의학석사 학위논문

Role of HLA-DP polymorphism in the pathogenesis of systemic sclerosis

전신성 경화증의 발병기전에서 인간백혈구항원-DP 다형성의 역할에 대한 연구

2015 년 2 월

서울대학교 대학원

의학과 면역학전공

이 정 석

A thesis of the Degree of Master of Science

전신성 경화증의 발병기전에서

인간백혈구항원-DP 다형성의 역할에 대한 연구

Role of HLA-DP polymorphism

in the pathogenesis of systemic sclerosis

February 2015

The Department of Medicine

Seoul National University

College of Medicine

Jeong Seok Lee

Role of HLA-DP polymorphism

in the pathogenesis of systemic sclerosis

by

Jeong Seok Lee

(Directed by Eun Bong Lee M.D,, Ph.D)

A thesis submitted to the

Faculty of the Graduate School of the

Seoul National University in partial fulfillment

Of the requirements for the degree of

Master of Science in Medicine (Immunology)

February 2015

Approved by Thesis Committee:

Professor Yeong Wook Song **Chairman**

Professor Eun Bong Lee **Vice chairman**

Professor Kyeong Cheon Jung

전신성 경화증의 발병기전에서

인간백혈구항원-DP 다형성의 역할에 대한 연구

지도교수 이 은 봉

이 논문을 의학석사 학위논문으로 제출함

2014년 10월

서울대학교 대학원

의학과 면역학전공

이 정 석

이정석의 의학석사 학위논문을 인준함

2015년 1월

위원장 송 영 욱 (인)

부위원장 이 은 봉 (인)

위원 정 경 천 (인)

ABSTRACT

Objectives: To identify common characteristics of amino acid sequences in the risk HLA-DPB1 alleles in systemic sclerosis (SSc) and to predict pathogenic peptide sequence of topoisomerase I.

Methods: A total of 127 Korean patients diagnosed with SSc and 548 healthy Korean controls were enrolled. Amino acid sequences of the risk HLA-DPB1 alleles were analyzed, with a focus on known HLA-DP binding motifs. The affinity of all possible 15 amino acid peptide fragments from topoisomerase I against HLA-DP risk alleles compared to protective alleles was examined by NetMHCIIpan-3.0.

Results: HLA-DPB1*0901, *1301, and *030101 were revealed as risk alleles for anti-topoisomerase I antibody (ATA)-positive SSc. In contrast, HLA-DPB1*020102 and *0501 were protective against ATA-positive SSc. HLA-DP6 had higher risk than any other supertype of HLA-DP. At amino acid positions 55-57, 67-69, and 82, 84-85 of HLA-DPB1, the number of negatively-charged triplets (NCTs) of amino acids (D, aspartic acid, and E, glutamic acid) were proportional to the odds ratio of ATA-positive SSc. The positions of the NCTs were also important and beta-82,84, and 85 at P1 of HLA-DP binding pocket has a key role. The pathogenic peptides with high affinity to HLA-DP risk alleles had partial concordance with those towards the risk allele of HLA-DQ but no concordance with known risk alleles of HLA-DR.

Conclusion: ATA-positive SSc-susceptible HLA-DPB1 alleles share NCTs at critical positions in the peptide binding groove. The peptide fragments of topoisomerase I with high affinity to HLA-DPB1 risk alleles were also favored by risk alleles of HLA-DQB1, suggesting that they may play a role in the pathogenesis of SSc.

Key word: Systemic sclerosis, HLA, Topoisomerase-I, Epitope

Student number: 2011-21896

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure 4. Predicted affinities of all possible 15-mer peptides of topoisomerase I (1-765) for high-risk alleles on other class II HLA (HLA-DRB1*1104 and DQB1*0301). (a) The ratio of the 1/IC50 of the risk HLA-DRB1*1104 allele to the protective HLA-DRB1 allele (*0401) is shown. (b) The ratio of the 1/IC50 of the risk HLA-DQB1*0301 allele to the protective HLA-DQB1 allele (*0202) is shown. All IC50s were calculated using the NetMHCIIpan-3.0 algorithm.

LIST OF ABBREVIATIONS

SSc	**Systemic sclerosis**
HLA	**Human leukocyte antigen**
ATA	**Anti-topoisomerase I antibody**
GWAS	**Genome wide association study**
OR	**Odds ratio**
NCT	**Negatively-charged triplet**

Introduction

Systemic sclerosis (SSc) is a chronic autoimmune disease of unknown etiology that is characterized by fibrosis of the skin and internal organs (1). It is associated with the presence of autoantibodies, including anti-topoisomerase I antibody (ATA) and anti-centromere antibody (1). Genetic factors appear to contribute to SSc pathogenesis (2). While several genetic polymorphisms have been identified in patients with SSc, polymorphisms in human leukocyte antigen (HLA) have been shown to be more closely associated with the presence of specific autoantibodies than with disease phenotype (2). Specifically, HLA-DQB1 and DRB1 loci are associated with ATA-positive SSc (3). The common structural features of the high risk alleles include changes to the tyrosine at position 26 and the five amino acid motif (FLEDR) at position 67-71 in the ß1 domain of DRB1 (3). The first genome wide association study (GWAS) of SSc found that HLA-DPB1 was the most strongly associated with ATA-positive SSc in Korean and Caucasian patients, with HLA-DPB1*0901, *1301, and *030101 as the risk alleles of ATA-positive SSc in Korean patients (4). Previous studies in other ethnic groups have also confirmed the strong association of HLA-DPB1 (5, 6).

HLA-DP polymorphisms have also been found to be associated with other autoimmune diseases, including juvenile rheumatoid arthritis, Graves' disease, Takayasu's arteritis, and chronic beryllium disease (7-10). Previous genetic studies have suggested an association between HLA-DP and systemic sclerosis in ATA-positive SSc (11, 12). Gilchrist *et al.* suggested that the glutamate residue at position 69 of the beta chain may be the unique feature of the risk allele HLA-DPB1*1301 (12). However, it remains unknown whether any common structural features of the risk alleles of HLA-DPB1 occur at peptide binding sites in ATA-positive SSc.

The interaction between antigenic peptide and HLA class II was mainly investigated in HLA-DR or –DQ because the crystal structure of HLA-DP was not elucidated until 2010 (13, 14). While computational predictions for peptides binding to HLA molecules are a powerful tool for identifying epitope candidates, the NetMHCIIpan-3.0 algorithm is the first method capable of predicting peptide binding to any HLA class II molecule with a defined protein sequence (15). Since human topoisomerase I has positively charged domains that allow it to bind to a negatively-charged DNA strand (16), we hypothesized that the risk alleles may result in negatively-charged amino acids in the epitopes of HLA-DP molecules, which would then bind and present positively-charged topoisomerase I. In this study, we analyzed the amino acid sequences of the peptide-binding sites in HLA-DPB1 alleles of ATA-positive SSc patients in order to determine whether the susceptible alleles shared any structural changes. In addition, we identified topoisomerase I-derived peptides that bind to HLA-DP molecules derived from susceptible alleles using the NetMHCIIpan-3.0 algorithm.

Patients and methods

Study subjects and HLA-DP typing

All of the study subjects were Korean and their allelic types and ATA status were previously reported (4). Briefly, the study population was composed of 127 patients with SSc and a control population of 548 subjects. Of the 127 cases, 79 were positive for ATA. For all of the patients, HLA-typing was genotyped by sequencing exon 2 of the HLA-DPB1 gene. This study was approved by the institutional review board of Seoul National University Hospital, and all patients and controls were provided with written informed consent.

Identification of common epitopes of HLA-DP molecules

Since the peptide binding sites of HLA-DPB1 are located at amino acid positions 55-57, 67-69, and 82-85, we searched for common amino acid structures at these strategic positions (17-19). Grouping by supertype of HLA-DP was also performed and the risk of each supertype was examined (20).

Molecular dynamic simulation

Amino acid sequences of HLA-DPB1*0901, *1301, *030101, and *0401 were obtained from the IMGT/HLA Database (www.ebi.ac.uk/ipd/imgt/hla/allele.html). The structure of each HLA-DPB1 molecule was derived from the crystal structure of HLA-DP2 (PDB ID: 3LQZ) (13). The amino acid sequence KDREHRHKEHKK was chosen as a binding peptide from topoisomerase I, representing the position of the most dominant positive charge in the molecule. A three-dimensional structure of the binding mode between the HLA-DP molecule and the selected peptide was simulated using Discovery Studio (Accelrys, San Diego, CA).

An initial systematic match of the protein-peptide interaction was performed using the ZDOCK protocol and further refinement of the docked-interaction was done by the RDOCK protocol (21, 22). The docking prediction of the HLA-DP molecule and peptides with the lowest binding free energy ΔΔG was selected.

