8	-441.5 ± 46.1
ΔE_{MM}	-672.6 ± 51.0	-760.5 ± 53.5	-700.0 ± 57.3	-928.0 ± 80.6	-551.8 ± 49.5
ΔE_{QM}	-62.1 ± 15.4	-41.6 ± 20.1	-41.6 ± 20.1	-120.3 ± 23.5	-43.7 ± 21.6
$-T\Delta S$	73.5 ± 15.9	66.5 ± 15.1	64.0 ± 17.4	70.5 ± 14.6	67.2 ± 16.2
Solvation term					
$\Delta G_{sol(PBSA)}$	572.1 ± 42.2	658.3 ± 49.1	606.4 ± 51.9	808.5 ± 69.9	472.9 ± 43.6
$\Delta G_{sol(GBSA)}$	569.4 ± 41.9	649.5 ± 48.2	595.3 ± 2.2	799.6 ± 2.7	470.0 ± 41.8
$\Delta G_{sol(QM-GBSA)}$	570.9 ± 58.0	629.4 ± 52.1	629.4 ± 52.1	819.2 ± 68.5	488.5 ± 38.5
Binding free energy					
$\Delta G_{bind(MM/PBSA)}$	-27.0 ± 14.4	-35.8 ± 12.9	-29.7 ± 14.7	-49.0 ± 16.3	-11.6 ± 13.8
$\Delta G_{bind(MM/GBSA)}$	-29.7 ± 15.7	-44.6 ± 12.5	-40.8 ± 13.6	-57.9 ± 14.4	-14.6 ± 13.7
$\Delta G_{bind(QM/MM-GBSA)}$	-20.4 ± 19.0	-33.4 ± 17.0	-30.0 ± 21.8	-44.5 ± 21.1	-3.0 ± 19.1

The complexes of Top1 self-peptide and HLA-DR molecules for [a] association, [b] suspect, and [c] non-association with ATASSc. Mean ± s.d. is estimated from independent simulations. The MM-PB(GB)SA and QM/MM-GBSA calculations are applied on 250 snapshots, whereas NMODE are performed into entropic terms on 50 snapshots.

To strengthen the validity of the calculated binding energies for HLA-DR/Top1 complexes, the higher-level computation was applied using QM/MM-GBSA method. In QM region, the nonameric core sequence of Top1 peptide was treated with SCC-DFTB implementation. The $\Delta G_{bind(QM/MM-GBSA)}$ confirmed that the major HLA-DR binding cleft interacts with the nine core residues (p1 to p9) of Top1 peptide. Nonameric residues, p1 to p9, have conserved interactions with the HLA-DR cleft of associated (in a range of -20.4 to -33.4 kcal/mol) and suspect ATASSc (-44.5 kcal/mol) types; in contrast, they are barely attracted on the non-associated allele at -3.0 kcal/mol. In addition, it is questionable whether HLA-DRB1*01:01 still keeps Top1 on its cleft (Figure 22B). Explanation for HLA-DR/Top1 binding is going to be more enlightened by per-residue energy decomposition and hydrogen bonding across protein-protein interface.

4.2.3 Per-residue energy decomposition

According to the structural database, the peptide motifs complexed to HLA class II were generally believed to have the same burial pattern with the key residue

contacts at p1, p4, p6 and p9[107]. To determine this point, the structural superimposition over all studied systems taken from the last MD snapshot was carried out. By a comparison in the side chain directions, *i.e.* carbon beta (CB) atom in Figure 23A, five HLA-DR molecules are predicted for possible key anchors of Top1 peptide. The CB atoms of p1, p4, p6 and p9 are likely to face the binding cleft.
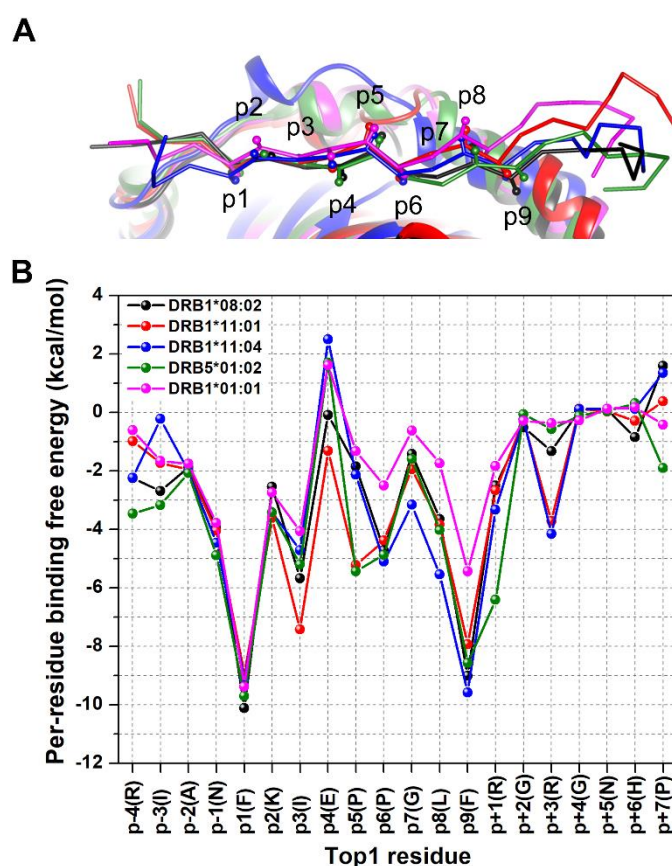
Figure 23. Top1 peptide binding with various HLA-DR molecules. (A) A superimposition of representative Top1 in five MD systems showing a side view of HLA-DR binding cleft. The Top1 binding to HLA-DRB1*08:02 is shown in black, HLA-DRB1*11:01 in red, HLA-DRB1*11:04 in blue, HLA-DRB5*01:02 in green and HLA-DRB1*01:01 in magenta. The Top1 peptide is depicted by the backbone chain and CB atoms of the side chains as a stick and a ball, respectively. (B) Per-residue binding free energy of Top1 peptide for all five HLA-DRs.

The relative binding affinity of Top1 in various HLA-DR types is examined by per-residue energy decomposition using implicit solvent model. The total binding free energy excluding entropic contribution is plotted per-residue in Figure 23B to illustrate the Top1 binding pattern. The binding free energy of individual amino acids within the core 9-mer is compared at identical position among all five systems. Energetic plot per residue shows similar tendency of Top1 binding interaction on diverse HLA-DR molecules (Figure 23B). Phenylalanine at the p1 and p9 positions clearly shows a strong interaction with HLA-DR with the binding free energy of < -8 kcal/mol. On the other hand, the p9 phenylalanine exhibits a weaker binding (~ -5 kcal/mol) to HLA-DRB1*01:01 insusceptible to ATASSc relative to the other antigenic presenter molecules. The residues concerning HLA-DRs' sub-pockets confer a significant effect to Top1 interaction; therefore, the residues within 5 Å of Top1 are focused in which the residues with total decomposition energy ≤ -1 kcal/mol are compared in supplemental Figure 24. With energy subdivision, the total energy is decomposed into two terms, polar ($\Delta E_{ele} + \Delta G_{polar}$) and nonpolar ($\Delta E_{vdW} + \Delta G_{nonpolar}$) contributions. Overall Top1's residues are mainly stabilized by HLA-DR protein through nonpolar interactions.

In Figure 25, energetic fingerprints are represented as color-coded surface models of various HLA-DR binding clefts interacting with Top1 in comparison to the HLA-DRB1*01:01/Top1 $\left(\Delta G_{residue}^{HLA-DR} - \Delta G_{residue}^{HLA-DRB1*01:01}\right)$. The HLA-DRs equally interact to p1 by hydrophobic residue cluster comprising of F24$_\alpha$, F32$_\alpha$, A52$_\alpha$, a hydrophilic S53$_\alpha$ on α-chain, and V85$_\beta$ on β-chain. Pi-pi stacking and pi-alkyl interactions are main stabilizations for the p1(F) side chain. The p2(K) and p3(I) are maintained on the cleft supported by N82$_\beta$ and pi-alkyl interaction with F54$_\alpha$, respectively. For some HLA-DR cases, a negative charge-charge repulsion between p4(E) and D70$_\beta$ side chains is found by a blue surface in Figure 25. Repulsive force on p4(E) is observed in HLA-DRs associated and suspected ATASSc, whereas Q70$_\beta$ for HLA-DR unassociated ATASSc exerts a polar attraction to p4. However, p4(E) is stabilized in the sub-pocket by Q9$_\alpha$ and/or R71$_\beta$ instead (Figure 24). A mere alkyl hydrophobic L74$_\beta$ belonging to HLA-DRB1*08:02 molecule has a longer side chain than A74$_\beta$ of others, which could additionally stabilize the CB-CG p4(E) side chain. Besides, auxiliary interactions come from the adjacent residues of p4(E) like p3(I) and p5(P).
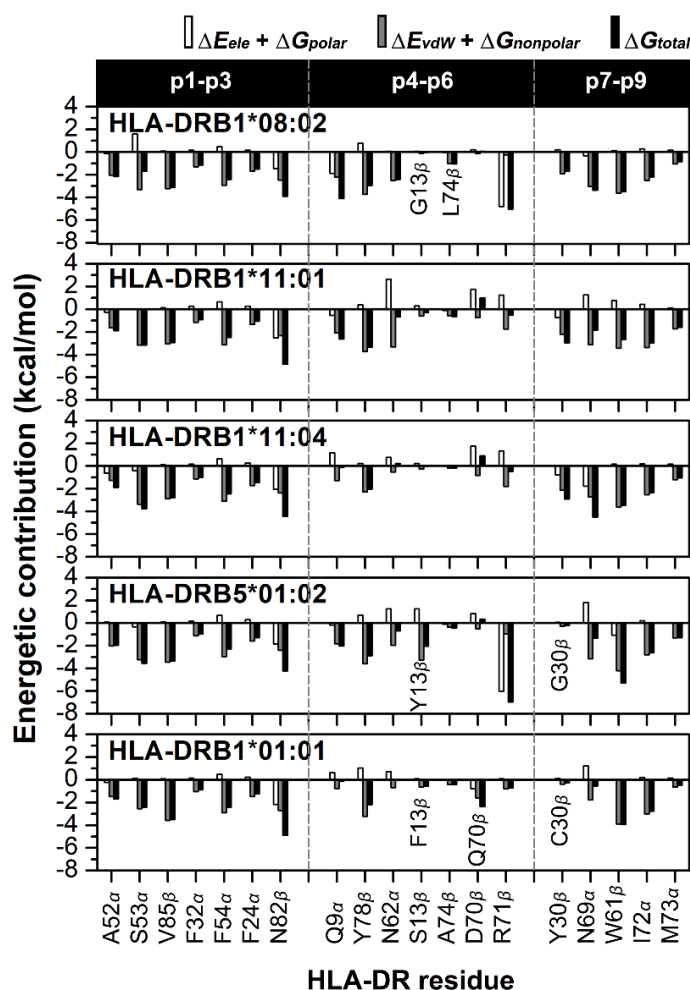
Figure 24. HLA-DR residues interacting with p1 to p9 of Top1 peptide. Total decomposition energy with per-residue contributed by polar (white bar) and nonpolar (gray bar) interactions.

The pyrrolidine side chain, a nonpolar group, of p6(P) on Top1 is mostly buried within the pocket with a total free energy in the range from -5 to -4 kcal/mol, while only the HLA-DRB1*01:01 case shows the free energy around -2 kcal/mol (Figure 23B). Energetic contribution for p6(P) is partially participated by the residues 13 and 30 of the β-chain depending on the protein type. One of polymorphic residues is G13$_\beta$ for HLA-DRB1*08:02, S13$_\beta$ for HLA-DRB1*11:01 and HLA-DRB1*11:04, Y13$_\beta$ for HLA-DRB5*01:02, and F13$_\beta$ for HLA-DRB1*01:01. Sharing contribution of 30$_\beta$ interaction between p6 and p9 is verified on the protein surface (Figure 25). Residue 30 partially constructing the p9-pocket plays an important role in efficient p9 holding of

each protein. This pocket consists of nonpolar or aromatic residues and could contribute to the p9(F) peptide interaction with equivalent binding level, excepting in the non-association case (Figure 23B). Exerting comparable nonpolar/aromatic effects, $Y30_\beta$ and $G30_\beta$ residues are respectively found for three HLA-DRs associated and suspected ATASSc, but changed to the polar $C30_\beta$ for non-ATASSc association leading to the loss of p9(F)-interaction. The remaining residues of Top1 (p5, p7 and p8) within HLA-DRB1*01:01 complex is estimated to have twice less binding efficiency than energetic trend of the other four HLA-DRs, as revealed in Figure 23B. According to the decomposition energies, the p4, p6 and p9 of HLA-DRB1*01:01/Top1 complex show a decrease in binding affinity caused by a divergent amino acid supporters 13, 30, 70 and 74. Overall energetic fingerprints normalized with HLA-DRB1*01:01/Top1 are displayed in red tone at the pocket surrounding p1 and p9 of four HLA-DRs, whereas the middle region of HLA-DRB1*11:04 groove is not different from HLA-DRB1*01:01 (Figure 25). Although the ATASSc-suspect system, HLA-DRB5*01:02/Top1 has quite unlike amino acids from ATASSc association type, it nevertheless presents a similar energy contribution to them.
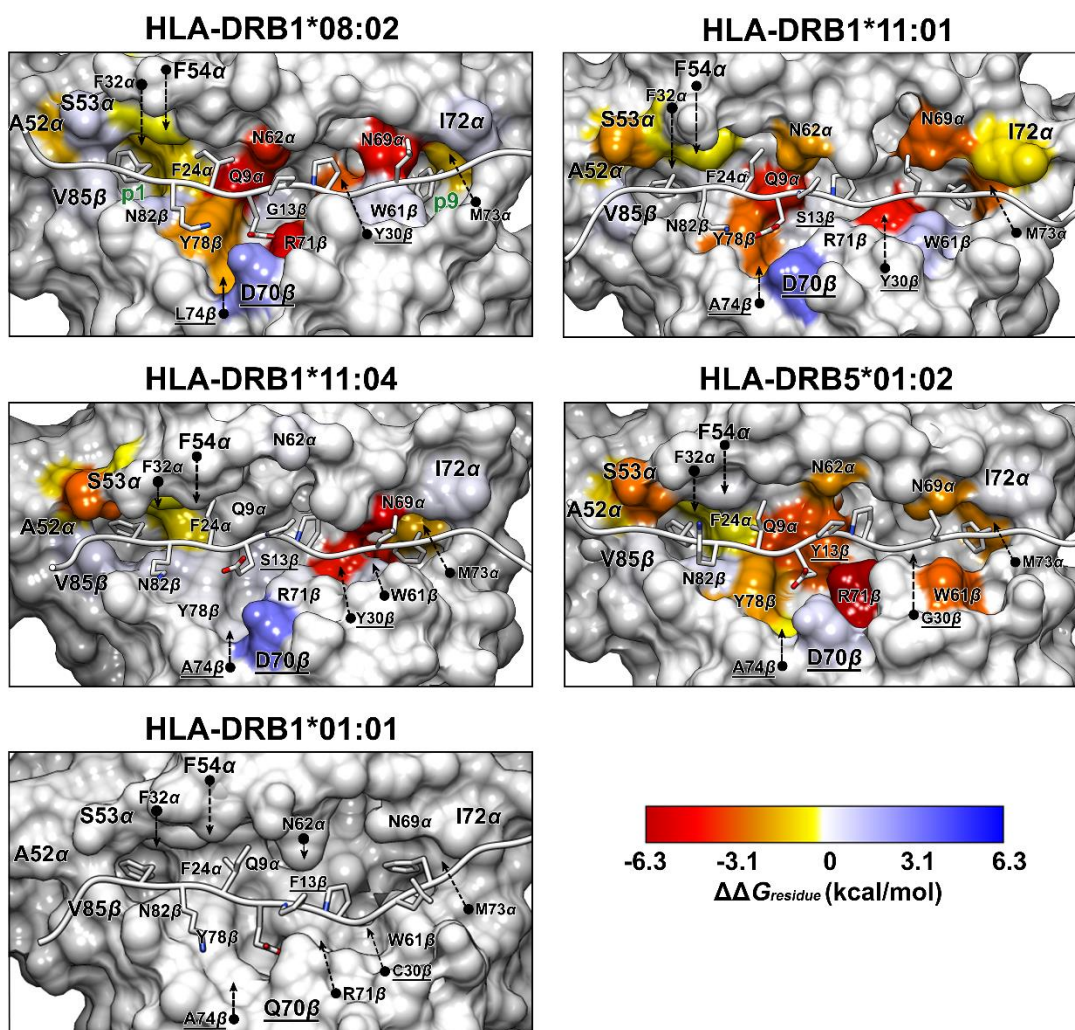
Figure 25. Energetic fingerprints of difference ($\Delta\Delta G_{residue}$) in per-residue free energy decomposition values of HLA-DR/Top1 complexes compared to non-associated with ATASSc HLA-DRB1*01:01. Energy contribution of HLA-DR residues for Top1 binding is illustrated by surface coloring. By comparing to per-residue decomposition energy of HLA-DRB1*01:01, the stronger binding is colored from yellow to red gradient and in contrast the weaker binding is shown in blue gradient.