HLA binding prediction against topoisomerase I

Human topoisomerase I consists of 765 amino acids and, in theory, up to 751 fragments of 15-mer peptides can exist. The affinity of each of the possible 751 15-mer peptides from topoisomerase I against HLA-DP risk alleles (HLA-DPB1*0901, *1301, and *030101) and protective alleles (HLA-DPB1*020102 and *0501) was examined by the NetMHCIIpan-3.0 algorithm (http://www.cbs.dtu.dk/services/NetMHCIIpan-3.0) (15).

The half maximal inhibitory concentration (IC50, nM) was obtained for each HLA-DPB1 allele against the 15-mer peptides and 1/IC50 (nM^{-1}) of each peptide was compared as a parameter of affinity. The ratio of the 1/IC50 for one risk allele to one protective allele was calculated to determine the selectiveness of the peptide fragment to the risk allele. A higher ratio indicates that the peptide favors the risk allele over the protective allele. The same comparison was conducted in HLA-DR (risk allele: HLA-DRB1*1104 vs. protective allele: *0401) and DQ (risk allele: HLA-DQB1*0301 vs. protective allele: *0202).

Statistical analysis

For each allele, the association between the number of negatively-charged triplets of amino acids and whether the patient had ATA-positive or ATA-negative SSc was analyzed using chi-square tests. The odds ratio (OR) of each amino acid sequence was determined by comparison with the control group and its relevant 95% confidence interval (95% CI) was

also determined. Bonferroni's correction was applied to adjust the multiple comparison and $p<0.05/n$ was used as cut-off value of statistical significance (n: the number of comparison). SAS, version 9.1.3 (SAS Institute, Cary, NC), were used for statistical analysis.

Results

Both risk and protective alleles of HLA-DPB1 occur in systemic sclerosis patients with positivity of anti-topoisomerase I antibody.

Of the 127 patients with systemic sclerosis, 79 were ATA positive and 48 patients were ATA negative. After Bonferroni's correction for multiple comparisons, the alleles that were significantly associated with risk of ATA-positive SSc included HLA-DPB1*0901 (OR=4.73, $p<10^{-4}$), *1301 (OR=4.56, $p<10^{-4}$), and *030101 (OR=2.69, $p=5.5x10^{-4}$) (Table 1). In contrast, the alleles HLA-DPB1*020102 (OR=0.38, $p<10^{-4}$) and *0501 (OR=0.48, $p=3.3x10^{-4}$) were protective against ATA-positive SSc (Table 1).

HLA-DP6 is the risk supertype and HLA-DP2 is the protective supertype of systemic sclerosis.

Based on an analysis of the HLA-DPB1 alleles, the supertype HLA-DP6, involving HLA-DPB1*0901, *1301, *1701, *1901, *2201, and *2901, is significantly associated with risk for systemic sclerosis (OR 5.16, $p<10^{-4}$). HLA-DP6 was previously defined as a supertype of the amino acid sequence signature in the beta chain of HLA-DP of the glutamic acid at position 69 and the aspartic acid at position 84 (20) (Table 2). In addition, HLA-DP2, which was previously defined as involving the glutamic acid at position 69 and the glycine or valine at position 84 of the beta chain of HLA-DP, had a significantly protective effect (OR 0.37, $p<10^{-4}$) (20) (Table 2).

Frequency and position of negatively-charged triplets of amino acids in three polymorphic sites of HLA-DP is associated with the risk of systemic sclerosis.

Negatively-charged amino acids, such as aspartic acid (D) or glutamic acid (E), were predominantly located in triplets (negatively-charged triplet, NCT) at positions 55/56/57

(DED or DEE), 67/68/69 (EEE), and 82/84/85 (EDE) of HLA-DPB1. The number of NCTs at the peptide binding groove of HLA-DP was significantly increased in ATA-positive SSc patients compared with healthy controls. In addition, the number of NCTs was proportional to the odds ratio of ATA-positive SSc. Patients with five or six NCTs showed an extremely high risk of SSc (OR=6.73, $p<10^{-4}$; OR=29.17, p=0.0027) (Table 3).

In addition to the frequency of NCTs, the position was also important. The NCT at position 82-85 in combination with one or more NCTs at the other positions was associated with increased risk of ATA positive SSc. The ORs for 82-85 with 55-57 was 2.41(p=0.0015) and 82-85 with 67-69 was 4.37 ($p<10^{-5}$). The OR of all three NCTs was 3.93 ($p<10^{-5}$), while the OR for one NCT or two NCTs excluding position 82-85 was similar to no NCT groups. Position 82-85 seems to play a significant role in binding the antigenic peptide in ATA-positive SSc patients (Table 4).

Interaction between a peptide of topoisomerase I and HLA-DP protein.

A unique peptide fragment of topoisomerase I (Peptide A, positions 29-40) was chosen as a binding peptide due to its ample positive charges (KDREHRHKEHKK, Figure 1a). The positively-charged amino acids lysine (K) and arginine (R) occur at all contact points (P1, P4, P6, P9) with the HLA-DP binding epitopes (position 55-57, 67-69, and 82-85) (13, 17-19) .

In molecular dynamic simulations, peptide A fit more closely into the binding cleft of the HLA-DP proteins that have more NCT amino acids than those without NCTs (Figure 1b), suggesting sustained peptide presentation by antigen presenting cells.

Binding prediction with HLA-DP molecules suggests potentially pathogenic peptide fragments of topoisomerase I.

NetMHCIIpan-3.0 was used to predict the affinity of all possible 15-mer peptides of topoisomerase I and the 1/IC50 represents the degree of affinity for the risk allele compared to the protective allele. The 1/IC50 of each peptide against the high-risk HLA-DPB1 alleles (*0901, *1301, *0301) was compared to that of the protective alleles (*0501 in Figure 2 and *0201 in Figure 3). Based on relative affinities (1/IC50 ratio), the upper 10% (75 of 751 fragments) of the 15-mer peptides exclusively favoring the risk alleles were compared. A total of 30 (against *0501) and 29 (against *0201) commonly appearing peptides were considered to be pathogenic peptides (Table 5), 10 of which (585-599 to 590-604, 694-708 to 697-711) overlapped between the two groups. By expanding the range of analysis to the upper 20% (150 of 751 fragments) of 1/IC50 ratio, 87 (against *0501) and 70 (against *0201) fragments were identified as potential peptides that selectively bind during disease-specific immune activation at the HLA-DP surface.

The peptide fragments favored by risk alleles of HLA-DP had partial overlap with those of HLA-DQ but no concordance with those of HLA-DR.

Using the same method, NetMHCIIpan-3.0 was used to predict the affinity of all possible 15-mer peptides of topoisomerase I against the known risk and protective alleles of HLA-DRB1 and DQB1 (4) (Figure 4). Among the upper 10% of peptides by 1/IC50 favoring the risk alleles, there were six common peptides (585-599 to 590-604) between HLA-DPB1 alleles and the risk allele of HLA-DQB1, but no common peptide fragments between HLA-DPB1 alleles and HLA-DRB1 (Table 6).

Table 1. The frequency of HLA-DPB1 alleles in SSc patients with ATA positive (n=79) and negative status (n=48) and healthy controls (n=548).

HLA-DPβ1 alleles	SSc patients with ATA(+) (n=79)			SSc patients with ATA(-) (n=48)			Control (n=548)
	Frequency	OR (95% CI)	*p-value	Frequency	OR (95% CI)	*p-value	Frequency
DPB1*0102	0	-	-	0.010	2.87 (0.32 – 26.0)	0.35	0.0036
DPB1*020102	0.14	**0.38** (0.24 – 0.61)	**<1.0x10^{-4}**	0.28	0.92 (0.58 – 1.47)	0.74	0.30
DPB1*0202	0.025	0.58 (0.21 – 1.63)	0.30	0.042	0.97 (0.34 – 2.75)	0.96	0.043
DPB1*030101	0.11	**2.69** (1.50 – 4.82)	**0.00055**	0.021	0.47 (0.11 – 1.99)	0.31	0.043
DPB1*040101	0.025	0.44 (0.16 – 1.23)	0.11	0.073	1.33 (0.59 – 3.00)	0.49	0.056
DPB1*0402	0.057	0.80 (0.39 – 1.63)	0.54	0.094	1.37 (0.66 – 2.82)	0.40	0.070
DPB1*0501	0.20	**0.48** (0.32 – 0.72)	**0.00033**	0.375	1.14 (0.74 – 1.75)	0.56	0.35
DPB1*0502	0.0063	-	-	0	-	-	0
DPB1*0602	0.0063	-	-	0	-	-	0
DPB1*0901	0.11	**4.73** (2.56 – 8.74)	**<1.0x10^{-4}**	0.010	0.39 (0.05 – 2.87)	0.35	0.026
DPB1*1102	0	-	-	0	-	-	0.00091
DPB1*1301	0.21	**4.56** (2.87 – 7.25)	**<1.0x10^{-4}**	0.052	0.95 (0.37 – 2.42)	0.91	0.055
DPB1*1401	0.019	1.16 (0.34 – 3.98)	0.81	0.042	2.60 (0.86 – 7.86)	0.089	0.016
DPB1*1501	0.0063	-	-	0	-	-	0
DPB1*1701	0.051	2.28 (1.01 – 5.16)	0.047	0	-	-	0.023
DPB1*1901	0.013	3.50 (0.64 – 19.3)	0.15	0	-	-	0.0036
DPB1*1902	0	-	-	0	-	-	0.0018
DPB1*2201	0	-	-	0	-	-	0.00091
DPB1*2901	0.0063	2.32 (0.24 – 22.4)	0.47	0	-	-	0.0027
DPB1*3801	0	-	-	0	-	-	0.0027
DPB1*410101	0	-	-	0	-	-	0.0018
DPB1*4501	0	-	-	0	-	-	0.0018
DPB1*4601	0	-	-	0	-	-	0.00091
DPB1*4801	0	-	-	0	-	-	0.00091
DPB1*5701	0.0063	-	-	0	-	-	0
DPB1*8301	0	-	-	0	-	-	0.00091
DPB1*9801	0.0063	3.48 (0.31 – 38.6)	0.31	0	-	-	0.0018

SSc: systemic sclerosis, ATA: anti-topoisomerase I antibody. OR: odds ratio. CI: confidence interval. * Following Bonferroni's correction for multiple comparisons, $p \leq 0.05/27=0.0019$ is considered significant.