4.2.4 Hydrogen bonds across protein-protein interface

A significant stabilizing contribution for HLA-DR/Top1 recognition occurrences is by salt bridge (SB) and hydrogen bond (HB) interactions at the protein-protein interface as seen in the gas phase energetic components for total binding free energy (Table 9) that the electrostatic interaction is highly contributed for HLA-

DR/Top1 binding. Stabilization by hydrogen bonding interactions is determined by the percentage of occupation based on the criteria for the distance of hydrogen bond donor and acceptor atoms (HD...A) of ≤ 3.5 Å and the D-H...A angle of $> 120°$. HB occupations between Top1 and sub-pocket residues are exhibited by grid cells in Figure 26. The hydrogen bond strengths are divided into 3 levels: low (10-39 %), moderate (40-69 %), and strong (70-100 %) interactions represented by the gradient of greenish, bluish and reddish grid cells, respectively. The nonameric core peptide consisting of –FKIEPPGLF– is buried in the antigenic binding cleft, while the rest of the residues on $-NH_3^+$ and $-COO^-$ termini stay away from the cleft.

A similar HB pattern across protein-protein interface is found in all complexes by a strong interaction between the backbone atoms of p1(F) to $S53_\alpha$ and p2(K) to the $N82_\beta$ amide group. In HLA-DRB1*08:02/Top1 complex, the acidic side chain of p4(E) is repulsed by the negative charge from $D70_\beta$; meanwhile, a salt bridge is found with the $R71_\beta$ guanidinium group. The charge-charge repulsion of $D70_\beta$–p4(E) leads to a decreased HLA-DR recognition according to the per-residue energy decomposition. However, the p4(E) backbone atoms form hydrogen bonds with the $N62_\alpha$ and $Q9_\alpha$ amide groups in the binding cleft which assists conservative flanking for p4(E) in case of HLA-DRB1*08:02 (strong), HLA-DRB1*11:01 (strong) and HLA-DRB5*01:02 (moderate). To compensate for Coulomb repulsion, neighboring p4(E) and p5(P) are stabilized by $R71_\beta$ for HLA-DRB1*11:01, HLA-DRB1*11:04 and HLA-DRB5*01:02. Since the amino acids at 11_β, 30_β and 70_β in the focused systems are distinct; the detected hydrogen bond strengths are different. The $Y30_\beta$ hydroxyl group in HLA-DRs with ATASSc-association strongly stabilizes p7(G), because it has a HB acceptor group and a longer side chain than the other two residues: $G30_\beta$ for HLA-DRB5*01:02 and $C30_\beta$ for HLA-DRB1*01:01, whose side chains are unable to form a HB. p7(G) is found to be weakly stabilized by $D11_\beta$ instead for the suspect system. Hydrogen bond networks on the $N69_\alpha$ amide group are linked to the backbone of the conserved amino acids of Top1 peptide like p7(G) and nonpolar p9(F) with strong, moderate and very weak hydrogen bonds for ATASSc-association, suspect and non-association systems, respectively. $Q70_\beta$ of HLA-DRB1*01:01 has weakly occupied hydrogen bond with p4(E) in contrast to unoccupied formation for the other HLA-DRs.
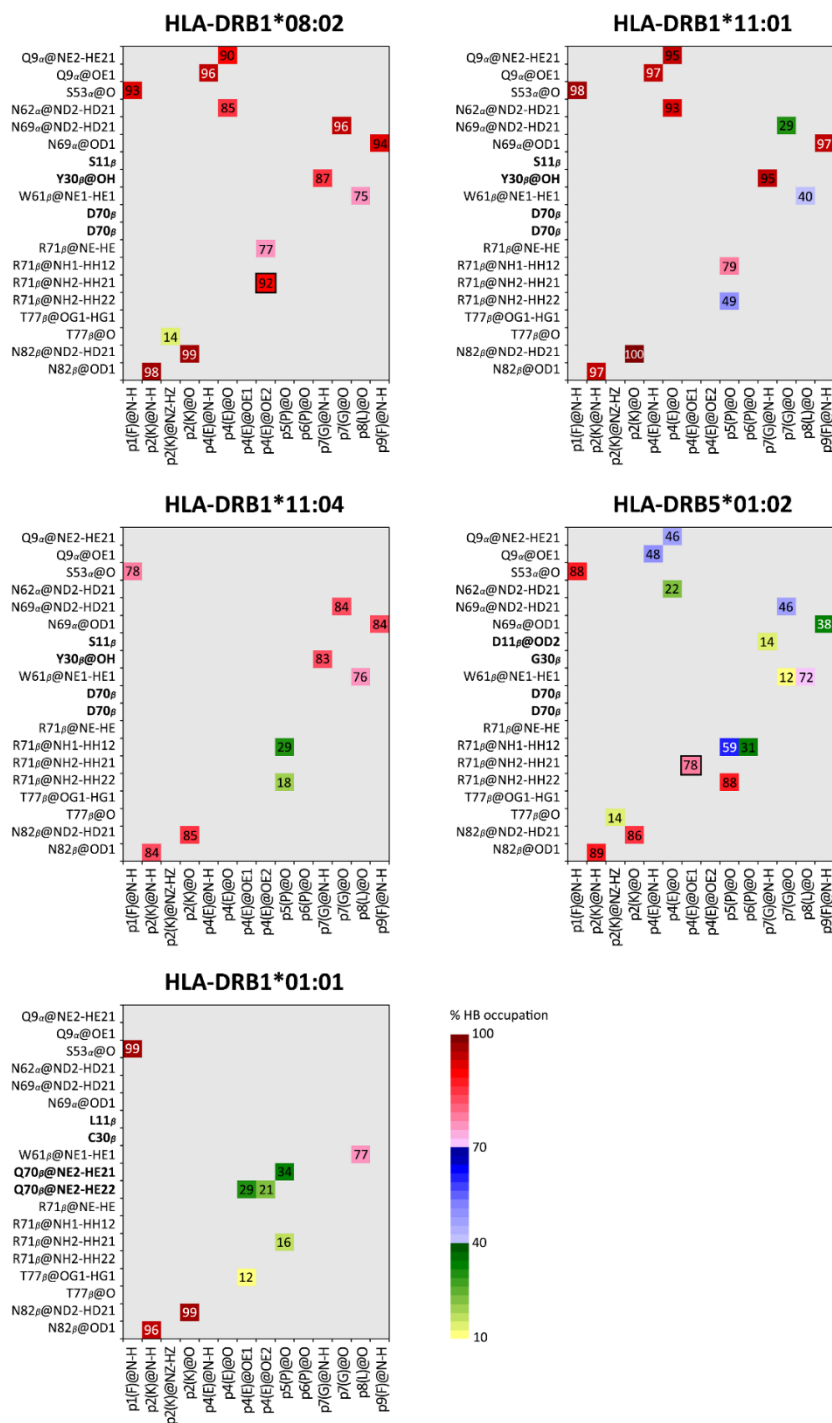
Figure 26. Hydrogen bond (HB) interactions of HLA-DR/Top1 complexes. HBs between HLA-DR (vertical axis) and Top1 (horizontal axis) residues are presented by grid map. Atom types are specified by tleap AMBER conversion. HB strength is defined by label and color in the grid cells, while salt bridge (SB) interaction is framed by a black border.

To examine the overall hydrogen bond interactions for different HLA-DR in recognizing Top1 peptide, the hydrogen bonds formed across the protein-protein interface in individual systems are counted and showed strong occupation ranging from 72 to 100 %, as illustrated in Fig. 6 with the numbers of reddish cells in the order of 12 (HLA-DRB1*08:02) > 9 (HLA-DRB1*11:01) > 7 (HLA-DRB1*11:04) > 6 (HLA-DRB5*01:02) > 4 (HLA-DRB1*01:01). Polar energy contribution is in agreement with hydrogen bond network and rather appeared in protein complexes of systematic ATASSc-association and suspect more than non-association. However, even though majority of Top1 from p1 to p9 are nonpolar residues, hydrogen bond and salt bridge interactions still appeared as additionally supporting interactions. Expansion beyond the nonameric core peptide at two sides mostly exhibits temporary occupation of hydrogen bonds with unexpectedly strong occupancy in a few bonds (Supplementary Figure S2).

4.2.5 Water solvation effect for self-peptide binding

Distribution of water accessibility at HLA-DR/Top1 interface was examined by 3D-RISM-KH molecular theory of solvation[120, 121] with TIP3P model using the rism3d.snglpnt module. Based on like-dissolve-like principle, a hydrophilic protein surface should find a higher amount of water than hydrophobic zone. As already pointed out before, the p1(F), p3(I)/p4(E), p6(P) and p9(F) of Top1 peptide are anchored on HLA-DR binding cleft. Water molecules accessibly inserted between these putative anchor residues and their surrounding residues of HLA in the binding cleft likely affect protein-protein binding strength. To predict the hydration sites around the antigenic binding, the stripped snapshots of peptide-protein complexes are inspected by 3D-RISM model. In 3D maps, the oxygen atoms of water molecules (in red flakes) are represented by 3D isosurfaces with a value > 3 of g(r) level (Figure 27), while hydrogen atoms are not shown. Water occupation is illustrated within HLA-DR sub-pockets around the nonameric core of Top1 peptide. Binding surface of HLAs class II is widely known to consist of flat and shallow sub-pockets, making it easy to be accessible by water molecules. Among all sub-pockets for Top1 binding in Figure 27, there are two rather deep sub-pockets supporting p1(F) and p9(F) anchors. Their phenyl ring is perfectly buried in the sub-pockets without any water molecules accessibility. In

contrast for non-associated ATASSc, HLA-DRB1*01:01 complex has a high possibility of accessible waters inside the p9-pocket to compete with the peptide-protein interaction that agrees well with a lesser binding affinity of p9 in Fig. 4B. This site of the ATASSc-insusceptible protein is constructed of specific hydrophilic residues $C30_\beta$ and $S37_\beta$, which differed from the other HLA-DR molecules. Even though the residue 37 on the β-chain of HLA-DRB5*01:02 is asparagine, its side chain points away from p9 making the hydrophobic pocket. For the three ATASSc-susceptible proteins, the –OH tail of $Y37_\beta$ also aligns away from p9(F).
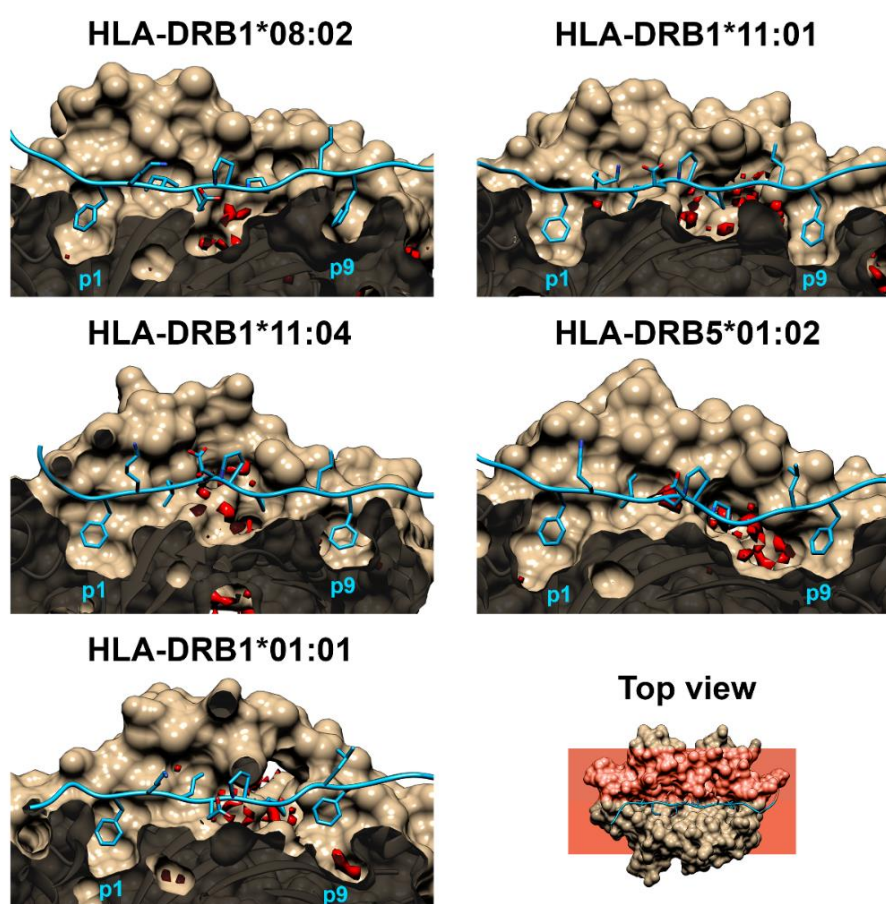
Figure 27. 3D distribution function of water oxygen atoms (red flakes) in the cross-sectional plane of HLA-DR/Top1 with g(r) > 3. HLA-DRs are sliced off at the binding cleft by a red plane perpendicular to a top view. A zoom-in of a side view represented by a tan surface is the sites specifically supporting p1 and p9 anchors. The core-nonameric residues of Top1 peptide embedded in the binding cleft is represented by a blue ribbon for the backbone and by a stick model for the side chains.

4.3 Part III: T-cell recognition

4.3.1 Structural properties at MD equilibria

To observe a binding stability of the simulated systems, the root mean square displacements (RMSDs) of the backbone atoms of HLA binding groove (black), peptide (red) and TCR (green) were measured along the simulation time with respect to their initial structures as plotted in Figure 28. In overview, RMSD values of HLA are observed in a range of 1.5-2.5 Å for all systems while RMSD values of peptide sequences are observed in a lower range of 0.7-1.5 Å. Interestingly, dynamical properties of TCR are observed in a similar magnitude as HLA at 1.5-2.5 Å. Except for the TCR/HLA-HIV8p system, the RMSD values are between 2.0 to 3.0 Å during the first 120 ns but changes a little to be between 2.0 to 2.5 Å afterwards. According to Figure 28, RMSD magnitudes of all simulated systems are fluctuating in a quite narrow range after 125-ns. Therefore, the production trajectories from the last 75 ns simulations are selected for further analysis in this work.
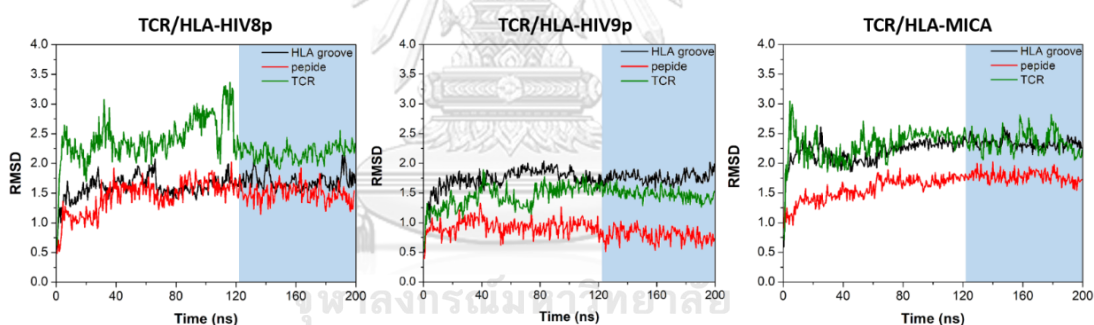
Figure 28. Root-mean square displacements (RMSDs) magnitudes of backbone atoms of HLA binding groove (black), peptide (red) and TCR (green) along 200-ns of simulation time

4.3.2 TCR rules of engagement

The first step in the T cell-mediated adaptive immune responses is presented by T-cell receptor (TCR) molecule recognizing the antigenic peptide laid on human leukocyte antigen presenter (pHLA). The pairwise residue contact map is simply marked as contacting or not contacting based on some criterion such as a cutoff distance. It is easy to analyze and suitable for the analysis of numerous residues. Therefore, the binding recognition pattern of TCR over the HLAp complex can be considered through pairwise

residue contact map analysis that covers intra- and inter-actions of protein configurations [122], protein-protein interactions, structural prediction and protein folding [123], successfully. Herein, inter atomic distance of Cα-Cα pairwise within 7 Å sphere are identified as interacting residues. The result of pairwise residue contact maps of each complexed systems is correlated and plotted in two dimension respective to the residue index while the distance magnitude of contact is given by red to blue color gradient between 4 to 7 Å as shown in Figure 29.

According to the plot, the residue index is defined as: 1-180 for HLA binding groove, 181-188 or 181-189 for peptide 8-meric or 9-meric residue, 190-297 for TCRα and 298-410 for TCRβ. Self-self contact or intra-folding interaction is clearly presented in the diagonal square region of the map. An information of pairwise contact in diagonal square region of HLA, TCRα and TCRβ clearly presents that there are many numbers of self-self contact interactions of both α-helix and β-sheet folding patterns within the protein domain (indicated by parallel and orthogonal characters with diagonal axial, respectively). In contrast, an information about the diagonal square region of HIV8p, HIV9p and MICA peptides is not observed suggesting the non-folding (linear-liked structure) character during the ternary complex formation. It should be noticed from this information that changing of the binding peptide sequence has no significant effect to the intra-folding character of HLA, TCRα and TCRβ domain as illustrated in quite similar self-self contact pattern.
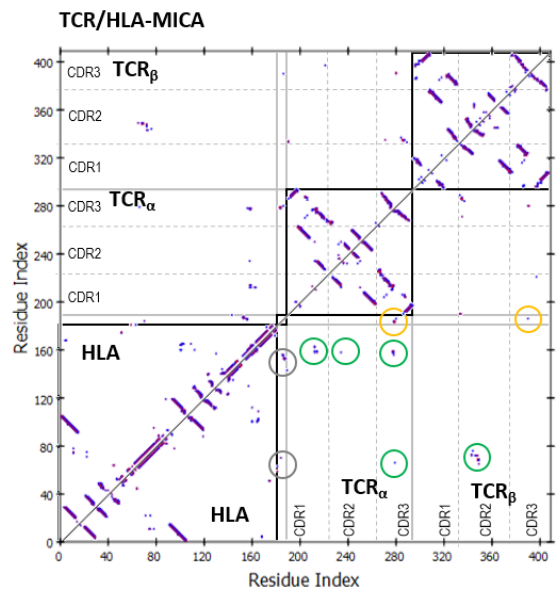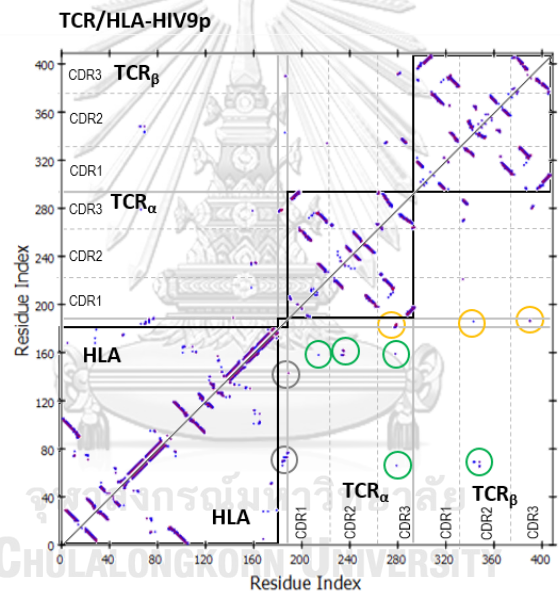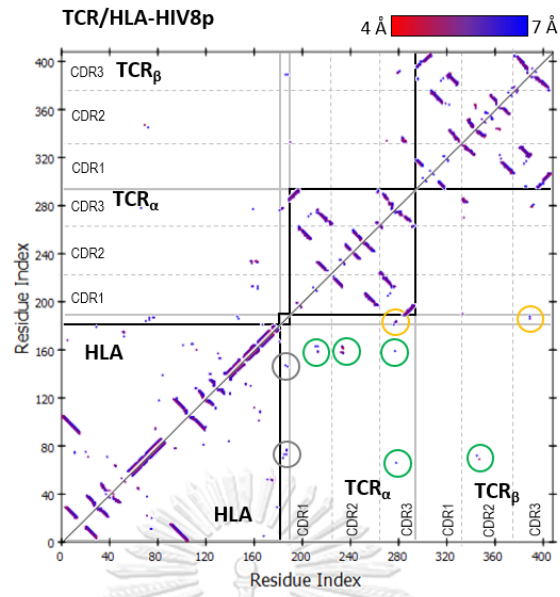
Figure 29. Pairwise residue contact maps of Cα-Cα distance of HLA binding groove, peptide and TCR$\alpha\beta$ displayed 4 to 7 Å by red to blue gradient, respectively. Residue index includes 1-180 for HLA binding groove, 181-189 for peptide, 190-297 for TCRα and 298-410 for TCRβ. (Note that the TCR residue of HIV8p (181-188) system starts at 189 to 409).

In principle, the "rule of engagement" introduces specific binding recognition pattern of the TCR/pHLA complex through triad restrictions for each TCR domain (CDR1 and/or CDR2 loops with HLA and CDR3 loop with antigenic peptide [124]). Importantly, many publications discovered the similar TCR/pHLA binding recognition pattern between autoimmune and infectious systems (like HIV) following the rule of the engagement criterion.

From Figure 29, all systems present somewhat similar groups of pairwise residue contacts as focused in gray, yellow and green circles for HLA-p, TCR-p and HLA-TCR, respectively. The most pairwise-residue contacts are observed between HLA and TCR proteins especially at the TCRα domain. Although TCRβ domain is not identified as the key region interacting with the HLA protein (in comparison to TCRα) during ternary complex formation, some different pattern between CDR2 of TCRβ and peptides is however observed (numbers of pairwise residue contact with TCRβ: MICA > HIV9p > HIV8p). In addition, pairwise residue contacts between CDR2 of TCRβ and 9-meric residue peptide is uniquely detected. These information allows us to suspect that tight binding order between TCRβ and peptide may be one of the important key factors for the selective binding recognition in autoimmune and/or antigenic infectious systems afterwards. To clarify this point, the details of chemical bonding interaction, decomposition energy and binding free energy will be investigated as described in the following sections.

4.3.3 Plasticity of TCR/HLAp complexes

To identify the plasticity of the protein regions, the flexibility per-residue with respect to its initial Cartesian coordinate position was investigated through root-mean-square-fluctuations (RMSF). For better understanding, the RMSF magnitude is

separately considered for the peptide, the HLA binding groove, the TCRα and the TCRβ domains as shown in Figure 30A-D.

It can be clearly seen from Figure 30A that the fluctuation behavior of peptide sequences is observed in a quite high rigidity in a range of 0.4 – 0.6 Å magnitude at the middle region. This result is somewhat fallen in our expectation because the peptide sequence is well fitted to the HLA binding grove and is additionally stabilized by the CDR3 domain of TCR as previously mentioned in the "rule of engagement" section. This information allows us to hypothesize that the binding of HLA-peptide with TCR may not influence the change in peptide conformation according to the high rigidity when the HLA-peptide binary complex is formed. Therefore, thermodynamics stability of the HLA-peptide binary complex may be considered as one of the important key factors that should be monitored for severity prediction of autoimmune or antigenic infectious diseases in our point of view. RMSF magnitudes of the two terminal edges of the peptide are presented by slightly higher values at 0.4 – 0.8 Å for p1 region and 0.6 – 1.0 Å for p8 or p9 regions. This is because these two edges of the peptide strain are located nearly outside of the binding groove pocket resulting in the higher degree of freedom for conformational flexibility compared to the middle region.

In Figure 30B, RMSF of all 180 residues along the binding groove domain of HLA protein are intensively analyzed. The key interacting residues with peptide and TCR fragments are specified in yellow and light blue areas for a clear understanding. From the results, RMSF magnitude of the HLA binding groove residues which exhibit some important interactions with peptide and TCR fragments seems to present a quite less plasticity (lower than 2.5 Å) compared to the other regions. RMSF magnitude of amino acid residues located on the free flexible loops (residues 18-20) and β-strand / α-helix junction loops (residues 39-45 and 83-86) are observed in quite high magnitude between 2.5 and 4.0 Å while the fluctuation behavior of the keys residues (α1 and α2 helices especially) in the HLA binding groove seems to show a higher rigidity and are less influenced by the peptide sequence variation in this case (HIV8p, HIV9p and MICA for black, green and yellow lines).
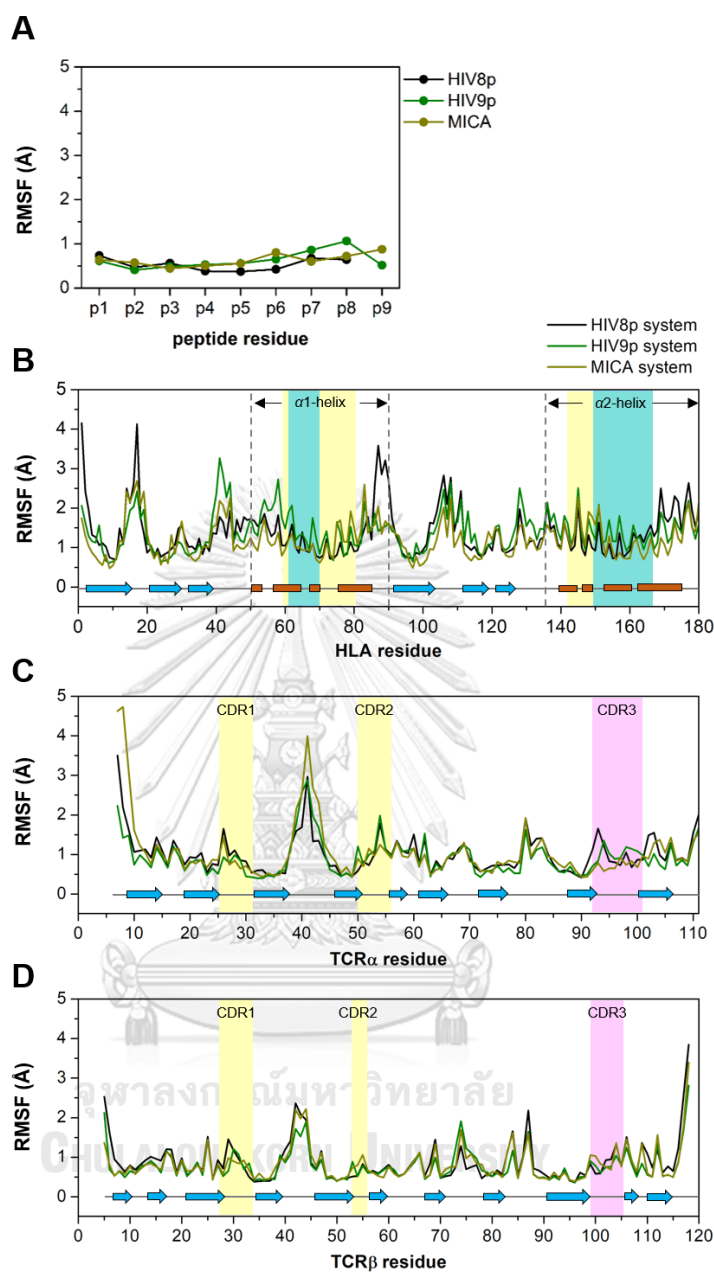
Figure 30. Root-mean-square-fluctuation (RMSF) analysis of TCR/HLAp complex. The fluctuation per residue of HIV8p, HIV9p and MICA peptide complexations is shown in A) the peptide, B) HLA binding groove, C) α-chain of TCR and D) β-chain of TCR.