Table 2. Grouping of HLA-DPB1 alleles based on the supertype of HLA-DP to identify risk and protective supertypes for SSc.

HLA-DP supertype	HLA-DPB1 alleles	SSc patients with ATA(+) (n=79)			SSc patients with ATA(-) (n=48)			Control (n=548)
		Frequency	OR (95% CI)	*p-value	Frequency	OR (95% CI)	*p-value	Frequency
DP1	DPB1*030101 DPB1*0501 DPB1*0502 DPB1*1401 DPB1*3801 DPB1*4501 DPB1*5701 DPB1*9801	0.35	0.75 (0.53–1.07)	0.11	0.44	1.14 (0.75–1.74)	0.53	0.41
DP2	DPB1*020102 DPB1*0202 DPB1*1102 DPB1*410101 DPB1*4601 DPB1*4801	**0.16**	**0.37** (0.24–0.58)	**<0.0001**	0.33	0.91 (0.58–1.42)	0.68	0.34
DP4	DPB1*0102 DPB1*040101 DPB1*0402 DPB1*0602 DPB1*1501 DPB1*1902 DPB1*8301	0.095	0.67 (0.38–1.19)	0.17	0.18	1.38 (0.78–2.42)	0.27	0.13
DP6	DPB1*0901 DPB1*1301 DPB1*1701 DPB1*1901 DPB1*2201 DPB1*2901	0.39	**5.16** (3.56–7.44)	**<0.0001**	0.063	0.52 (0.22–1.21)	0.13	0.11

SSc: systemic sclerosis, ATA: anti-topoisomerase I antibody. OR: odds ratio. CI: confidence interval. D: aspartic acid. E: glutamic acid. K: lysine, G: glycine, V: valine. * Following Bonferroni's correction for multiple comparisons, $p \leq 0.05/4 = 0.013$ is considered significant.

Table 3. The association between number of negatively-charged triplets (NCTs) at three polymorphic sites of the beta chain of HLA-DPB1 and risk of ATA-positive SSc. The three sites included DED(E) at positions 55/56/57, EEE at positions 67/68/69, and DED at positions 82/84/85.

Number of NCTs in each patient	SSc patients with ATA(+) (n=79)			SSc patients with ATA(-) (n=48)			Control (n=548)
	Frequency	OR (95% CI)	*p-value	Frequency	OR (95% CI)	*p-value	Frequency
6	0.05	**29.17** (3.22–265)	**0.0027**	0	-	-	0.0018
5	0.18	**6.73** (3.17–14.3)	**<0.0001**	0	-	-	0.031
4	0.32	1.30 (0.78–2.16)	0.32	0.19	0.65 (0.31–1.37)	0.26	0.26
3	0.27	0.61 (0.36–1.04)	0.067	0.42	1.20 (0.66–2.13)	0.54	0.37
2	0.19	0.58 (0.32–1.06)	0.075	0.29	1.03 (0.54–1.96)	0.94	0.29
1	0	-	-	0.10	2.65 (0.96–7.33)	0.060	0.042
0	0	-	-	0	-	-	0.0036

SSc: systemic sclerosis, ATA: anti-topoisomerase I antibody. OR: odds ratio. CI: confidence interval. D: aspartic acid. D: Aspartic acid. E: Glutamic acid. * Following Bonferroni's correction for multiple comparisons, $p \leq 0.05/7 = 0.0071$ is considered significant.

Table 4. Position of negatively-charged triplets (NCTs) in three polymorphic site of HLA-DPB1 (55/56/57, 67/68/69, 82/84/85) was associated with the risk of SSc. The alleles with two or more NCTs including NCT at 82/84/85 have significantly increased OR with ATA positive SSc compared to healthy control.

Polymorphic sequence of HLA-DPβ1 chain			SSc patients with ATA(+) (n=79)			SSc patients with ATA(-) (n=48)			Control (n=548)
55/56/57	67/68/69	82/84/85	Frequency	OR (95% CI)	*p-value	Frequency	OR (95% CI)	*p-value	Frequency
-	-	-	0.032	0.55 (0.22-1.40)	0.25	0.073	1.33 (0.59-3.00)	0.49	0.056
DED(E)	-	-	0.057	0.79 (0.39-1.61)	0.62	0.094	1.35 (0.65-2.78)	0.41	0.071
-	EEE	-	0.025	0.58 (0.21-1.63)	0.39	0.042	0.97 (0.34-2.75)	1.00	0.043
-	-	EDE	0.21	**0.48** (0.32-0.72)	**2.9×10^{-4}**	0.39	1.15 (0.75-1.76)	0.58	0.35
DED(E)	EEE	-	0.14	**0.38** (0.24-0.60)	**1.1×10^{-5}**	0.28	0.91 (0.57-1.44)	0.73	0.30
DED(E)	-	EDE	0.14	**2.41** (1.44-4.02)	**0.0015**	0.063	0.99 (0.42-2.35)	1.00	0.063
-	EEE	EDE	0.22	**4.37** (2.79-6.85)	**$<1.0 \times 10^{-5}$**	0.052	0.84 (0.33-2.15)	1.00	0.061
DED(E)	EEE	EDE	0.18	**3.93** (2.41-6.39)	**$<1.0 \times 10^{-5}$**	0.010	0.19 (0.026-1.40)	0.081	0.052

SSc: systemic sclerosis, ATA: anti-topoisomerase I antibody. OR: odds ratio. CI: confidence interval. D: aspartic acid. D: Aspartic acid. E: Glutamic acid. * Following Bonferroni's correction for multiple comparisons, $p \leq 0.05/8 = 0.0063$ is considered significant.

Table 5. List of the 15-mer peptide fragments from topoisomerase I favoring risk alleles of HLA-DPB1. Based on relative affinities (1/IC50 ratio), the upper 10% (75 of 751 fragments) of the 15-mer peptides exclusively favoring the risk alleles were compared. A total of 30 (against *0501) and 29 (against *0201) commonly appearing peptides were considered to be pathogenic peptides.

Risk alleles versus DPB1*0501 (position start / end / sequence)			Risk alleles versus DPB1*0201 (position start / end / sequence)		
223	237	KGPVFAPPYEPLPEN	419	433	NIQGSIKYIMLNPSS
256	270	VATFFAKMLDHEYTT	421	435	QGSIKYIMLNPSSRI
257	271	ATFFAKMLDHEYTTK	422	436	GSIKYIMLNPSSRIK
			424	438	IKYIMLNPSSRIKGE
404	418	VRHDNKVTWLVSWTE	425	439	KYIMLNPSSRIKGEK
405	419	RHDNKVTWLVSWTEN			
406	420	HDNKVTWLVSWTENI	562	576	DDLFDRLNTGILNKH
407	421	DNKVTWLVSWTENIQ	563	577	DLFDRLNTGILNKHL
408	422	NKVTWLVSWTENIQG	564	578	LFDRLNTGILNKHLQ
409	423	KVTWLVSWTENIQGS			
410	424	VTWLVSWTENIQGSI	585	599	TAKVFRTYNASITLQ *
411	425	TWLVSWTENIQGSIK	586	600	AKVFRTYNASITLQQ *
412	426	WLVSWTENIQGSIKY	587	601	KVFRTYNASITLQQQ *
			588	602	VFRTYNASITLQQQL *
575	589	KHLQDLMEGLTAKVF	589	603	FRTYNASITLQQQLK *
			590	604	RTYNASITLQQQLKE *
585	599	TAKVFRTYNASITLQ *			
586	600	AKVFRTYNASITLQQ *	598	612	LQQQLKELTAPDENI