Flexibility behavior of the TCR protein is here separately considered in TCRα and TCRβ profiles with yellow and magenta highlighted areas for key binding residues with peptide and HLA domains as given in Figure 30C and D In analogous to HLA

character, RMSF magnitudes of amino acid residues located on the free flexible loop are quite high (~ 4.0 Å) according to free solvent accessibility (α-chain residue 38-45 and β-chain residue 40-45) while flexibility of the binding residue domains (CDR1, CDR2 and CDR3 of both α and β subunits) is presented in a value smaller than 1.5 Å for all systems with almost similar pattern. The high rigidity of these fragments are in accordance with the previous Cα-Cα close contact information which indicates the significant distance relationship between HLA, peptide and TCR of both α and β subunits (circles in Figure 29). In addition, all these observations well support the "rule of engagement" concept that binding domain regions of TCR protein play an important role in stabilizing HLA protein and peptide sequence through CDR1, CDR2 and CDR3 fragments during HLA/peptide/TCR ternary complex formation in an early activation stage of autoimmune or antigenic infectious diseases recognition.

4.3.4 Interaction analysis of TCR/HLAp complexes
Decomposition Energy

In principle, the energy contributions of the contact residues which support TCR/HLAp ternary complex formation can be separately considered by decomposition energy (DC) analysis by MM/PBSA calculation (average magnitude over 150 frames of production phase). The focused binding interaction is identified by using DC magnitude ≤ -2 kcal/mol as a criterion. For better understanding, the selected per-residue DC energy distribution between HLAp and TCR domains are illustrated in three-dimensional structure with colors ranged by energy levels as given in Figure 31 (surface and ribbon stick models for HLA and peptide domain, respectively).

For HLA/MICA system (Figure 31A), the key binding residues of HLA (color surface) involved in TCR (stick model and green label) complexation are identified as *i)* HLAα1: R62, Q65, I66, K68, T69 and Q72 with TCRβ: Q53β, A58β, D59β and D60β, and *ii)* HLAα2: L163 with TCRα: I29α. In overview, TCR specific β chain prefers to bind HLAα1 than HLAα2. Our finding agrees well with the "rule of engagement" since CDR3 residue, R97α is identified as the key importance for peptide recognized at p5(A) and p6(I) by -2.2 to -2.7 kcal/mol. The involved binding residues of HLA in TCR/HLA-HIV9p system also play a role for TCR stabilization in the similar pattern with TCR/HLA-MICA system but the level of DC stabilization energies are

observed in a higher magnitude of ~ 0.5-1.5 kcal/mol (decreased stabilization). The same residue, R97α of CDR3 interacts to p4(V) with -2.8 kcal/mol.
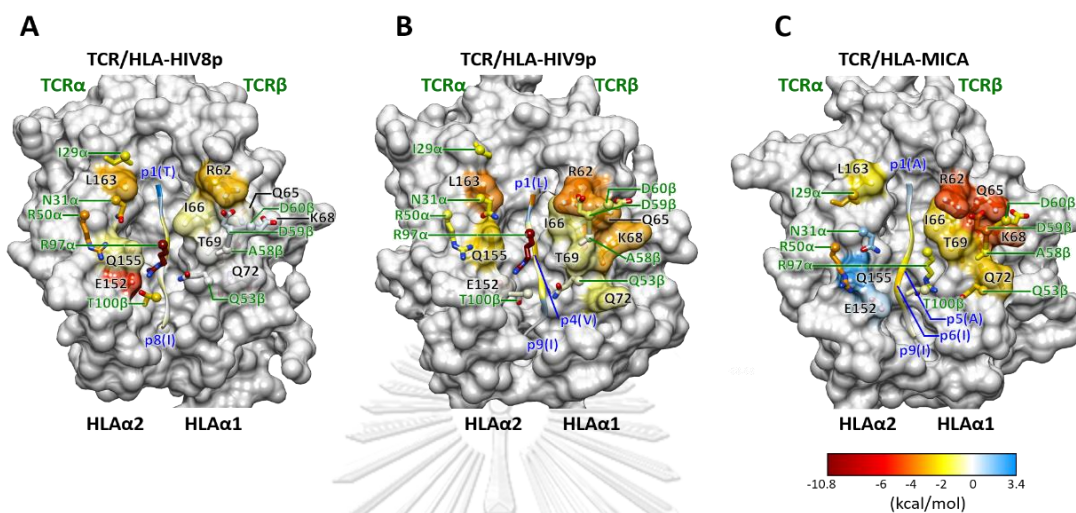
Figure 31. Decomposition energies per TCR/HLAp residue fingerprint painting are displayed for TCR and HLAp interactions. The TCR and HLAp are represented by stick and surface-ribbon models, respectively. The attractive binding is colored from yellow to red gradient in contrast to the repulsive binding which is shown in blue gradient in the HLA binding groove.

In Figure 31B and C, HLA-HIV9p and HLA-MICA complexes dominantly interact with the TCR protein through HLAα1 domain while HLA-HIV8p complex uses HLAα2 domain instead (red region). For HLA/HIV8p system, the key binding residues of HLA for TCR complexation are identified as *i)* HLAα2: E152, Q155 and L163 with TCRα: I29α, N31α, R50α and R97α, and *ii)* HLAα1: R62 and I66 with TCRβ: Q53β, A58β, D59β, D60β and T100β. R97α of CDR3 has low stabilization to p4(T) at -1.8 kcal/mol because the attraction is shared by E152 at -5.3 kcal/mol. It can be clearly seen that R62 and L163 residues are observed as the key common strong interactions with TCR in all cases. However, stabilization energy of these residues with TCR protein seems to present a higher magnitude when the number of peptide monomer is increased (9-mer: HIV9p and MICA, 8-mer: HIV8p). In fact, it is fallen in our scope of expectation since R62 and L163 residues are specified as the key residues component for p1 or p2 binding pocket where the binding is initiated [107].

Hydrogen Bonding Interaction

To investigate the detail of binding interaction pattern, the number of hydrogen bond occupation (HB) is determined (using donor–acceptor atoms cutoff distance < 3.5 Å, donor–H–acceptor angle > 120°). The results are separately given in two different characters as: *i)* HB interaction between TCR and HLAp, Figure 32A, and *ii)* HB interaction between HLA and peptide fragment, Figure 32B. In addition, the strength of HB characters are defined in 3 types as weak, moderate and strong occupation with percentage of 10%-39%, 40-69% and 70-100%, respectively.

According to Figure 32A, the total number of strong HB interactions between HLAp and TCR and domains in MICA, HIV9p and HIV8p are shown in 9, 6 and 5 positions while the moderate HB interactions are found in 4, 2 and 4 positions, respectively. It should be noted from the results that the HLA domain plays an important role in TCR/HLAp ternary complex formation indicated by the higher number of HB interactions in comparison with the peptide fragment. This finding is well in accordance with the clinical observation which confirms the dominant role of HLA protein to trap and exhibit the strange antigenic peptide to TCR protein for recognition.

In TCR/HLA-MICA system, the number of HB occupations are observed on both TCRα and TCRβ domains through the interaction network of Q98α-T69, A58β-T69 and D59β-Q65 pairs for strong interaction and Q98α-R62, Q53β-Q72, D59β-R62, D50β-Q65 and S61β-R62 pairs for moderate interaction. Two strong HB occupations in TCR/HLA-HIV9p system are observed through N31α-Q155 pair while the medium HB interactions are observed in quite similar region to TCR/HLA-MICA system (R97α-E152, T55β-Q72, D59β-Q65 and D50β-Q65 pairs). The pattern of HB occupation of HLA-HIV8p system is quite different from both MICA and HIV9p which exhibits N31α-Q155, R97α-E152 for strong interactions and R50α-E154, A58β-Q65, T100β-E152, T100β-Q155 for moderate interactions.

Focusing on the peptide fragment, active residues related to recognition by CDR3 loop of TCR protein (especially on R97α position) are investigated at p4-p7 regions which also supports the "rule of engagement" as well. The unique additional interactions in

MICA peptide fragment are detected between p4(A) and p6(I) regions to TCR: S95α with 87% and T100β with 58% of HB occupation, respectively.
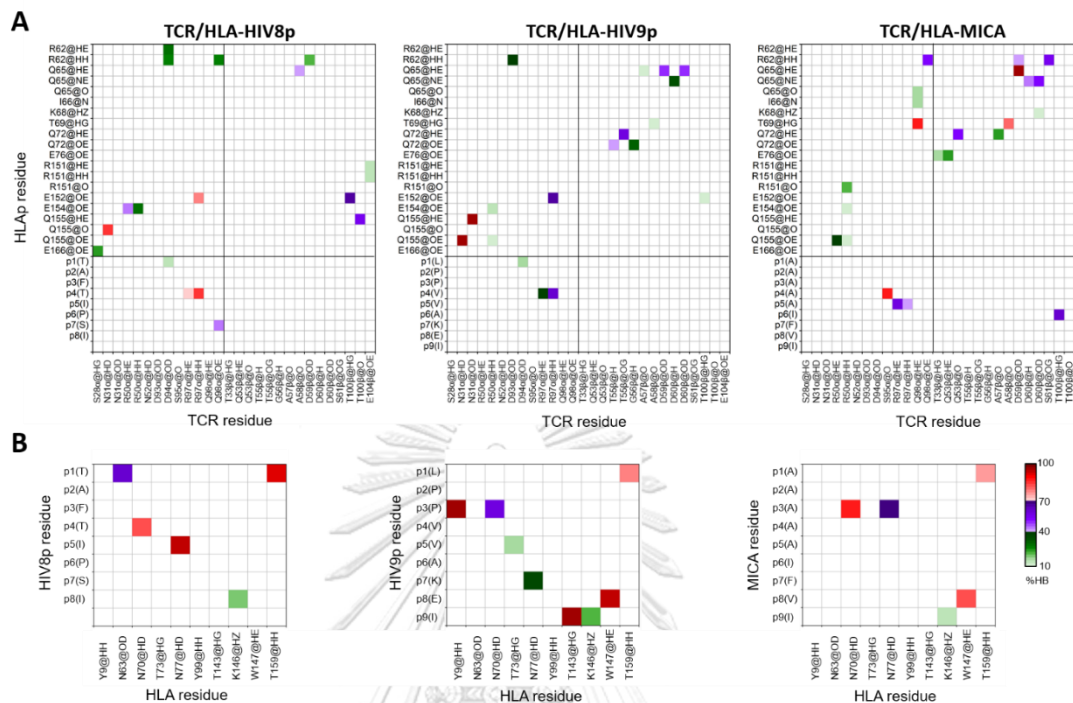
Figure 32. Hydrogen bond (HB) occupations of A) TCR/HLAp and B) HLA/p interactions. HBs atom types are specified by tleap AMBER conversion.

HB formations between HLA and peptide are shown in Figure 32B. Antigenic peptides are commonly mouth at termini residues. HIV9p and MICA show similarly restricted dominant HB at p1, p3 and p8 with T159, Y9/N70 and W147 of HLA groove. A shorter peptide (HIV8p) shows outstanding HB at p1, p4 and p5 with T159, N70 and N77. Termini of nonameric peptides bonded to HLA antigenic presenter, while the middle region is exposed to recognize by TCR. The HB pattern on 8-meric (HIV8p) appears at p4 and p5, instead at the COO⁻-terminus. However, it is still restricted by TCR. In comparison of HB formations between HLA-MICA bound and unbound TCR, the previous research reported p3-Y9, p5-W73, p8-W147 and p9-T143 with high occupation. The effective binding of TCR leads to some HB changing or loss, specifically, p5 changes to bind R97α of TCR.

4.3.5 Binding free energy

Binding free energy (ΔG_{bind}) of TCR with different peptides on HLA is calculated from the energetic component of MM energy (ΔE_{MM}), the solvation free energy (ΔG_{sol}) by the molecular mechanics Poisson-Boltzmann (or Generalized Born) surface area (MM/PB(GB)SA) and the entropic term (TΔS) using normal mode (NMODE) approaches under continuum solvent models. Analysis over the same set of 150 trajectories were performed using the MMPBSA.py script provided in the Amber 14 package [109]. The total energy components are reported in Table S2. In the gas phase, binding affinity order of TCR protein with various HLAp complex in (ΔE_{MM}) is ranked as: MICA (-518 kcal/mol) > HIV9p (-384 kcal/mol) > HIV8p (-373 kcal/mol). Binding affinity order is also observed in the similar character when solvation environment is applied over the systems (MM/PBSA: -81, -73 and -69 kcal/mol and MM/GBSA: -62, -52 and -44 kcal/mol). This i nformation suggests that the HLA binding groove supports 9-meric peptide (MICA and HIV9p) with a stronger interaction than 8-meric peptide (HIV8p) consequently resulting in an increased binding recognition with TCR protein. Based on the binding pattern of HLA class I that mounts at the two terminals with bulging middle region of antigenic peptide, it directly affects the TCR binding. According to HIV8p crystallographic research, it was reported that TCR binding to HLA-B*51:01-HIV8p with very weak affinity of K_D = 81.8 μM or -5.7 kcal/mol [125]. On the other hand, the affinity of most pathogen-specific TCR and HLA class I interaction lies in the region of K_D = 1-10 μM (-6.8 to -8.2 kcal/mol) depending on the amino acid sequence of pathogenic peptide [126]. A longer antigenic peptide sequence might have a closer TCR contact than a shorter one.

The orientation of TCR has a sensitive cross-reactivity of multiple antigens. The free energy landscapes are considered respective to the dihedral angles and tilt angles of the TCR/HLAp complexes. The free energy calculation at each point is performed by Poisson-Boltzmann model and normal mode of entropy. The cpptraj module was applied to compute the dihedral angles and tilt angles. The dihedral angle is measured according to the four centers of gravity (CG) of (a) HLAα1-helix (residue 57-85), (b) HLAα2-helix (residue 138-175), (c) TCRα chain (residue 7-111) and (d) TCRβ chain (residue 5-118), that implies for TCR rotation on the HLAp molecule (Figure 33). A tilt angle is obtained from the three CG of HLAα1-helix, HLAα2-helix and TCRα chain,

or a-b-c angle. For initial structure, the heterodimer of TCR molecule is restricted on HLAp with the dihedral angle of -7⁰ and the tilt angle of 97⁰. Under equilibrium MD stage, dihedral scannings for TCR/HLA-HIV8p, TCR/HLA-HIV9p and TCR/HLA-MICA are respectively manifested in the range of -13⁰ to -0.2⁰, -17⁰ to 3⁰ and -14⁰ to 4⁰. TCR in both HIV peptide complexes produces same tilt angle as found in the crystal structure (-7⁰) as they are found to have tight binding (-56 and -77 kcal/mol) with small cluster. Only TCR/HLA-HIV9p reveals strong interaction around -70 kcal/mol with quite similar TCR rotation in comparison to the initial structure. The global minimum of TCR/HLA-MICA complex is located at (2⁰, 78⁰) with -92 kcal/mol. The overview tilt angle of TCR within HIV8p and HIV9p peptide complexes do not change significantly and TCRα prefers to bind to HLAα2 domain with an angle > 97⁰. In comparison, TCR/HLA-MICA mostly shows sharper angle of a-b-c than the others, that HLAα1 domain is in close contacted with TCRβ chain. However, each structure of three systems has various orientation and local minima energy. Considering the energy profiles, TCR/HLA-HIV8p, TCR/HLA-HIV9p and TCR/HLA-MICA systems show a large diversity of orientations of the local energy of -10, -28 and -38 kcal/mol. The system of MICA peptide is found to have the strongest binding, which can unmistakably report to self-peptide recognition leading to the autoimmune disease. Nevertheless, TCR binding is not specific to MICA peptide (Behçet's disease). A TCR has an adaptive binding for diversity of peptides on the HLA groove. This work could predict and reveal how TCR recognizes self-peptide as a foreign-peptide.
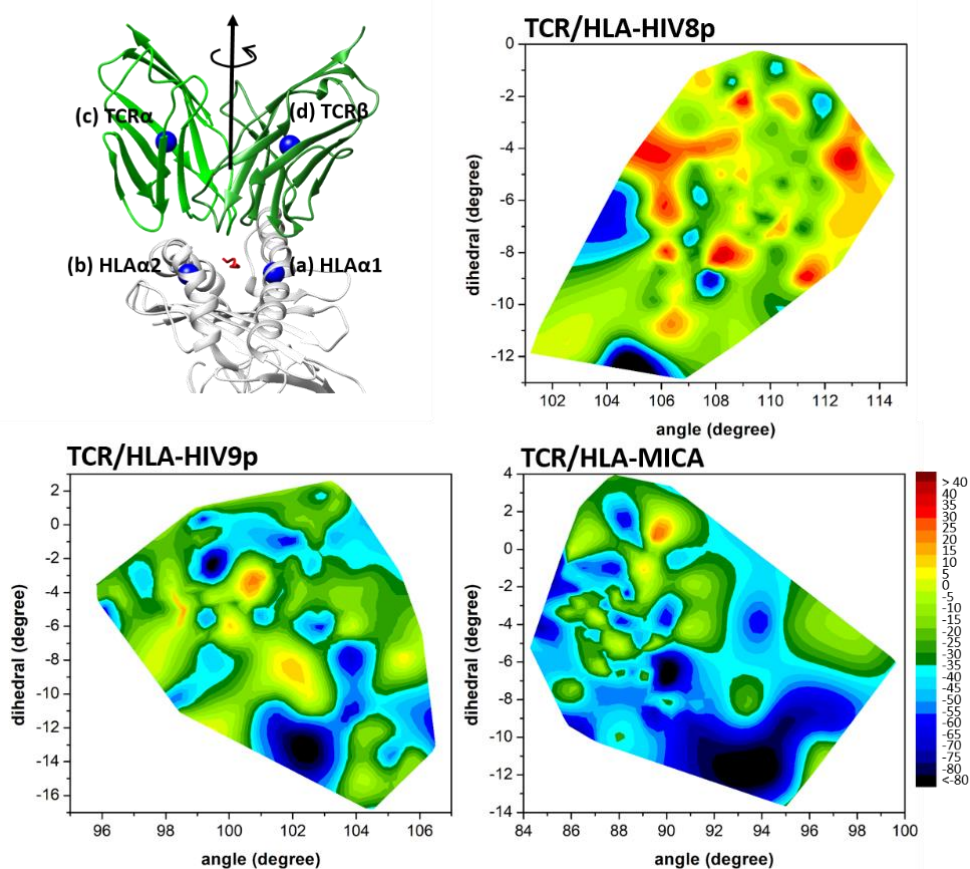
Figure 33. Contour maps displaying MM/PBSA binding free energies (kcal/mol) of TCR/HLAp as a function of the dihedral angle and tilt angle. The center of gravity for (a) HLAα1, (b) HLAα2, (c) TCRα and (d) TCRβ are defined by the blue balls of TCR/HLAp molecule, whereas dihedral angle and tilt angle are measured according to (a)-(b)-(c)-(d) and (a)-(b)-(c), respectively.

CHAPTER V
CONCLUSIONS

Part I: Behçet's disease

BD is a multi-organ inflammatory disorder with vasculitis, in which the main cause and mechanism of action are still not well understood. From clinical investigations, the HLAB51:01 and HLA-A26:01 with MICA009 containing the nonamer peptide (AAAAAIFVI or MICA-TM) are most frequently detected in BD patients. Herein, we aimed to search for the selective correlation between the two BD-associated HLA alleles (HLA-A26:01 and HLAB51:01) and the MICA-TM peptide in comparison with two class matched non-BD-associated HLA alleles (HLA-A11:01 and HLA-B35:01) by MD simulations. From the simulations, more contact residues at the binding groove of the BD-associated HLA alleles stabilized the incoming MICA-TM peptide than at the non-associated alleles of the same class (22 and 16 residues for B51:01 and A26:01; 18 and 14 residues for B35:01 and A11:01). The vdW force was found to be the main protein-protein interaction, but in addition strong H-bonds (>70% occupation) were likely formed with the backbone of the nonpolar peptide, and these were stronger in the BD-associated HLA alleles. The P2/P3 and P9 residues (close to and at the peptide ends, respectively) acted as the anchor for the peptide accommodation at the binding groove of the BD-associated HLAs. The total binding free energy of the HLA/peptide complex suggested a significantly stronger binding strength of the MICA-TM peptide with the BD-associated HLA alleles in the same class (B51:01 > B35:01 and A26:01 > A11:01). The residues 70, 73, 99, 146, 147 and 159 provided the conserved interaction of the two BD-associated HLAs with the MICA-TM peptide. All the structural, dynamics and energetics information somewhat explain the recognition and selective binding of the MICA-TM peptide towards the specific HLAs related to the BD pathogenesis.

Part II: Systemic sclerosis disease

Complexation of HLA/peptide is a prerequisite for T-cell recognition leading to an immune response against antigens. In this study, the five different HLA-DRs with a binding of the self-antigen of Top1 protein (RIANFKIEPPGLFRGRGNHP) for

ATASSc were characterized. From the obtained dynamical behaviors and binding interactions, HLA-DRB1*01:01 was likely unspecific to Top1 self-peptide which was a candidate for auto-antigenic ATASSc. In comparison between the binding clefts of HLA/peptide complexes and identical HLA free forms, the binding of Top1 peptide has led to a more stable protein similar to the ankylosing spondylitis-associated HLAs with bound peptide[127]. A higher flexibility of the whole complex and the opening of the binding cleft for HLA-DRB1*01:01/Top1 (insusceptible ATASSc), while this flexible magnitude was presented in quite smaller values for the other complexes. The determined Top1 rigidity on various binding clefts suggested that HLA (-DRB1*08:02, -DRB1*11:01 and -DRB1*11:04) associated and (-DRB5*01:02) suspect ATASSc could keep a tighter binding to the core self-peptide than that of the non-associated one. The estimation of the total binding free energy of the peptide and HLA interactions based on either MM-PB(GB)SA for the whole Top1 (20 amino acids) or QM/MM-GBSA for the nonameric core peptide treated by QM method showed a binding strength in the order of suspect ATASSc: HLA-DRB5*01:02 > susceptible ATASSc: HLA-DRB1*11:01, HLA-DRB1*11:04, HLA-DRB1*08:02 >> insusceptible ATASSc: HLA-DRB1*01:01. The question how different interactions could influence the binding of HLA-DR/Top1 complexes was investigated. We found that minimal side chain with beta carbon atoms of p1, p4, p6 and p9 properly faced at the HLA-DR interface, while the rest turned to the T-cell receptor recognition in accord with the frequent finding for HLA class II binding. Analysis of energetic per-residue decomposition of Top1 indeed identified the important anchors for p1, p3, p6 and p9. Top1 peptide binding was predominantly contributed by HLA-DRs' nonpolar interactions, which were included into the total binding energy fingerprint of the interface. By investigation of the core nonameric residues of Top1 binding to five different HLA-DRs, the results showed the tendency of p6-p9 binding reduction for HLA-DRB1*01:01 complex based on polymorphic residues at 13, 30, 70 and 74 within pockets. Especially, diversification of hydrophilic residues ($C30_\beta$ and $S37_\beta$) for the HLA-DRB1*01:01 p9-pocket leads to attract water accessibility across protein-protein interface with a consequence of the interruption of Top1-p9 binding. Moreover, the strong hydrogen bond formation between proteins was mostly found in cases of associated and suspected ATASSc. HLA-DRB5*01:02 was reported to have the strongest binding affinity for Top1 peptide

indicating that there should be an appropriate consideration to include this allele into genetic risk for ATASSc development. In our point of views, this information allows us to summarize that HLA-DRs susceptible with ATASSc presents the strong binding affinity for Top1, which well explains the direct relationship between HLA/peptide specific recognition and pathogenicity of SSc disease in a similar manner with the clinically observed information. One of the autoimmunity theories mentioned that the high conservative epitopes between foreign-antigen and self-peptide might be involved in the effective autoimmune disease [27, 128, 129]. Identification of foreign-antigen as well as this Top1 self-peptide sequence is necessary to avoid triggering of ATASSc from pathogenic environment, especially for those who had HLA-DR genetic risk for the SSc disease with ATA.

Part III: T-cell recognition

Both binding interaction pattern and binding free energy of MICA peptide (self-peptide) in BD system was found in a minimal criteria of TCR recognition in HIV infection. The binding interaction of TCR/HLA-MICA complex was presented dominantly between TCRβ and HLAα1 domains, while TCR/HLA-HIV9p had lower stabilization and TCR/HLA-HIV8p shifted to the opposite site. The powerful cross-reactivity of TCR to HLA-MICA molecule was clearly summarized into total binding free energy. This might be the answer to "How does T-cell recognize self-antigen?" and could explain how it develops to the autoimmune disease. However, TCR antigenic binding loops employed different conformations when interacting with various peptides. This finding might support Behçet's disease mechanism leading to a more understanding and guideline on the treatment for autoimmune diseases.

REFERENCES

1.   Benacerraf, B.; McDevitt, H.O. "*Histocompatibility linked immune response genes*" *Science* **1972,** *175*, 273–9.

2.   Rammensee, H.G. "*Chemistry of peptides associated with MHC class I and class II molecules*" *Curr. Opin. Immunol.* **1995,** *7* (1), 85-96.

3.   Parham, P.; Lomen, C.E.; Lawlor, D.A.; Ways, J.P.; Holmes, N.; Coppin, H.L.; Salter, R.D.; Wan, A.M.; Ennis, P.D. "*Nature of polymorphism in HLA-A, -B, and -C molecules*" *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85*, 4005-9.

4.   Nakagawa, M.; Kim, K.H.; Gillam, T.M.; Moscicki, A.B. "*HLA class I binding promiscuity of the CD8 T-cell epitopes of human papillomavirus type 16 E6 protein*" *J. Virol.* **2007,** *81* (3), 1412-23.

5.   Ghosh, P.; Amaya, M.; Mellins, E.; Wiley, D.C. "*The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3*" *Nature* **1995,** *378* (6556), 457-62.

6.   van der Merwe, P.A.; Dushek, O. "*Mechanisms for T cell receptor triggering*" *Nat. Rev. Immunol.* **2011,** *1* (11), 47-55.

7.   Wooldridge, L.; Ekeruche-Makinde, J.; Hugo, A.; van den Berg, H.A.; Skowera, A.; John, J.; Miles, J.J.; Tan, M.P.; Dolton, G.; Clement, M.; Llewellyn-Lacey, S.; Price, D.A.; Peakman, M.; Sewell, A.K. "*A Single autoimmune T cell receptor recognizes more than a million different peptides*" *Journal of biological chemistry* **2012,** *287* (2), 1168 –1177.

8.   Gascoigne, N.R.; Rybakin, V.; Acuto, O.; Brzostek, J. "*TCR Signal Strength and T Cell Development*" *Annu. Rev. Cell Dev. Biol.* **2016,** *32*, 327-48.

9.   Kim, S.T.; Shin, Y.; Brazin, K.; Mallis, R.J.; Sun, Z.Y.; Wagner, G.; Lang, M.J.; Reinherz, E.L. "*TCR mechanobiology: torques and tunable structures linked to earlyT cell signaling*" *Front. Immunol.* **2012,** *3* (76), 1-8.

10.   Wooldridge, L.; Ekeruche-Makinde, J.; van den Berg, H.A.; Skowera, A.; Miles, J.J.; Tan, M.P.; Dolton, G.; Clement, M.; Llewellyn-Lacey, S.; Price, D.A.; Peakman, M.; Sewell, A.K. "*A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides*" *J. Biol. Chem.* **2012,** *287* (2), 1168-77.

11.   Rossjohn, J.; Gras, S.; Miles, J.J.; Turner, S.J.; Godfrey, D.I.; McCluskey, J. "*T cell antigen receptor recognition of antigen-presenting molecules*" *Annu. Rev. Immunol.* **2015,** *33*, 169-200.

12.   Sewell, A.K. "*Why must T cells be cross-reactive?*" *Nat. Rev. Immunol.* **2012,** *12* (9), 669-77.

13.   Luckheeram, R.V.; Zhou, R.; Verma, A.D.; Xia, B. "*CD4(+)T cells: differentiation and functions*" *Clin. Dev. Immunol.* **2012,** *2012*, 925135.

14.   Townsend, A.; Bodmer, H. "*Antigen recognition by class I-restricted T lymphocytes*" *Annu. Rev. Immunol.* **1989,** *7*, 601-24.

15.   Maenaka, K.; Maenaka, T.; Tomiyama, H.; Takiguchi, M.; Stuart, D.I.; Jones, E.Y. "*Nonstandard peptide binding revealed by crystal structures of HLA-B*5101 complexed with HIV immunodominant epitopes*" *J. Immunol.* **2000,** *165* (16), 3260-7.

16.   Cole, D.K.; Pumphrey, N.J.; Boulter, J.M.; Sami, M.; Bell, J.I.; Gostick, E.; Price, D.A.; Gao, G.F.; Sewell, A.K.; Jakobsen, B.K. "*Human TCR-binding affinity is governed by MHC class restriction*" *J. Immunol.* **2007,** *178* (9), 5727-34.

17.   Poletaev, A.B.; Churilov, L.P.; Stroev, Y.I.; Agapov, M.M. "*Immunophysiology versus immunopathology: Natural autoimmunity in human health and disease*" *Pathophysiology* **2012,** *19* (3), 221-31.

18.   Colf, L.A.; Bankovich, A.J.; Hanick, N.A.; Bowerman, N.A.; Jones, L.L.; Kranz, D.M.; Garcia, K.C. "*How a single T cell receptor recognizes both self and foreign MHC*" *Cell* **2007,** *129* (1), 135-46.

19.   Wolfson, M.Y.; Nam, K.; Chakraborty, A.K. "*The effect of mutations on the alloreactive T cell receptor/peptide-MHC interface structure: a molecular dynamics study*" *J. Phys. Chem. B* **2011,** *115* (25), 8317-27.

20.   Cuendet, M.A.; Zoete, V.; Michielin, O. "*How T cell receptors interact with peptide-MHCs: a multiple steered molecular dynamics study*" *Proteins* **2011,** *79* (11), 3007-24.

21.   Yin, L.; Huseby, E.; Scott-Browne, J.; Rubtsova, K.; Pinilla, C.; Crawford, F.; Marrack, P.; Dai, S.; Kappler, J.W. "*A single T cell receptor bound to major histocompatibility complex class I and class II glycoproteins reveals switchable TCR conformers*" *Immunity* **2011,** *35* (1), 23-33.