587	601	KVFRTYNASITLQQQ *			
588	602	VFRTYNASITLQQQL *	617	631	LSYNRANRAVAILCN
589	603	FRTYNASITLQQQLK *	618	632	SYNRANRAVAILCNH
590	604	RTYNASITLQQQLKE *	619	633	YNRANRAVAILCNHQ
591	605	TYNASITLQQQLKEL	620	634	NRANRAVAILCNHQR
694	708	LEEQLMKLEVQATDR *	639	653	TFEKSMMNLQTKIDA
695	709	EEQLMKLEVQATDRE *	640	654	FEKSMMNLQTKIDAK
696	710	EQLMKLEVQATDREE *	641	655	EKSMMNLQTKIDAKK
697	711	QLMKLEVQATDREEN *	642	656	KSMMNLQTKIDAKKE
			643	657	SMMNLQTKIDAKKEQ
746	760	KTQREKFAWAIDMAD			
747	761	TQREKFAWAIDMADE	694	708	LEEQLMKLEVQATDR *
748	762	QREKFAWAIDMADED	695	709	EEQLMKLEVQATDRE *
749	763	REKFAWAIDMADEDY	696	710	EQLMKLEVQATDREE *
750	764	EKFAWAIDMADEDYE	697	711	QLMKLEVQATDREEN *
751	765	KFAWAIDMADEDYEF	698	712	LMKLEVQATDREENK

position start/end: position of the first and last amino acids of 15-mer peptide on the sequence of topoisomerase I sequence.

*indicates peptides that overlapped between the two groups

Table 6. The upper 10% by 1/IC50 ratio of the 15-mer peptide fragments from topoisomerase I that favor the risk alleles of HLA-DRB1 (risk allele: HLA-DRB1*1104 vs. protective allele: *0401) and DQB1 (risk allele: HLA-DQB1*0301 vs. protective allele: *0202).

High risk DRB1*1104 specific peptides (position start / end / sequence)			High risk DQB1*0301 specific peptides (position start / end / sequence)		
127	141	VPPKEDIKPLKRPRD	88	102	KRKEEKVRASGDAKI
128	142	PPKEDIKPLKRPRDE	89	103	RKEEKVRASGDAKIK
129	143	PKEDIKPLKRPRDED	90	104	KEEKVRASGDAKIKK
130	144	KEDIKPLKRPRDEDD	91	105	EEKVRASGDAKIKKE
131	145	EDIKPLKRPRDEDDV	92	106	EKVRASGDAKIKKEK
132	146	DIKPLKRPRDEDDVD	93	107	KVRASGDAKIKKEKE
			94	108	VRASGDAKIKKEKEN
242	256	YDGKVMKLSPKAEEV			
243	257	DGKVMKLSPKAEEVA	103	117	KKEKENGFSSPPQIK
244	258	GKVMKLSPKAEEVAT	104	118	KEKENGFSSPPQIKD
245	259	KVMKLSPKAEEVATF	105	119	EKENGFSSPPQIKDE
			106	120	KENGFSSPPQIKDEP
268	282	YTTKEIFRKNFFKDW			
269	283	TTKEIFRKNFFKDWR	384	398	INCSKDAKVPSPPPG
270	284	TKEIFRKNFFKDWRK			
271	285	KEIFRKNFFKDWRKE	414	428	VSWTENIQGSIKYIM
272	286	EIFRKNFFKDWRKEM	415	429	SWTENIQGSIKYIML
273	287	IFRKNFFKDWRKEMT			
274	288	FRKNFFKDWRKEMTN	467	481	SKEMKVRQRAVALYF
275	289	RKNFFKDWRKEMTNE			
			493	507	KEEGETADTVGCCSL
363	377	GRGNHPKMGMLKRRI	494	508	EEGETADTVGCCSLR
364	378	RGNHPKMGMLKRRIM	495	509	EGETADTVGCCSLRV

365	379	GNHPKMGMLKRRIMP	496	510	GETADTVGCCSLRVE
366	380	NHPKMGMLKRRIMPE	497	511	ETADTVGCCSLRVEH
367	381	HPKMGMLKRRIMPED			
368	382	PKMGMLKRRIMPEDI	565	579	FDRLNTGILNKHLQD
369	383	KMGMLKRRIMPEDII			
370	384	MGMLKRRIMPEDIII	575	589	KHLQDLMEGLTAKVF*
			576	590	HLQDLMEGLTAKVFR
439	453	KDWQKYETARRLKKC	577	591	LQDLMEGLTAKVFRT
440	454	DWQKYETARRLKKCV	578	592	QDLMEGLTAKVFRTY
441	455	WQKYETARRLKKCVD	579	593	DLMEGLTAKVFRTYN
442	456	QKYETARRLKKCVDK	580	594	LMEGLTAKVFRTYNA
443	457	KYETARRLKKCVDKI	581	595	MEGLTAKVFRTYNAS
444	458	YETARRLKKCVDKIR	583	597	GLTAKVFRTYNASIT
445	459	ETARRLKKCVDKIRN	**585**	**599**	**TAKVFRTYNASITLQ***
446	460	TARRLKKCVDKIRNQ	**586**	**600**	**AKVFRTYNASITLQQ***
447	461	ARRLKKCVDKIRNQY	**587**	**601**	**KVFRTYNASITLQQQ***
448	462	RRLKKCVDKIRNQYR	**588**	**602**	**VFRTYNASITLQQQL***
449	463	RLKKCVDKIRNQYRE	**589**	**603**	**FRTYNASITLQQQLK***
450	464	LKKCVDKIRNQYRED	**590**	**604**	**RTYNASITLQQQLKE***
451	465	KKCVDKIRNQYREDW			
			605	619	LTAPDENIPAKILSY
463	477	EDWKSKEMKVRQRAV	606	620	TAPDENIPAKILSYN
464	478	DWKSKEMKVRQRAVA	607	621	APDENIPAKILSYNR
465	479	WKSKEMKVRQRAVAL	608	622	PDENIPAKILSYNRA
466	480	KSKEMKVRQRAVALY	609	623	DENIPAKILSYNRAN
467	481	SKEMKVRQRAVALYF	611	625	NIPAKILSYNRANRA
468	482	KEMKVRQRAVALYFI	612	626	IPAKILSYNRANRAV
469	483	EMKVRQRAVALYFID	613	627	PAKILSYNRANRAVA
470	484	MKVRQRAVALYFIDK	614	628	AKILSYNRANRAVAI

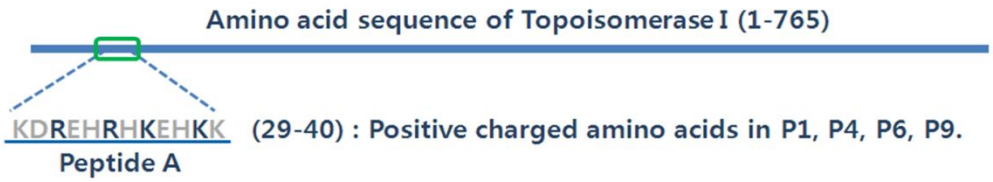
			615	629	KILSYNRANRAVAIL
538	552	YNKVPVEKRVFKNLQ	616	630	ILSYNRANRAVAILC
539	553	NKVPVEKRVFKNLQL	617	631	LSYNRANRAVAILCN*
540	554	KVPVEKRVFKNLQLF	618	632	SYNRANRAVAILCNH*
541	555	VPVEKRVFKNLQLFM	619	633	YNRANRAVAILCNHQ*
			620	634	NRANRAVAILCNHQR*
593	607	NASITLQQQLKELTA	621	635	RANRAVAILCNHQRA
594	608	ASITLQQQLKELTAP	622	636	ANRAVAILCNHQRAP
595	609	SITLQQQLKELTAPD	623	637	NRAVAILCNHQRAPP
610	624	ENIPAKILSYNRANR	659	673	ADARRDLKSAKADAK
			660	674	DARRDLKSAKADAKV
670	684	ADAKVMKDAKTKKVV	661	675	ARRDLKSAKADAKVM
671	685	DAKVMKDAKTKKVVE	662	676	RRDLKSAKADAKVMK
672	686	AKVMKDAKTKKVVES	663	677	RDLKSAKADAKVMKD
673	687	KVMKDAKTKKVVESK	664	678	DLKSAKADAKVMKDA
			665	679	LKSAKADAKVMKDAK
676	690	KDAKTKKVVESKKKA	666	680	KSAKADAKVMKDAKT
677	691	DAKTKKVVESKKKAV	667	681	SAKADAKVMKDAKTK
678	692	AKTKKVVESKKKAVQ	668	682	AKADAKVMKDAKTKK
679	693	KTKKVVESKKKAVQR	669	683	KADAKVMKDAKTKKV
680	694	TKKVVESKKKAVQRL			
681	695	KKVVESKKKAVQRLE	705	719	ATDREENKQIALGTS
682	696	KVVESKKKAVQRLEE	706	720	TDREENKQIALGTSK
683	697	VVESKKKAVQRLEEQ	707	721	DREENKQIALGTSKL
			708	722	REENKQIALGTSKLN
724	738	LDPRITVAWCKKWGV	709	723	EENKQIALGTSKLNY
725	739	DPRITVAWCKKWGVP	710	724	ENKQIALGTSKLNYL
726	740	PRITVAWCKKWGVPI	711	725	NKQIALGTSKLNYLD

727	741	RITVAWCKKWGVPIE	712	726	KQIALGTSKLNYLDP
739	753	PIEKIYNKTQREKFA	721	735	LNYLDPRITVAWCKK
740	754	IEKIYNKTQREKFAW	722	736	NYLDPRITVAWCKKW
741	755	EKIYNKTQREKFAWA	723	737	YLDPRITVAWCKKWG
742	756	KIYNKTQREKFAWAI			

*indicates peptides that overlapped with the peptides bound preferentially by risk alleles of HLA-DPB1.