22.   Marshall, S.E. "*Behçet's disease*" *Best Pract. Res. Clin. Rheumatol.* **2004,** *18* (3), 291–311.

23.   Kongkaew, S.; Yotmanee, P.; Rungrotmongkol, T.; Kaiyawet, N.; Meeprasert, A.; Kaburaki, T.; Noguchi, H.; Takeuchi, F.; Kungwan, N.; Hannongbua, S. "*Molecular Dynamics Simulation Reveals the Selective Binding of Human Leukocyte Antigen Alleles Associated with Behçet's Disease*" *PLoS One* **2015,** *10* (9), e0135575.

24.   Yasuoka, H.; Okazaki, Y.; Kawakami, Y.; Hirakata, M.; Inoko, H.; Ikeda, Y.; Kuwana, M. "*Autoreactive CD8+ cytotoxic T lymphocytes to major histocompatibility complex class I chain-related gene A in patients with Behcet's disease*" *Arthritis Rheum.* **2004,** *50* (11), 3658-62.

25.   Ombrello, M.J.; Kirino, Y.; de Bakker, P.I.; Gül, A.; Kastner, D.L.; Remmers, E.F. "*Behçet disease-associated MHC class I residues implicate antigen binding and regulation of cell-mediated cytotoxicity*" *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (24), 8867-72.

26.   Yasuoka, H.; Okazaki, Y.; Kawakami, Y.; Hirakata, M.; Inoko, H.; Ikeda, Y.; Kuwana, M. "*Autoreactive CD8+ cytotoxic T lymphocytes to major histocompatibility complex class I chain–related gene A in patients with Behçet's disease*" *Arthritis Rheum.* **2004,** *50* (11), 3658–62.

27.   Cusick, M.F.; Libbey, J.E.; Fujinami, R.S. "*Molecular Mimicry as a Mechanism of Autoimmune Disease*" *Clin. Rev. Allergy Immunol.* **2012,** *42* (11), 102–11.

28.   Saadoun, D.; Wechsler, B.; Desseaux, K.; Le Thi Huong, D.; Amoura, Z.; Resche-Rigon, M.; Cacoub, P. "*Mortality in Behçet's disease*" *Arthritis Rheum.* **2010,** *62* (9), 2806-12.

29.   Kaya, T.I. "*Genetics of Behçet's disease*" *Patholog. Res. Int.* **2012,** *2012*.

30.   Kaburaki, T.; Araki, F.; Takamoto, M.; Okinaga, K.; Yoshida, A.; Numaga, J.; Fujino, Y.; Kawashima, H. "*Best-corrected visual acuity and frequency of ocular attacks during the initial 10 years in patients with Behçet's disease*" *Graefes. Arch. Clin. Exp. Ophthalmol.* **2010,** *248* (5), 709-14.

31.   Hughes, T.; Coit, P.; Adler, A.; Yilmaz, V.; Aksu, K.; Düzgün, N.; Keser, G.; Cefle, A.; Yazici, A.; Ergen, A.; Alpsoy, E.; Salvarani, C.; Casali, B.; Kötter, I.; Gutierrez-Achury, J.; Wijmenga, C.; Direskeneli, H.; Saruhan-Direskeneli, G.;

Sawalha, A.H. "*Identification of multiple independent susceptibility loci in the HLA region in Behçet's disease*" *Nat. Genet.* **2013,** *45* (3), 319–24.

32.   Kaneko, F.; Oyama, N.; Nishibu, A. "*Streptococcal infection in the pathogenesis of Behçet's disease and clinical effects of minocycline on the disease symptoms*" *Yonsei. Med. J.* **1997,** *38* (6), 444-54.

33.   Tanaka, T.; Yamakawa, N.; Koike, N.; Suzuki, J.; Mizuno, F.; Usui, M. "*Behçet's disease and antibody titers to various heat-shock protein 60s*" *Ocul. Immunol. Inflamm.* **1999,** *7* (2), 69-74.

34.   Chambers, J.C.; Haskard, D.O.; Kooner, J.S. "*Vascular endothelial function and oxidative stress mechanisms in patients with Behçet's syndrome*" *J. Am. Coll. Cardiol.* **2001,** *37* (2), 517-20.

35.   Remmers, E.F.; Cosan, F.; Kirino, Y.; Ombrello, M.J.; Abaci, N.; Satorius, C.; Le, J.M.; Yang, B.; Korman, B.D.; Cakiris, A.; Aglar, O.; Emrence, Z.; Azakli, H.; Ustek, D.; Tugal-Tutkun, I.; Akman-Demir, G.; Chen, W.; Amos, C.I.; Dizon, M.B.; Kose, A.A.; Azizlerli, G.; Erer, B.; Brand, O.J.; Kaklamani, V.G.; Kaklamanis, P.; Ben-Chetrit, E.; Stanford, M.; Fortune, F.; Ghabra, M.; Ollier, W.E.; Cho, Y.H.; Bang, D.; O'Shea, J.; Wallace, G.R.; Gadina, M.; Kastner, D.L.; Gül, A. "*Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease*" *Nat. Genet.* **2010,** *42* (8), 698-702.

36.   Ohno, S.; Ohguchi, M.; Hirose, S.; Matsuda, H.; Wakisaka, A.; Aizawa, M. "*Close association of HLA-Bw51 with Behçet's disease*" *Arch. Ophthalmol.* **1982,** *100* (9), 1455-8.

37.   Mizuki, N.; Inoko, H.; Ando, H.; Nakamura, S.; Kashiwase, K.; Akaza, T.; Fujino, Y.; Masuda, K.; Takiguchi, M.; Ohno, S. "*Behçet's disease associated with one of the HLA-B51 subantigens, HLA-B*5101*" *Am. J. Ophthalmol.* **1993** *116* (4), 406-9.

38.   de Menthon, M.; Lavalley, M.P.; Maldini, C.; Guillevin, L.; Mahr, A. "*HLA-B51/B5 and the risk of Behçet's disease: a systematic review and meta-analysis of case-control genetic association studies*" *Arthritis Rheum.* **2009,** *61* (10), 1287-96.

39.   Pineton de Chambrun, M.; Wechsler, B.; Geri, G.; Cacoub, P.; Saadoun, D. "*New insights into the pathogenesis of Behçet's disease*" *Autoimmun. Rev.* **2012,** *11* (10), 687-98.

40.   Ergun, T.; Ince, U.; Ekşioğlu-Demiralp, E.; Direskeneli, H.; Gürbüz, O.; Gürses, L.; Aker, F.; Akoğlu, T. "*HSP 60 expression in mucocutaneous lesions of Behçet's disease*" *J. Am. Acad. Dermatol.* **2001,** *45* (6), 904-9.

41.   Mizuki, N.; Meguro, A.; Tohnai, I.; Gül, A.; Ohno, S.; Mizuki, N. "*Association of major histocompatibility complex class I chain-related gene a and HLA-B alleles with Behçet's disease in Turkey*" *Jpn. J. Ophthalmol.* **2007,** *51* (6), 431-6.

42.   Mizuki, N.; Ota, M.; Katsuyama, Y.; Yabuki, K.; Ando, H.; Goto, K.; Nakamura, S.; Bahram, S.; Ohno, S.; Inoko, H. "*Association analysis between the MIC-A and HLA-B alleles in Japanese patients with Behçet's disease*" *Arthritis Rheum.* **1999,** *42* (9), 1961-6.

43.   Mizuki, N.; Ota, M.; Kimura, M.; Ohno, S.; Ando, H.; Katsuyama, Y.; Yamazaki, M.; Watanabe, K.; Goto, K.; Nakamura, S.; Bahram, S.; Inoko, H. "*Triplet repeat polymorphism in the transmembrane region of the MICA gene: a strong association of six GCT repetitions with Behçet disease*" *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94* (4), 1298-303.

44.   Wallace, G.R.; Verity, D.H.; Delamaine, L.J.; Ohno, S.; Inoko, H.; Ota, M.; Mizuki, N.; Yabuki, K.; Kondiatis, E.; Stephens, H.A.; Madanat, W.; Kanawati, C.A.; Stanford, M.R.; Vaugha, R.W. "*MIC-A allele profiles and HLA class I associations in Behçet's disease*" *Immunogenetics* **1999,** *49* (7-8), 613-7.

45.   Lehner, T. "*Immunopathogenesis of Behçet's disease*" *Ann. Med. Interne. (Paris)* **1999,** *150* (6), 483-7.

46.   Baharav, E.; Weinberger, A. "*The HLA-B*5101 Molecule-Binding Capacity to Antigens Used in Animal Models of Behçet's Disease: A Bioinformatics Study*" *Isr. Med. Assoc. J.* **2012,** *14* (7), 424-8.

47.   Alvarez-Márquez, A.; Aguilera, I.; Gentil, M.A.; Caro, J.L.; Bernal, G.; Fernández Alonso, J.; Acevedo, M.J.; Cabello, V.; Wichmann, I.; Gonzalez-Escribano, M.F.; Núñez-Roldán, A. "*Donor-specific antibodies against HLA, MICA, and GSTT1 in patients with allograft rejection and C4d deposition in renal biopsies*" *Transplantation* **2009,** *87* (1), 94-9.

48.   Terasaki, P.I.; Ozawa, M.; Castro, R. "*Four-year follow-up of a prospective trial of HLA and MICA antibodies on kidney graft survival*" *Am. J. Transplant* **2007,** *7* (2), 408-15.

49.   Suárez-Alvarez, B.; López-Vázquez, A.; Díaz-Peña, R.; Díaz-Molina, B.; Blanco-García, R.M.; Alvarez-López, M.R.; López-Larrea, C. "*Post-transplant soluble MICA and MICA antibodies predict subsequent heart graft outcome*" *Transpl. Immunol.* **2006,** *17* (1), 43-6.

50.   Yabuki, K.; Mizuki, N.; Ota, M.; Katsuyama, Y.; Palimeris, G.; Stavropoulos, C.; Koumantaki, Y.; Spyropoulou, M.; Giziaki, E.; Kaklamani, V.; Kaklamani, E.; Inoko, H.; Ohno, S. "*Association of MICA gene and HLA-B*5101 with Behçet's disease in Greece*" *Invest. Ophthalmol. Vis. Sci.* **1999,** *40*, 9.

51.   Kaya, T.I.; Dur, H.; Tursen, U.; Gurler, A. "*Association of class I HLA antigens with the clinical manifestations of Turkish patients with Behcet's disease*" *Clin. Exp. Dermatol.* **2002,** *27* (6), 498-501.

52.   Mizuki, N.; Ohno, S.; Ando, H.; Chen, L.; Palimeris, G.D.; Stavropoulos-Ghiokas, E.; Ishihara, M.; Goto, K.; Nakamura, S.; Shindo, Y.; Isobe, K.; Ito, N.; Inoko, H. "*A strong association between HLA-B*5101 and Behçet's disease in Greek patients*" *Tissue Antigens* **1997,** *50* (1), 57-60.

53.   Itoh, Y.; Inoko, H.; Kulski, J.K.; Sasaki, S.; Meguro, A.; Takiyama, N.; Nishida, T.; Yuasa, T.; Ohno, S.; Mizuki, N. "*Four-digit allele genotyping of the HLA-A and HLA-B genes in Japanese patients with Behcet's disease by a PCR-SSOP-Luminex method*" *Tissue Antigens* **2006,** *67* (5), 390-4.

54.   Chung, Y.M.; Yeh, T.S.; Sheu, M.M.; Chen, M.S.; Wen, M.S.; Tsai, H.Y.; Liao, H.R. "*Behcet's disease with ocular involvement in Taiwan: a joint survey of six major ophthalmological departments*" *J. Formos. Med. Assoc.* **1990,** *89* (5), 413-7.

55.   Kurata, R.; Yonezawa, T.; Inoko, H. "*Association analysis of the polymorphism of human leukocyte antigen-A, -B and -E gene with Behcet's disease in Japanese cohort using sequencing-based typing method*" *MOJ Immunol.* **2014,** *1* (3), 00013.

56.   Kaburaki, T.; Takamoto, M.; Numaga, J.; Kawashima, H.; Araie, M.; Ohnogi, Y.; Harihara, S.; Kuwata, S.; Takeuchi, F. "*Genetic association of HLA-A*2601 with*

ocular Behçet's disease in Japanese patients" Clin. Exp. Rheumatol. **2010,** *4 Suppl 60,* S39-44.

57.   Arnett, F.C.; Gourh, P.; Shete, S.; Ahn, C.W.; Honey, R.E.; Agarwal, S.K.; Tan, F.K.; McNearney, T.; Fischbach, M.; Fritzler, M.J.; Mayes, M.D.; Reveille, J.D. "*Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls*" Ann. Rheum. Dis. **2010,** *69* (5), 822–7.

58.   Jimenez, S.A. "*Role of endothelial to mesenchymal transition in the pathogenesis of the vascular alterations in systemic sclerosis*" ISRN Rheumatol. **2013,** *2013*, 1-15.

59.   Nihtyanova, S.I.; Schreiber, B.E.; Ong, V.H.; Rosenberg, D.; Moinzadeh, P.; Coghlan, J.G.; Wells, A.U.; Denton, C.P. "*Prediction of pulmonary complications and long-term survival in systemic sclerosis*" Arthritis Rheumatol. **2014,** *66* (6), 1625-35.

60.   Chung, L.; Domsic, R.T.; Lingala, B.; Alkassab, F.; Bolster, M.; Csuka, M.E.; Derk, C.; Fischer, A.; Frech, T.; Furst, D.E.; Gomberg-Maitland, M.; Hinchcliff, M.; Hsu, V.; Hummers, L.K.; Khanna, D.; Medsger, T.A.J.; Molitor, J.A.; Preston, I.R.; Schiopu, E.; Shapiro, L.; Silver, R.; Simms, R.; Varga, J.; Gordon, J.K.; Steen, V.D. "*Survival and predictors of mortality in systemic sclerosis-associated pulmonary arterial hypertension: outcomes from the pulmonary hypertension assessment and recognition of outcomes in scleroderma registry*" Arthritis Care Res. (Hoboken) **2014,** *66* (3), 489-95.

61.   Rubio-Rivas, M.; Royo, C.; Simeón, C.P.; Corbella, X.; Fonollosa, V. "*Mortality and survival in systemic sclerosis: Systematic review and meta-analysis*" Semin. Arthritis Rheum. **2014,** *44* (2), 208-19.

62.   Rodriguez-Reyna, T.S.; Mercado-Velazquez, P.; Yu, N.; Alosco, S.; Ohashi, M.; Lebedeva, T.; Cruz-Lagunas, A.; Nunez-Alvarez, C.; Cabiedes-Contreras, J.; Vargas-Alarcon, G.; Granados, J.; Zuniga, J.; Yunis, E. "*HLA Class I and II Blocks Are Associated to Susceptibility, Clinical Subtypes and Autoantibodies in Mexican Systemic Sclerosis (SSc) Patients*" PLoS One **2015,** *10* (5), e0126727.