Figure 1. Molecular dynamic simulation shows different modes of binding between the HLA-DP protein and a positively charged peptide. (a) A positively charged 12-mer peptide (Peptide A, KDREHRHKEHKK) from topoisomerase I had four positively charged amino acids at all contact positions (P1, P4, P6, and P9) interacting with the HLA-DP binding pocket. (b) The HLA-DP protein with three NCTs (HLA-DPB1*0901) and other two alleles with two NCTs (HLA-DPB1*1301 and 030101) interact more strongly with the topoisomerase I peptide than the HLA-DP protein without NCT (HLA-DPB1*0401).

(a)



(b)

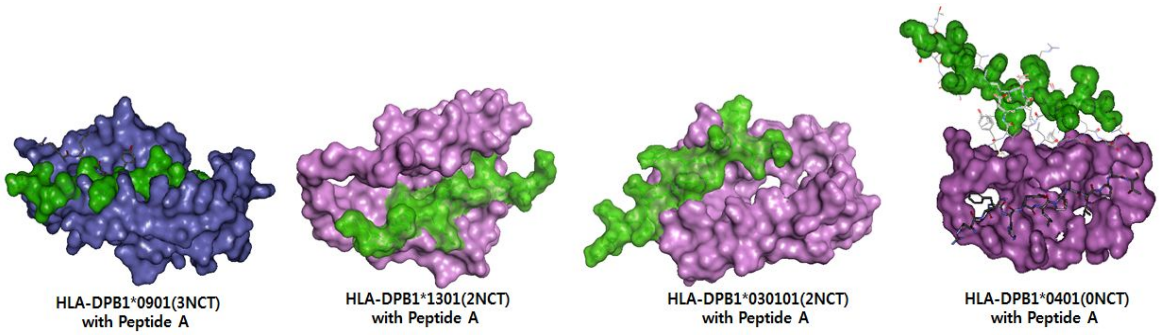
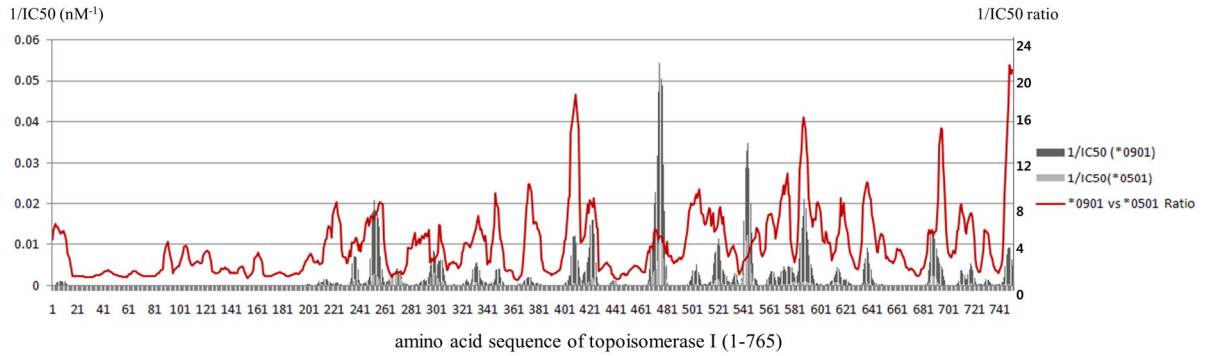
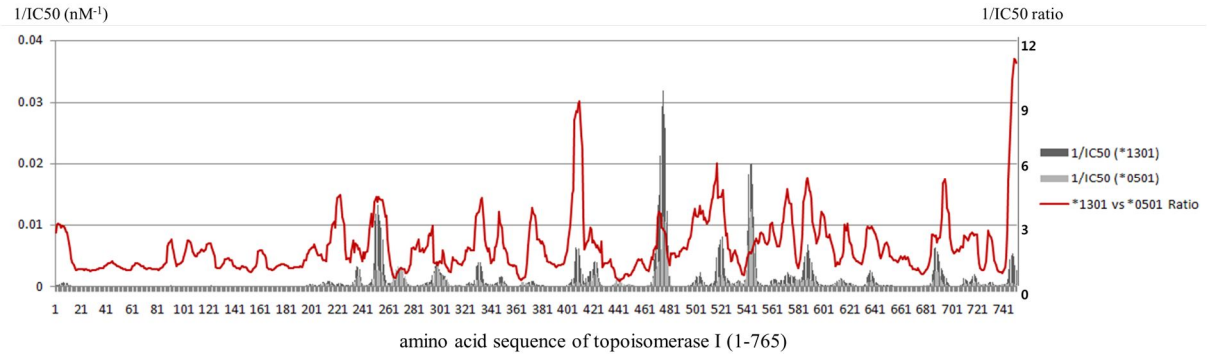


Figure 2. Predicted affinities of all possible 15-mer peptides of topoisomerase I (1-765). (a)-(c) The ratio (red line) of the 1/IC50 of the risk HLA-DPB1 alleles (*0901, *1301, *030101) (dark grey bars) to the protective HLA-DPB1 allele (*0501, light grey bars) is shown. All IC50s were calculated using the NetMHCIIpan-3.0 algorithm. (d) Venn diagram showing the number of peptides commonly favored by the risk alleles of ATA-positive SSc for 10% (75) or 20% (150) of the possible 751 15-mer peptides with the highest 1/IC50 ratio (selectively favors risk alleles).

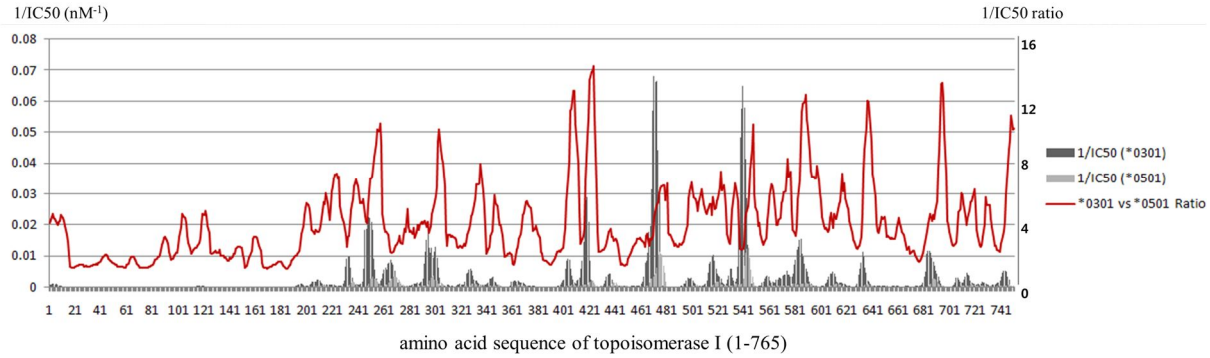
(a)



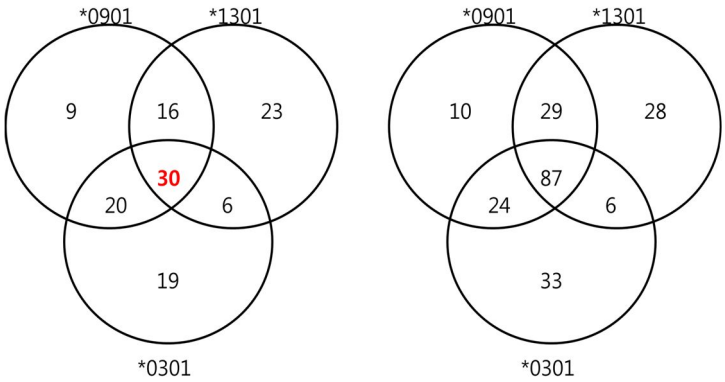
(b)



(c)



(d)

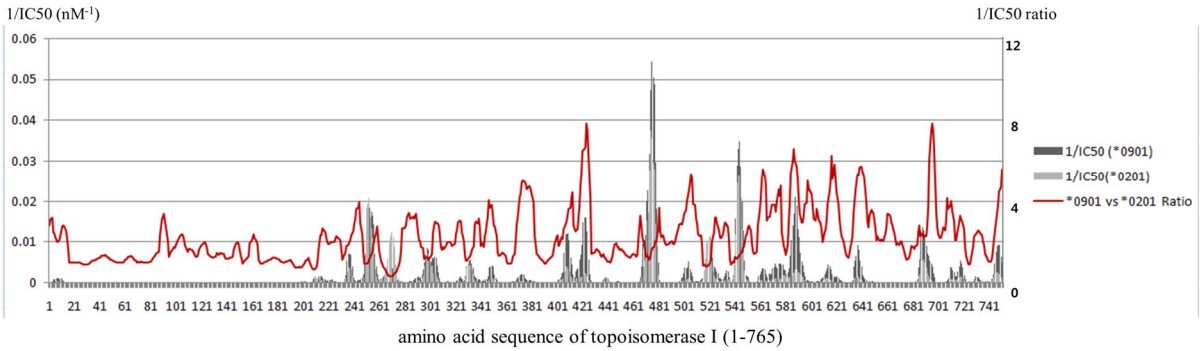


Concordance in upper 10% (75 peptides)

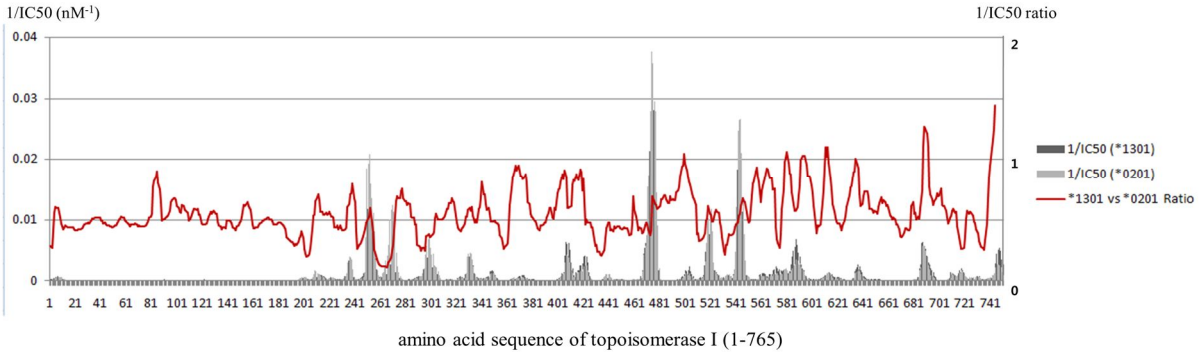
Concordance in upper 20% (150 peptides)

Figure 3. Predicted affinities of all possible 15-mer peptides of topoisomerase I (1-765). (a)-(c) The ratio (red line) of the 1/IC50 of the risk HLA-DPB1 alleles (*0901, *1301, *030101) (dark grey bars) to the protective HLA-DPB1 allele (*0201, light grey bars) is shown. All IC50s were calculated using the NetMHCIIpan-3.0 algorithm. (d) Venn diagram showing the number of peptides commonly favored by the risk alleles of ATA-positive SSc for 10% (75) or 20% (150) of the possible 751 15-mer peptides with the highest 1/IC50 ratio (selectively favors risk alleles).

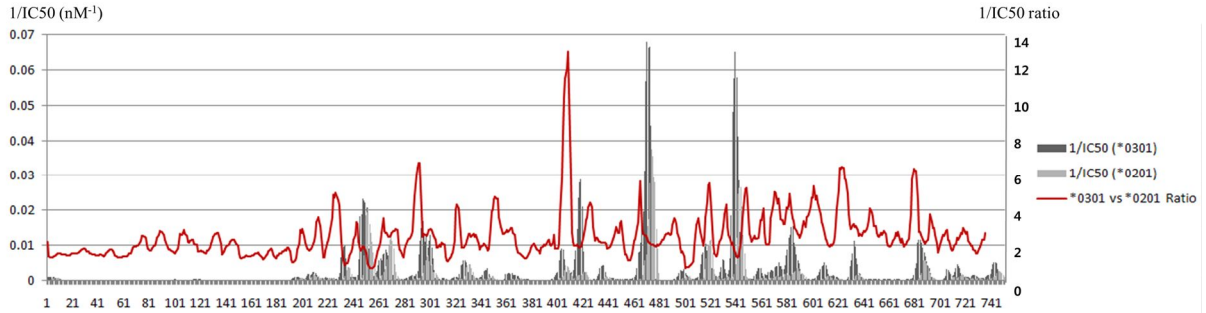
(a)



(b)

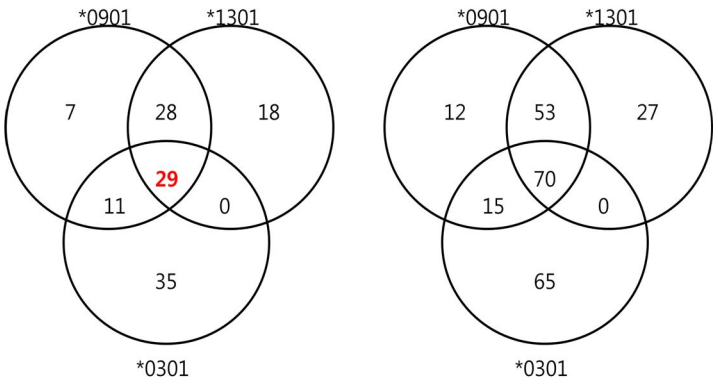


(c)



amino acid sequence of topoisomerase I (1-765)

(d)

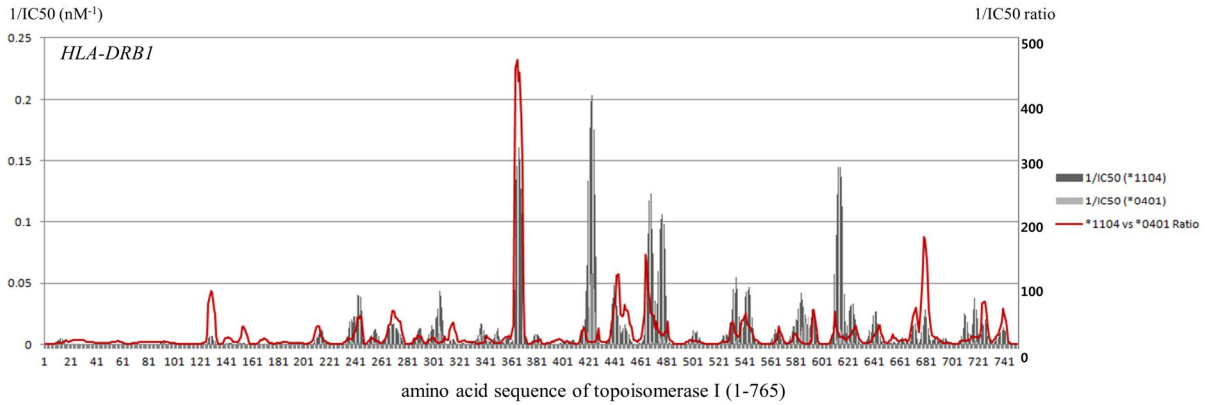


Concordance in upper 10% (75 peptides)

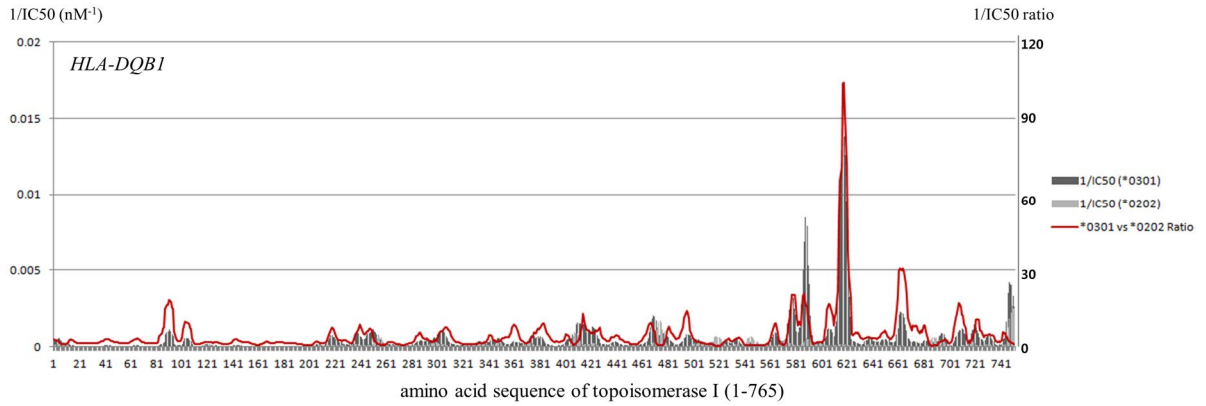
Concordance in upper 20% (150 peptides)

Figure 4. Predicted affinities of all possible 15-mer peptides of topoisomerase I (1-765) for high-risk alleles on other class II HLA (HLA-DRB1*1104 and DQB1*0301). (a) The ratio (red line) of the 1/IC50 of the risk HLA-DRB1*1104 allele (dark grey bars) to the protective HLA-DRB1 allele (*0401, light grey bars) is shown. (b) The ratio (red line) of the 1/IC50 of the risk HLA-DQB1*0301 allele (dark grey bars) to the protective HLA-DQB1 allele (*0202, light grey bars) is shown. All IC50s were calculated using the NetMHCIIpan-3.0 algorithm.

(a)



(b)



Discussion

Systemic sclerosis (SSc) with anti-topoisomerase I antibody (ATA) positivity was associated with the presence of specific class II HLA alleles. The HLA-DRB1 region is the most extensively studied and HLA-DRB1*1502 and *1104 have been found to be associated with ATA-positive SSc (3, 6). In addition, in Japanese and Korean patients, an association with the FLEDR sequence was reported (2, 23). However, previous genetic studies have found that the most susceptible loci for ATA-positive SSc is HLA-DP (4, 6). An extensive HLA polymorphism study in Americans revealed that the odds ratio at the HLA-DP region was highest among the different HLA-regions (6). Therefore, we focused on the HLA-DP region in this study.

Here we identified HLA-DPB1*0901, *1301, and *030101 as risk alleles for ATA-positive SSc and HLA-DPB1*020102 and *0501 as protective alleles against ATA-positive SSc. When we examined HLA-DP alleles in the context of supertype, HLA-DP6 was revealed to be a risk type while HLA-DP2 was found to be a protective type. At a fundamental level, SSc-associated HLA-DPB1 alleles shared a common binding motif against antigenic peptides. Negatively-charged triplets (NCTs) of amino acids appeared repeatedly at three critical places for binding antigenic peptides. Importantly, a dose-response relationship between the frequency of NCTs and the risk of ATA-positive SSc was observed. Along with the frequency of NCTs, position of NCTs was also important. NCT at position 82-85 of HLA-DPB1 played an anchoring role and an additional NCT at either position 67-69 or 82-85 was necessary to maximize the affinity between the HLA-DP protein and antigenic peptides. Molecular interactions between these negatively-charged amino acids on the HLA-DP protein and the positively charged fragment (amino acids 29-40, KDREHRHKEHKK) of topoisomerase

were modeled using molecular dynamic simulation. However, the peptide had a low affinity and less preference to the risk alleles of HLA-DPB1 in the predictions using the NetMHCIIpan3.0 algorithm. Moreover, the six peptide fragments with high selectivity against the risk alleles of HLA-DPB1 did not have intense positive charges. Since the binding peptide has a much larger molecular weight than a single element (e.g. beryllium) or small molecule drugs, spatial factors as well as electrical interactions between HLA-DP and binding peptides may also play an important role in the prediction process.

The significance of electrical charge in the interaction between HLA-DP and peptides was previously studied in chronic beryllium disease (17, 18). Since beryllium is a very small element with a positive charge, it plays a critical role in the binding pocket of HLA-DP2, particularly by interacting with the glutamic acid at position 69 of the HLA-DPB1 protein (17, 18). Abnormal immune activation via $CD4^+$ T cells has been reported previously and beryllium itself presented as a part of an antigenic peptide (24). It remains controversial whether SSc-specific T cells stimulated by topoisomerase I have pathologic roles or not. Kuwana *et al*. showed that topoisomerase I-specific T cells were stimulated by large topoisomerase I fragments (25). But the strengths of stimulation were not different among SSc patients with ATA positivity, SSc patients with ATA negativity, and healthy controls (25). In contrast, using exact fragmented peptides, Veeraraghaven *et al.* successfully demonstrated significant differences in epitope reactivity between SSc patients and healthy controls (26). Therefore, identifying the antigenic peptides preferred dominantly by the risk alleles for ATA-positive SSc is essential for elucidating the role of $CD4^+$ T cells in the autoimmunity that occurs in SSc.

In order to determine the exact fragments of topoisomerase I that may act as antigenic peptides and contribute to the autoimmune response of SSc, the maximal discrepancy of

peptide binding affinity between the risk and protective alleles of HLA-DPB1 was predicted using computational analysis software (NetMHCIIpan-3.0 algorithm). As a result, six peptides originating from topoisomerase I (585-599 to 590-604: TAKVFRTYNASITLQ, AKVFRTYNASITLQQ, KVFRTYNASITLQQQ, VFRTYNASITLQQQL, FRTYNASITLQ-QQLK, and RTYNASITLQQQLKE) were predicted to bind selectively to the risk alleles of HLA-DPB1. Previously identified HLA-DQB1 alleles associated with increased risk of ATA-positive SSc also had significant affinity to the six fragments. Interestingly, one of these fragments, AKVFRTYNASITLQQQL (586-602), was previously shown to frequently display reactivity against SSc patients (26). Therefore, this fragment may have strong immunogenicity in the SSc patients with ATA and a risk allele.

In a recent study of sclerotic graft-versus-host disease (GVHD), GPM (glycine-proline-methionine) motifs at positions 85-87 of the DP beta chain were linked to decreased risk (27). This is the same position that has anchoring role in the risk of ATA-positive SSc. At positions 85-87, only two versions of amino acid sequences have been found. The first is the GPM sequence and the other is EAV (glutamic acid-alanine-valine), which constitutes NCT of positions 82/84/85 of the DP beta chain. Therefore, sclerotic GVHD is also associated with an NCT at the same critical position of HLA-DPB1 as systemic sclerosis. The role of topoisomerase I or ATA status of sclerotic GVHD patients remains unknown (27). Expression of the sclerotic phenotype itself may not be exclusively related to topoisomerase I, but related to other antigens presented in the GVHD reaction.

Future studies are needed to demonstrate activation of $CD4^+$ T cells by the predicted fragments of topoisomerase I and the antigen presenting cells with the risk alleles of HLA-DPB1. Attenuation of the stimulation by an anti-HLA-DP antibody would also be necessary as in a previous study of HLA-DR and –DQ (3). Whether those activated $CD4^+$ T cells can

act as triggers of autoimmune and fibrotic reactions is also an interesting topic. Topoisomerase I-reactive T cells are activated and clonally expanded in the SSc patient (25, 28). However, the difference in phenotypic and functional properties of the activated T cells obtained from the SSc patient and the healthy control was subtle (28). In one mouse model of SSc, topoisomerase I injection along with a strong adjuvant induced skin and lung fibrosis and autoimmunity with a high ATA titer (29).

This study has several limitations. First, the fragments of topoisomerase I in this study were predicted to preferentially bind to the specific risk alleles of HLA-DPB1 as a part of an autoimmune reaction among HLA-DP and pathogenic $CD4^+$ T cells. However, this remains to be demonstrated in vitro or in vivo with SSc-specific pathogenic $CD4^+$ T-cells or topoisomerase I specific $CD4^+$ T cells. Second, all polymorphism data was obtained from a Korean population. Ethnic differences in SSc-related HLA alleles have been well documented in previous studies (6, 30). Although HLA-DPB1*1301 displayed the strongest association in two previous GWAS studies, other HLA-DRB1 and DQB1 alleles were also positively or negatively correlated to SSc in non-Asian populations (4, 6). Finally, HLA-DR and DQ genotypes of the SSc patients and healthy controls were not investigated in this study. Gene-gene interactions among HLA-DR, DQ, and DP have been previously investigated in rheumatoid arthritis (31, 32). Considering the evidence of an association between HLA-DR/DQ and ATA positive-SSc, the gene-gene interactions may account for the discrepancies between the results of former genetic studies and GWAS (3-6, 11, 12).

In conclusion, ATA-positive SSc-susceptible HLA-DPB1 alleles share NCTs at critical positions in the peptide binding groove. A group of peptide fragments of topoisomerase I was identified based on the discrepancy of affinity between risk alleles and protective alleles of ATA positive SSc. A subset of those peptides was also favored by risk alleles of HLA-DQB1,

suggesting that they may play a role in the pathogenesis of SSc.

References

1. Gu YS, Kong J, Cheema GS, Keen CL, Wick G, Gershwin ME. The immunobiology of systemic sclerosis. Semin Arthritis Rheum. 2008;38(2):132-60.

2. Broen JC, Radstake TR, Rossato M. The role of genetics and epigenetics in the pathogenesis of systemic sclerosis. Nature reviews Rheumatology. 2014;10(11):671-81.

3. Kuwana M, Kaburaki J, Okano Y, Inoko H, Tsuji K. The HLA-DR and DQ genes control the autoimmune response to DNA topoisomerase I in systemic sclerosis (scleroderma). J Clin Invest. 1993;92(3):1296-301.

4. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum. 2009;60(12):3807-14.

5. Gorlova O, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS genetics. 2011;7(7):e1002178.

6. Arnett FC, Gourh P, Shete S, Ahn CW, Honey RE, Agarwal SK, et al. 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.

7. Begovich AB, Bugawan TL, Nepom BS, Klitz W, Nepom GT, Erlich HA. A specific HLA-DP beta allele is associated with pauciarticular juvenile rheumatoid arthritis but not adult rheumatoid arthritis. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(23):9489-93.

8. Dong RP, Kimura A, Okubo R, Shinagawa H, Tamai H, Nishimura Y, et al. HLA-A

and DPB1 loci confer susceptibility to Graves' disease. Human immunology. 1992;35(3):165-72.

9. Dong RP, Kimura A, Numano F, Yajima M, Hashimoto Y, Kishi Y, et al. HLA-DP antigen and Takayasu arteritis. Tissue antigens. 1992;39(3):106-10.

10. Richeldi L, Sorrentino R, Saltini C. HLA-DPB1 glutamate 69: a genetic marker of beryllium disease. Science. 1993;262(5131):242-4.

11. Rihs HP, Conrad K, Mehlhorn J, May-Taube K, Welticke B, Frank KH, et al. Molecular analysis of HLA-DPB1 alleles in idiopathic systemic sclerosis patients and uranium miners with systemic sclerosis. International archives of allergy and immunology. 1996;109(3):216-22.

12. Gilchrist FC, Bunn C, Foley PJ, Lympany PA, Black CM, Welsh KI, et al. Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. Genes and immunity. 2001;2(2):76-81.

13. Dai S, Murphy GA, Crawford F, Mack DG, Falta MT, Marrack P, et al. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(16):7425-30.

14. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC bioinformatics. 2010;11:568.

15. Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65(10):711-24.

16. Stewart L, Ireton GC, Champoux JJ. The domain organization of human topoisomerase I. The Journal of biological chemistry. 1996;271(13):7602-8.

17. Scott BL, Wang Z, Marrone BL, Sauer NN. Potential binding modes of beryllium

with the class II major histocompatibility complex HLA-DP: a combined theoretical and structural database study. J Inorg Biochem. 2003;94(1-2):5-13.

18. Diaz G, Amicosante M, Jaraquemada D, Butler RH, Guillen MV, Sanchez M, et al. Functional analysis of HLA-DP polymorphism: a crucial role for DPbeta residues 9, 11, 35, 55, 56, 69 and 84-87 in T cell allorecognition and peptide binding. Int Immunol. 2003;15(5):565-76.

19. Cano P, Fernandez-Vina M. Two sequence dimorphisms of DPB1 define the immunodominant serologic epitopes of HLA-DP. Human immunology. 2009;70(10):836-43.

20. Doytchinova IA, Flower DR. In silico identification of supertypes for class II MHCs. J Immunol. 2005;174(11):7085-95.

21. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins. 2003;52(1):80-7.

22. Li L, Chen R, Weng Z. RDOCK: refinement of rigid-body protein docking predictions. Proteins. 2003;53(3):693-707.

23. Kang SH, Park MH, Song EY, Kang SJ, Lee EB, Song YW, et al. Association of HLA class II genes with systemic sclerosis in Koreans. The Journal of rheumatology. 2001;28(7):1577-83.

24. Fontenot AP, Torres M, Marshall WH, Newman LS, Kotzin BL. Beryllium presentation to CD4+ T cells underlies disease-susceptibility HLA-DP alleles in chronic beryllium disease. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(23):12717-22.

25. Kuwana M, Medsger TA, Jr., Wright TM. T cell proliferative response induced by DNA topoisomerase I in patients with systemic sclerosis and healthy donors. J Clin Invest. 1995;96(1):586-96.

26. Veeraraghavan S, Renzoni EA, Jeal H, Jones M, Hammer J, Wells AU, et al. Mapping of the immunodominant T cell epitopes of the protein topoisomerase I. Ann Rheum Dis. 2004;63(8):982-7.

27. Inamoto Y, Martin PJ, Flowers ME, Lee SJ, Carpenter PA, Warren III EH, et al. Genetic Risk Factors for Sclerotic Graft-Versus-Host Disease. 56th ASH annual meeting and exposition. 2014;Oral and Poster Abstracts(Session: 722):665.

28. Kuwana M, Feghali CA, Medsger TA, Jr., Wright TM. Autoreactive T cells to topoisomerase I in monozygotic twins discordant for systemic sclerosis. Arthritis Rheum. 2001;44(7):1654-9.

29. Yoshizaki A, Yanaba K, Ogawa A, Asano Y, Kadono T, Sato S. Immunization with DNA topoisomerase I and Freund's complete adjuvant induces skin and lung fibrosis and autoimmunity via interleukin-6 signaling. Arthritis Rheum. 2011;63(11):3575-85.

30. Kuwana M, Kaburaki J, Arnett FC, Howard RF, Medsger TA, Jr., Wright TM. Influence of ethnic background on clinical and serologic features in patients with systemic sclerosis and anti-DNA topoisomerase I antibody. Arthritis Rheum. 1999;42(3):465-74.

31. Seidl C, Kasser UR, Fischer B, Koch U, Meier L, Fischer P, et al. HLA-DR/DQ interaction in patients with erosive rheumatoid arthritis presenting articular and extraarticular disease manifestations. European journal of immunogenetics : official journal of the British Society for Histocompatibility and Immunogenetics. 1999;26(1):19-27.

32. Seidl C, Koch U, Brunnler G, Buhleier T, Frank R, Moller B, et al. HLA-DR/DQ/DP interactions in rheumatoid arthritis. European journal of immunogenetics : official journal of the British Society for Histocompatibility and Immunogenetics. 1997;24(5):365-76.

국문 초록

목표: 전신성 경화증의 위험이 높은 인간 백혈구 항원(HLA)-DPB1 대립인자 유형들을 비교하여 아미노산 서열의 공통적인 특징을 규명하고, 자가항원으로 의심되는 토포아이소머레이스-I의 아미노산 서열 중 병원성 펩티드로 작용할 수 있는 부분을 예측하는 것을 이 연구의 목표로 하였다.

방법: 127명의 한국인 전신성 경화증 환자와 548명의 건강한 한국인 대조군을 연구의 대상으로 하였다. 전신성 경화증 위험이 높은 HLA-DPB1 대립인자들의 아미노산 서열에서, 펩티드 결합부위를 중심으로 공통점을 분석하였다. 또한 토포아이소머레이스-I의 전체 아미노산 서열을 15개 아미노산 단위로 나눈 조각들을 대상으로, HLA-DP 각 대립인자 유형들과의 결합친화력을 NetMHCIIpan-3.0 프로그램을 통해 계산하였다. 이러한 기법을 이용하여, 전신성 경화증에 보호 효과를 갖는 HLA-DPB1 대립인자 대비 위험이 높은 대립인자에 대한 결합친화력 간의 비율을 산출하였고, 이 비율이 높은 토포아이소머레이스 조각을 병원성 펩티드로 작용할 수 있을 것으로 예측하였다.

결과: HLA-DPB1*0901, *1301, *030101 대립인자들이 항-토포아이소머레이스-I 항체(ATA) 양성을 보이는 전신성 경화증에서 고위험 대립인자들이었다. 대조적으로, HLA-DPB1*020102, *0501 대립인자들은 전신성 경화증에 대하여 보호 효과를 갖는 것으로 나타났다. HLA-DP의 슈퍼타입을 기준으로 재분석했을 때, HLA-DP6가 고위험 슈퍼타입이었다. HLA-DPB1의 아미노산 서열 중 55-57, 67-69, 82, 84-85번째 위치에서 음전하를 띄는 아미노산(아스파르트산,

글루탐산)이 연속하여 세 개씩 나타나는(이하 음전하 트리플렛) 대립인자들이 발견되었는데, 이러한 음전하 트리플렛이 많이 나타날수록 ATA양성 전신성 경화증의 위험이 증가하는 것으로 보였다. 음전하 트리플렛의 위치도 중요했으며, 특히 HLA-DPB1 아미노산 서열의 82, 84, 85번째 위치에 나타나는 음전하 트리플렛이 전신성 경화증 위험과의 연관성이 깊었다. HLA-DP의 고위험 대립인자들과 강한 친화도를 보이는 병인성 펩티드 조각들은 과거 알려졌던 HLA-DQ의 고위험 대립인자들과도 일부 강한 친화도를 보였으나 HLA-DR의 고위험 대립인자들과는 관련이 없었다.

결론: ATA양성 전신성 경화증의 위험이 높은 HLA-DPB1 대립인자들은 펩티드와 결합하는 부위에 음전하 트리플렛을 공유하고 있었다. 토포아이소머-레이스-I의 펩티드 조각들 중 HLA-DPB1과 관련하여 병인성을 가진다고 알려진 것들은 HLA-DQB1와도 관련이 있었으며, 이 펩티드들은 전신성 경화증의 발병과정에 일정한 역할을 할 것으로 보인다.

주요어: 전신성 경화증, 인간 백혈구 항원, 토포아이소머레이스-I, 항원 결정기

학번: 2011-21896