63.   Denton, C.P.; Hachulla, E. "*Risk factors associated with pulmonary arterial hypertension in patients with systemic sclerosis and implications for screening*" *Eur. Respir. Rev.* **2011,** *20* (122), 270-6.

64.   Ciechomska, M.; van Laar, J.; O'Reilly, S. "*Current frontiers in systemic sclerosis pathogenesis*" *Exp. Dermatol.* **2015,** *24* (6), 401-6.

65.   Pattanaik, D.; Brown, M.; Postlethwaite, B.C.; Postlethwaite, A.E. "*Pathogenesis of Systemic Sclerosis*" *Front. Immunol.* **2015,** *6*, 272.

66.   LeRoy, E.C.; Black, C.; Fleischmajer, R.; Jablonska, S.; Krieg, T.; Medsger, T.A.J.; Rowell, N.; Wollheim, F. "*Scleroderma (systemic sclerosis): classification, subsets and pathogenesis*" *J. Rheumatol.* **1988,** *15* (2), 202-5.

67.   Gabrielli, A.; Avvedimento, E.V.; Krieg, T. "*Scleroderma*" *N. Engl. J. Med.* **2009,** *360* (19), 1989-2003.

68.   Berger, M.; Steen, V.D. "*Role of anti-receptor autoantibodies in pathophysiology of scleroderma*" *Autoimmun. Rev.* **2017,** *16* (10), 1029-35.

69.   Foocharoen, C.; Watcharenwong, P.; Netwijitpan, S.; Mahakkanukrauh, A.; Suwannaroj, S.; Nanagara, R. "*Relevance of clinical and autoantibody profiles in systemic sclerosis among Thais*" *Int. J. Rheum. Dis.* **2017,** *20* (10), 1572-81.

70.   Douvas, A. "*Doses Scl-70 modulate collagen production in systemic sclerosis?*" *Lancet* **1988,** *2* (8609), 475-7.

71.   Scheja, A.; Wildt, M.; Wollheim, F.A.; Akesson, A.; Saxne, T. "*Circulating collagen metabolites in systemic sclerosis. Differences between limited and diffuse form and relationship with pulmonary involvement*" *Rheumatology (Oxford)* **2000,** *39* (10), 1110-3.

72.   Rizou, C.; Ioannidis, J.P.; Panou-Pomonis, E.; Sakarellos-Daitsiotis, M.; Sakarellos, C.; Moutsopoulos, H.M.; Vlachoyiannopoulos, P.G. "*B-Cell epitope mapping of DNA topoisomerase I defines epitopes strongly associated with pulmonary fibrosis in systemic sclerosis*" *Am. J. Respir. Cell Mol. Biol.* **2000,** *22* (3), 344-51.

73.   Redinbo, M.R.; Champoux, J.J.; Hol, W.G. "*Novel insights into catalytic mechanism from a crystal structure of human topoisomerase I in complex with DNA*" *Biochemistry* **2000,** *39* (23), 6832-40.

74.   Reveille, J.D.; Durban, E.; MacLeod-St Clair, M.J.; Goldstein, R.; Moreda, R.; Altman, R.D.; FC, A. "*Association of amino acid sequences in the HLA-DQB1 first domain with antitopoisomerase I autoantibody response in scleroderma (progressive systemic sclerosis)*" *J. Clin. Invest.* **1992,** *90* (3), 973-80.

75.   Takeuchi, F.; Nakano, K.; Yamada, H.; Hong, G.H.; Nabeta, H.; Yoshida, A.; Matsuta, K.; Bannai, M.; Tokunaga, K.; Ito, K. "*Association of HLA-DR with progressive systemic sclerosis in Japanese*" *J. Rheumatol.* **1994,** *21* (5), 857-63.

76.   Takeuchi, F.; Nabeta, H.; Füssel, M.; Conrad, K.; Frank, K.H. "*Association of the TNFa13 microsatellite with systemic sclerosis in Japanese patients*" *Ann. Rheum. Dis.* **2000,** *59* (4), 293-6.

77.   Radstake, T.R.D.J.; Gorlova, O.; Rueda, B.; Martin, J.-E.; Alizadeh, B.Z.; Palomino-Morales, R.; Mayes, M.D. "*Genome-wide association study of systemic sclerosis identifies CD247 as a novel susceptibility locus*" *Nat. Genet.* **2010,** *42* (5), 426–429.

78.   Kang, S.H.; Park, M.H.; Song, E.Y.; Kang, S.J.; Lee, E.B.; Song, Y.W.; Takeuchi, F. "*Association of HLA class II genes with systemic sclerosis in Koreans*" *J. Rheumatol.* **2001,** *28* (7), 1577-83.

79.   Hu, P.Q.; Fertig, N.; Medsger, T.A., Jr.; Wright, T.M. "*Molecular recognition patterns of serum anti-DNA topoisomerase I antibody in systemic sclerosis*" *J. Immunol.* **2004,** *173* (4), 2834-41.

80.   Simeon, C.P.; Fonollosa, V.; Tolosa, C.; Palou, E.; Selva, A.; Solans, R.; Armadans, L.; Moreno, E.; Marsal, S.; Vilardell, M. "*Association of HLA class II genes with systemic sclerosis in Spanish patients*" *J. Rheumatol.* **2009,** *36* (12), 2733-6.

81.   Hasegawa, M.; Imura-Kumada, S.; Matsushita, T.; Hamaguchi, Y.; Fujimoto, M.; Takehara, K. "*Anti-topoisomerase I antibody levels as serum markers of skin sclerosis in systemic sclerosis*" *J. Dermatol.* **2013,** *40* (2), 89-93.

82.   Kuwana, M.; Kaburaki, J.; Arnett, F.C.; Howard, R.F.; Medsger, T.A.J.; Wright, T.M. "*Influence of ethnic background on clinical and serologic features in patients with systemic sclerosis and anti-DNA topoisomerase I antibody*" *Arthritis Rheum.* **1999,** *42*, 465–74.

83.   Beretta, L.; Rueda, B.; Marchini, M.; Santaniello, A.; Simeon, C.P.; Fonollosa, V.; Caronni, M.; Rios-Fernandez, R.; Carreira, P.; Rodriguez-Rodriguez, L.; Spanish Systemic Sclerosis, G.; Moreno, A.; Lopez-Nevot, M.A.; Escalera, A.; Gonzalez-Escribano, M.F.; Martin, J.; Scorza, R. "*Analysis of Class II human leucocyte antigens in Italian and Spanish systemic sclerosis*" *Rheumatology (Oxford)* **2012,** *51* (1), 52-9.

84.   Louthrenoo, W.; Kasitanon, N.; Wichainun, R.; Wangkaew, S.; Sukitawut, W.; Ohnogi, Y.; Nakaue, N.; Kuwata, S.; Takeuchi, F. "*Association of HLA-DRB1*15:02 and DRB5*01:02 allele with the susceptibility to systemic sclerosis in Thai patients*" *Rheumatol. Int.* **2013,** *33* (8), 2069-77.

85.   He, D.; Wang, J.; Yi, L.; Guo, X.; Guo, S.; Guo, G.; Tu, W.; Wu, W.; Yang, L.; Xiao, R.; Li, Y.; Chu, H.; Lai, S.; Jin, L.; Zou, H.; Reveille, J.D.; Assassi, S.; Mayes, M.D.; Zhou, X. "*Association of the HLA-DRB1 with scleroderma in Chinese population*" *PLoS One* **2014,** *9* (9), e106939.

86.   Azzouz, D.F.; Rak, J.M.; Fajardy, I.; Allanore, Y.; Tiev, K.P.; Farge-Bancel, D.; Martin, M.; Kanaan, S.B.; Pagni, P.P.; Hachulla, E.; Harlé, J.R.; Didelot, R.; Granel, B.; Cabane, J.; Roudier, J.; Lambert, N.C. "*Comparing HLA shared epitopes in French Caucasian patients with scleroderma*" *PLoS One* **2012,** *7* (5), e36870.

87.   Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. "*A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*" *J. Comput. Chem.* **2003,** *24* (16), 1999-2012.

88.   Nutho, B.; Khuntawee, W.; Rungnim, C.; Pongsawasdi, P.; Wolschann, P.; Karpfen, A.; Kungwan, N.; Rungrotmongkol, T. "*Binding mode and free energy prediction of fisetin/beta-cyclodextrin inclusion complexes*" *Beilstein J. Org. Chem.* **2014,** *10*, 2789-99.

89.   Sangpheak, W.; Khuntawee, W.; Wolschann, P.; Pongsawasdi, P.; Rungrotmongkol, T. "*Enhanced stability of a naringenin/2,6-dimethyl beta-cyclodextrin inclusion complex: molecular dynamics and free energy calculations based on MM- and QM-PBSA/GBSA*" *J. Mol. Graph. Model.* **2014,** *50*, 10-5.

90.   Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. "*A self-consistent charge density-functional based tight-binding scheme for large biomolecules*" *Physica Status Solidi (b)* **2000,** *217* (357), 357–76.

91.   Moreira, I.S.; Fernandes, P.A.; Ramos, M.J. "*Computational Alanine Scanning Mutagenesis—An Improved Methodological Approach*" *J. Comput. Chem.* **2006,** *28* (3), 644-54.

92.   da Costa, L.M.; Hayaki, S.; Stoyanov, S.R.; Gusarov, S.; Tan, X.; Gray, M.R.; Stryker, J.M.; Tykwinski, R.; Carneiro, J.W.; Sato, H.; Seidl, P.R.; Kovalenko, A. "*3D-RISM-KH molecular theory of solvation and density functional theory investigation of the role of water in the aggregation of model asphaltenes*" *Phys. Chem. Chem. Phys.* **2012,** *14* (11), 3922-34.

93.   Blicher, T.; Kastrup, J.S.; Buus, S.; Gajhede, M. "*High-resolution structure of HLA-A*1101 in complex with SARS nucleocapsid peptide*" *Acta Crystallogr. D Biol. Crystallogr.* **2005,** *61* (Pt 8), 1031-40.

94.   Menssen, R.; Orth, P.; Ziegler, A.; Saenger, W. "*Decamer-like conformation of a nona-peptide bound to HLA-B*3501 due to non-standard positioning of the C terminus*" *J. Mol. Biol.* **1999,** *285* (2), 645-53.

95.   Madrigal, J.A.; Hildebrand, W.H.; Belich, M.P.; Benjamin, R.J.; Little, A.M.; Zemmour, J.; Ennis, P.D.; Ward, F.E.; Petzl-Erler, M.L. "*Structural diversity in the HLA-A10 family of alleles: correlations with serology*" *Tissue Antigens* **1993,** *41* (2), 72-80.

96.   Case, D.A.; Darden, T.A.; III Cheatham, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Mathews, G.C.D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A., *AMBER10*. University of California: San Francisco, 2008.

97.   Olsson, M.H.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. "*PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions*" *J. Chem. Theory Comput.* **2011,** *2*, 525-37.

98.   Guillot, B. "*A reappraisal of what we have learnt during three decades of computer simulations on water*" *J. Mol. Liq.* **2002,** *101* (1-3), 219–60.

99.   Berendsen, H.J.C.; Postma, J.P.M.; van Gunsteren, W.F.; DiNola, A.; Haak, J.R. "*Molecular dynamics with coupling to an external bath*" *J. Chem. Phys.* **1984,** *81*, 3684–90.

100.   Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. "*Numerical integration of Cartesian equations of motion of system with constraints: molecular dynamics of n-alkanes*" *J. Comput. Phys.* **1997,** *23* (3), 327-41.

101.   Darden, T.; York, D.; Pedersen, L. "*Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*" *J. Chem. Phys.* **1993,** *98* (10089), 10089-92.

102.   Jonsson, A.K.; Andersson, L.; Rask, L. "*A cellular and functional split in the DRw8 haplotype is due to a single amino acid replacement (DR beta ser 57- asp 57)*" *Immunogenetics* **1989,** *29* (5), 308-16.

103.   Bergstrom, T.F.; Josefsson, A.; Erlich, H.A.; Gyllensten, U. "*Recent origin of HLA-DRB1 alleles and implications for human evolution*" *Nat. Genet.* **1998,** *18* (3), 237-42.

104.   Longhi, E.; Frison, S.; Andreini, E.; Favoino, B.; Poli, F.; Scalamogna, M. "*Description of a new HLA-DRB1*1104, DRB1*110403*" *Tissue Antigens* **2005,** *66* (6), 700-1.

105.   Ozaki, Y.; Suzuki, S.; Shigenari, A.; Okudaira, Y.; Kikkawa, E.; Oka, A.; Ota, M.; Mitsunaga, S.; Kulski, J.K.; Inoko, H.; Shiina, T. "*HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput sequencing*" *Tissue Antigens* **2014,** *83* (1), 10-6.

106.   Li, Y.; Li, H.; Martin, R.; Mariuzza, R.A. "*Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins*" *J. Mol. Biol.* **2000,** *304* (2), 177-88.

107.   Holland, C.J.; Cole, D.K.; Godkin, A. "*Re-Directing CD4(+) T Cell Responses with the Flanking Residues of MHC Class II-Bound Peptides: The Core is Not Enough*" *Front. Immunol.* **2013,** *4* (172).

108.   Murthy, V.L.; Stern, L.J. "*The class II MHC protein HLA-DR1 in complex with an endogenous peptide: implications for the structural basis of the specificity of peptide binding*" *Structure* **1997,** *5* (10), 1385-96.

109.   Case, D.A.; Babin, V.; Berryman, J.T.; Betz, R.M.; Cai, Q.; Cerutti, D.S.; TCheatham, I., .E.; Darden, T.A.; Duke, R.E.; Gohlke, H.; Goetz, A.W.; Gusarov, S.;

omeyer, N.H.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.M.; Paesani, F.; D.R. Roe; A. Roitberg; C. Sagui; R. Salomon-Ferrer; G. Seabra; C.L. Simmerling; W. Smith; J. Swails; R.C. Walker; J. Wang; R.M. Wolf; X. Wu; P.A. Kollman *AMBER 14*, University of California: San Francisco, 2014.

110.   Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. "*Numerical integration of Cartesian equations of motion of system with constraints: molecular dynamics of n-alkanes*" *J. Comput. Phys.* **1997,** *23* (3), 327-41.

111.   Miyamoto, S.; Kollman, P.A. "*SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models*" *J. Comput. Chem.* **1992,** *13*, 952–62.

112.   Grubmüller, H.; Heller, H.; Windemuth, A.; Schulten, K. "*Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions*" *Mol. Simulat.* **1991,** *6*, 121–42.

113.   Ren, P.; Ponder, J.W. "*Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations*" *J. Comput. Chem.* **2002,** *23*, 1497–506.

114.   Decha, P.; Rungrotmongkol, T.; Intharathep, P.; Malaisree, M.; Aruksakunwong, O.; Laohpongspaisan, C.; Parasuk, V.; Sompornpisut, P.; Pianwanit, S.; Kokpol, S.; Hannongbua, S. "*Source of high pathogenicity of an avian influenza virus H5N1: why H5 is better cleaved by furin*" *Biophys. J.* **2008,** *95* (1), 128-3.

115.   Macdonald, I.K.; Harkiolaki, M.; Hunt, L.; Connelley, T.; Carroll, A.V. "*MHC class I bound to an immunodominant Theileria parva epitope demonstrates unconventional presentation to T cell receptors*" *PLoS Pathog.* **2010,** *6* (10), e1001149.

116.   Holland, C.J.; Cole, D.K.; Godkin, A. "*Re-directing CD4(+) T cell responses with the flanking residues of MHC class II-bound peptides: The core is not enough*" *Front Immunol* **2013,** *4* (172), 1-9.

117.   Nunthaboot, N.; Rungrotmongkol, T.; Malaisree, M.; Kaiyawet, N.; Decha, P.; Sompornpisut, P.; Poovorawan, Y.; Hannongbua, S. "*Evolution of human receptor binding affinity of H1N1 hemagglutinins from 1918 to 2009 pandemic influenza A virus*" *J. Chem. Inf. Model.* **2010,** *50* (8), 1410-7.

118. Rungrotmongkol, T.; Nunthaboot, N.; Malaisree, M.; Kaiyawet, N.; Yotmanee, P.; Meeprasert, A.; Hannongbua, S. "*Molecular insight into the specific binding of ADP-ribose to the nsP3macro domains of chikungunya and venezuelan equine encephalitis viruses: Molecular dynamics simulations and free energy calculations*" *J. Mol. Graphics. Model.* **2010,** *29*, 347−53.

119. Wang, Q.; Pang, Y.P. "*Normal-mode-analysis-monitored energy minimization procedure for generating small-molecule bound conformations*" *PLoS One* **2007,** *2* (10), e1025.

120. Kovalenko, A.; Hirata, F. "*Three-dimensional density profiles of water in contact with a solute of arbitrary shape: a RISM approach*" *Chem. Phys. Lett.* **1998,** *290* (1-3), 237–44.

121. Phanich, J.; Rungrotmongkol, T.; Sindhikara, D.; Phongphanphanee, S.; Yoshida, N.; Hirata, F.; Kungwan, N.; Hannongbua, S. "*A 3D-RISM/RISM study of the oseltamivir binding efficiency with the wild-type and resistance-associated mutant forms of the viral influenza B neuraminidase*" *Protein Sci.* **2016,** *25* (1), 147-58.

122. Melo, R.C.; Ribeiro, C.; Murray, C.S.; Veloso, C.J.; da Silveira, C.H.; Neshich, G.; Meira, W.J.; Carceroni, R.L.; Santoro, M.M. "*Finding Protein-Protein Interaction Patterns by Contact Map Matching*" *Genet. Mol. Res.* **2007,** *6* (4), 946-63.

123. Gianni, S.; Ivarsson, Y.; De Simone, A.; Travaglini-Allocatelli, C.; Brunori, M.; Vendruscolo, M. "*Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain*" *Nat. Struct. Mol. Biol.* **2010,** *17* (12), 1431-7.

124. Tynan, F.E.; Burrows, S.R.; Buckle, A.M.; Clements, C.S.; Borg, N.A.; Miles, J.J.; Beddoe, T.; Whisstock, J.C.; Wilce, M.C.; Silins, S.L.; Burrows, J.M.; Kjer-Nielsen, L.; Kostenko, L.; Purcell, A.W.; McCluskey, J.; Rossjohn, J. "*T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide*" *Nat. Immunol.* **2005,** *6* (11), 1114-22.

125. Motozono, C.; Kuse, N.; Sun, X.; Rizkallah, P.J.; Fuller, A.; Oka, S.; Cole, D.K.; Sewell, A.K.; Takiguchi, M. "*Molecular basis of a dominant T cell response to an HIV reverse transcriptase 8-mer epitope presented by the protective allele HLA-B*51:01*" *J. Immunol.* **2014,** *192* (7), 3428-34.

126.   Cole, D.K.; Pumphrey, N.J.; Boulter, J.M.; Sami, M.; Bell, J.I.; Gostick, E.; Price, D.A.; Gao, G.F.; Sewell, A.K.; Jakobsen, B.K. "*Human TCR-binding affinity is governed by MHC class restriction*" *J. Immunol.* **2007,** *178*, 5727–5734.

127.   Narzi, D.; Becker, C.M.; Fiorillo, M.T.; Uchanska-Ziegler, B.; Ziegler, A.; Bockmann, R.A. "*Dynamical characterization of two differentially disease associated MHC class I proteins in complex with viral and self-peptides*" *J. Mol. Biol.* **2012,** *415* (2), 429-42.

128.   Riedhammer, C.; Weissert, R. "*Antigen Presentation, Autoantigens, and Immune Regulation in Multiple Sclerosis and Other Autoimmune Diseases*" *Front. Immunol.* **2015,** *6*, 322.

129.   Sousa, M.A.; Paraná, R.; Andrade, L.J. "*Sequence similarity between thyroid self-protein and hepatitis C virus polyprotein: Possible triggering mechanism of autoimmune thyroiditis*" *Arq. Gastroenterol.* **2016,** *53* (3), 185-91.
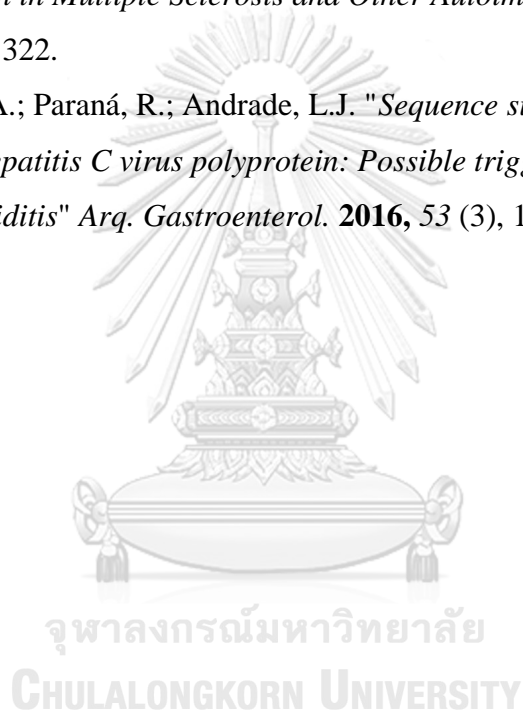
จุฬาลงกรณ์มหาวิทยาลัย
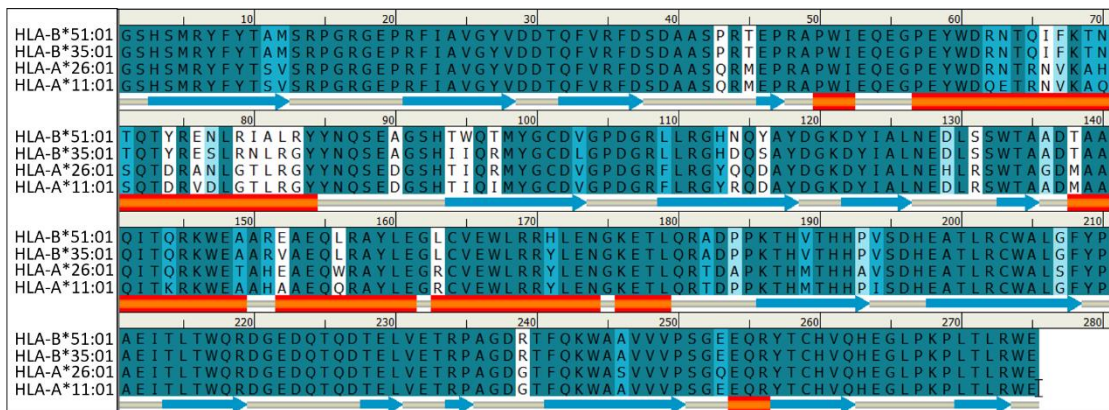CHULALONGKORN UNIVERSITY

APPENDIX I

Figure 34. Sequence alignment of HLAs in this study. The conserve residues represent by dark cyan shade, while different are show in white.

Table S1. HLA residues with a total decomposition (DC) free energy ($\Delta G_{\text{residue}}$) of less than -0.5 kcal/mol.

B*51:01 [a]		B*35:01 [b]		A*26:01 [a]		A*11:01 [b]	
residue	$\Delta G_{\text{residue}}$	residue	$\Delta G_{\text{residue}}$	residue	$\Delta G_{\text{residue}}$	residue	$\Delta G_{\text{residue}}$
α1 helix							
I66	-1.69	I66	-2.06	Y59	-0.62	E63	-3.33
T69	-0.71	T69	-0.68	N63	-2.82	N66	-0.72
N70	-1.84	N70	-0.54	H70	-2.89	T73	-1.58
T73	-2.19	T73	-1.91	T73	-1.79	D77	-0.76
Y74	-0.66	S77	-2.89	N77	-0.90		
E76	-0.55	N80	-4.53	T80	-0.78		
N77	-0.93	L81	-0.92	L81	-0.93		
I80	-1.46			Y84	-0.76		
Y84	-0.84						
α2 helix							
I142	-0.51	T143	-1.08	T143	-0.78	K146	-3.19
T143	-1.75	K146	-3.33	K146	-3.96	W147	-2.40
K146	-4.57	W147	-3.58	W147	-3.17	A150	-0.63
W147	-3.02	V152	-1.49	W156	-2.38	A152	-0.68
A150	-0.71	L156	-1.29	Y159	-1.91	Q155	-0.54
Q155	-0.52	Y159	-2.67	W167	-2.70	Y159	-2.51
L156	-1.16	W167	-2.17			W167	-1.69
Y159	-3.72						
L163	-0.71						
β-strand							
Y9	-0.64	Y7	-1.67	Y7	-1.46	Y7	-2.24
W95	-0.55	R97	-2.34	Y99	-1.87	Y9	-0.63
Y99	-2.03	Y99	-2.38			Y99	-2.89
Y123	-0.54	Y123	-0.54				

HLA alleles are [a] associated or [b] not associated with Behçet's disease (BD)

Table S2. The MM/PB(GB)SA method for energy components and binding free energy (kcal/mol) for TCR/HLAp complexes, whereas p are HIV8p, HIV9p and MICA peptides.

Energy Component	TCR/HLAp		
	HIV8p	HIV9p	MICA
ΔE_{vdW}	-93.3 ± 9.4	-89.2 ± 8.2	-106.5 ± 10.3
ΔE_{ele}	-280.2 ± 49.9	-295.5 ± 54.7	-412.2 ± 48.0
ΔE_{MM}	-373.5 ± 53.9	-384.7 ± 51.2	-518.6 ± 53.2
MM/GBSA			
ΔE_{GB}	342.5 ± 48.9	345.4 ± 45.0	470.7 ± 47.0
ΔE_{SURF}	-13.6 ± 1.2	-13.3 ± 0.8	-14.5 ± 1.4
ΔG_{solv}	328.8 ± 48.0	332.1 ± 44.9	456.2 ± 46.0
$\Delta G_{total(GB)}$	-44.7 ± 8.9	-52.7 ± 9.6	-62.5 ± 10.6
MM/PBSA			
ΔE_{PB}	320.4 ± 49.7	326.9 ± 42.6	452.6 ± 50.5
$\Delta E_{non\text{-}polar}$	-16.4 ± 1.2	-15.4 ± 0.8	-15.8 ± 1.2
ΔG_{solv}	304.0 ± 49.1	311.4 ± 42.6	436.8 ± 49.6
$\Delta G_{total(PB)}$	-69.5 ± 11.4	-73.3 ± 12.0	-81.8 ± 13.2
TΔS	-59.2 ± 17.0	-45.7 ± 16.9	-43.5 ± 16.2
$\Delta G_{binding(MM/GBSA)}$	14.6 ± 19.2	-7.0 ± 19.5	-18.9 ± 19.3
$\Delta G_{binding(MM/PBSA)}$	-10.3 ± 20.5	-27.6 ± 20.8	-38.3 ± 20.9
$\Delta G_{binding(Exp.)}$	-5.6		

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

APPENDIX II

Conferences (Oral and Poster presentations)

1. The 18th Annual Symposium on Computational Science and Engineering (ANSCSE18), Pattaya, Chonburi, Thailand, March 17-19, 2014.

   Topic: Inhibition of HPV-16 E2 protein by gallotannin and its derivatives

2. ICBCBBE 2014 : 16th International Conference on Bioinformatics, Computational Biology and Biomedical Engineering, Kyoto, Japan, November 13-14, 2014

   Topic: MICA-TM peptide selectively binds to HLAs associated with Behçet's disease

3. The 11th Thai Summer School of Computational Chemistry (11th-TS$_2$C$_2$), Nan, Thailand, January 4-7, 2015.

   Topic: The Selective Binding of MICA-TM Peptide to HLAs Associated with Behçet's Disease

4. The 20th Annual Symposium on Computational Science and Engineering (ANSCSE20), Kasetsart University, Bangkok, Thailand, July 27-29, 2016.

   Topic: The Dynamics and Binding Interactions of HLA-DR and Topoisomerase-I Peptide Risk in Systemic Sclerosis Diseases: Molecular Computations

5. The 21st Annual Symposium on Computational Science and Engineering (ANSCSE21), Thailand Science Park, Pathum Thani, Thailand, August 3-4, 2017.

   Topic: How a T-cell Antigen Receptor Recognizes Self-peptide Leading Autoimmune

6. The 43rd Congress on Science and Technology of Thailand (STT 43), Chulalongkorn University, Bangkok, Thailand, October 17-19, 2017.

   Topic: Autoimmunity: How does T-Cell Recognize Self-Antigen?

7. The 1st Taiwan-Thailand-Vietnam Workshop on Theoretical and Computational Chemistry, The National Taiwan University of Science and Technology, Taipei, Taiwan, March 22-23, 2018.

   Topic: Interactions of HLA-DR and Topoisomerase I Epitope Modulated Genetic Risk for Systemic Sclerosis

Publications

1. Kongkaew, S.; Yotmanee, P.; Rungrotmongkol, T.; Kaiyawet, N.; Meeprasert, A.; Kaburaki, T.; Noguchi, H.; Takeuchi, F.; Kungwan, N.; Hannongbua, S. "*Molecular Dynamics Simulation Reveals the Selective Binding of Human Leukocyte Antigen Alleles Associated with Behçet's Disease*" *PLoS One* **2015,** *10* (9), e0135575.

2. Kongkaew, S.; Rungrotmongkol, T.; Punwong, C.; Noguchi, H.; Takeuchi, F.; Kungwan, N.; Wolschann, P.; Hannongbua, S. *"Interactions of HLA-DR and Topoisomerase I Epitope Modulated Genetic Risk for Systemic Sclerosis"* *Sci. Rep.* **2018**. (in revision)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

Personal data:

Name   Sirilak Kongkaew

Nationality   Thai

Date of birth   22 Febuary, 1987

Place of birth   Saraburi, Thailand

E-mail address   Sirilak.k6548@gmail.com; silkyo_j@hotmail.com

Educations:

2005-2009   B.Sc. (Chemistry), Silpakorn University, Thailand

2010-2014   M.Sc. (Inorganic chemistry), Chulalongkorn University, Thailand

2015-2018   Ph.D. (Biotechnology), Chulalongkorn University, Thailand

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY