# T-Cell Epitopes of the La/SSB Autoantigen: Prediction Based on the Homology Modeling of HLA-DQ2/DQ7 with the Insulin-B Peptide/HLA-DQ8 Complex

AGGELIKI KOSMOPOULOU,[1] METAXIA VLASSI,[2] ATHANASSIOS STAVRAKOUDIS,[1,3] CONSTANTINOS SAKARELLOS,[1] MARIA SAKARELLOS-DAITSIOTIS[1]

[1]*Laboratory of Peptide Chemistry, Department of Chemistry, University of Ioannina, Ioannina 44110, Greece*

[2]*Institute of Biology, National Center for Research "Demokritos," P.O. Box 60228, Agia Paraskevi 15310, Greece*

[3]*Department of Economics, University of Ioannina, Ioannina 44110, Greece*

*Received 1 June 2005; Accepted 28 January 2006*
*DOI 10.1002/jcc.20422*
*Published online in Wiley InterScience (www.interscience.wiley.com).*

**Abstract:** T-cell epitopes are important components of the inappropriate response of the immune system to self-proteins in autoimmune diseases. In this study, the candidate T-cell epitopes of the La/SSB autoantigen, the main target of the autoimmune response in patients with Sjogren's Syndrome (SS), and Systemic Lupus Erythematosus (SLE) were predicted using as a template the HLA-DQ2 and DQ7 molecules, which are genetically linked to patients with SS and SLE. Modeling of DQ2 and DQ7 was based on the crystal structure of HLA-DQ8, an HLA molecule of high risk factor of type I diabetes, which is also an autoimmune disease. The quality and reliability of the modeled DQ2 and DQ7 was confirmed by the Ramachandran plot and the TINKER molecular modeling software. Common and/or similar candidate T-cell epitopes, obtained by comparing three different approaches the Taylor's sequence pattern, the TEPITOPE quantitative matrices, and the MULTIPRED artificial neural network, were subjected to homology modeling with the crystal structure of the insulin-B peptide complexed with HLA-DQ8, and the best superposed candidate epitopes were placed into the modeled HLA-DQ2 and DQ7 binding grooves to perform energy minimization calculations. Six T-cell epitopes were predicted for HLA-DQ7 and nine for HLA-DQ2 covering parts of the amino-terminal and the central regions of the La/SSB autoantigen. Residues corresponding to the P1, P4, and P9 pockets of the HLA-DQ2 and DQ7 binding grooves experience very low SASA because they are less exposed to the microenvironment of the groove. The proposed T-cell epitopes complexed with HLA-DQ2/DQ7 were further evaluated for their binding efficiency according to their potential interaction energy, binding affinity, and $IC_{50}$ values. Our approach constitutes the ground work for a rapid and reliable experimentation concerning the T-cell epitope mapping of autoantigens, and could lead to the development of T-cell inhibitors as immunotherapeutics in autoimmune diseases.

 J Comput Chem 27: 1033–1044, 2006

**Key words:** prediction of T-cell epitopes; La/SSB autoantigen; modeling of HLA-DQ2 and HLA-DQ7 molecules; Sjogren's Syndrome (SS); Systemic Lupus Erythematosus (SLE)

## Introduction

T cells play a crucial role in specific immune responses against diseases such as infectious diseases, cancer, allergies, and autoimmune diseases. Identification of T-cell epitopes is an important step in the design of vaccines and in developing immunotherapeutic drug-like inhibitors of T-cell epitopes against the inappropriate response of the immune system to self-antigens. However, T cells recognize antigen only after it is processed and presented on the antigen presenting cells (APCs) in association with histocompatibility glycoproteins. This process involves major histocompatibility complex (MHC) molecules, which first bind short peptides and then display them to the T cell-specific receptors resulting in the initiation of the immune response.[1–4]

***Correspondence to:*** A. Stavrakoudis; e-mail: astavrak@cc.uoi.gr

Contract/grant sponsor: Greek General Secretariat for Research and Technology (PENED 2001)

The ability of the immune system to respond to a particular antigen varies between individuals according to their different pattern of MHC genes, which shows a remarkable polymorphism with a great number of variants of HLA class I and II molecules that have been characterized and named to date.[5] Peptide binding to the peptide-binding groove of MHC is effected by a network of hydrogen bonds between the peptide backbone and the binding groove, and through interactions of peptide side chains, anchors, and pockets of the MHC cavity. By contrast to the class I MHC binding groove, which has closed ends restricting the lengths of bound peptides to 8–11 amino acids, class II MHCs have open ends, and thus present peptides as long as 30 amino acids, although the binding core of the peptide bound within the groove is only nine amino residues. Although peptide binding to MHC is a prerequisite for T-cell recognition, it is not sufficient by itself, as it also depends on the interaction of the displayed peptide with the T-cell receptor[3,6–8] (TCR).

However, several approaches have traditionally been used for HLA-based epitope prediction.[9,10] For example, the sequence similarity between a known MHC-binding peptide and known T-cell epitopes or the binding motifs that describe amino acids commonly occurring at particular positions within peptides that bind to a specific MHC molecule.[11–13] Sequence similarity and binding motifs are simple to implement and easy to understand, but they are of relatively modest accuracy.[9,10]

Quantitative matrices provide coefficients for each amino acid and each position within the peptide, and the method is based on the assumption that (1) each position within the peptide contributes independently to binding to an MHC, and (2) a residue located at a given peptide position contributes an equal amount to binding, even within different peptides. The quantitative matrix method represents an extension of binding motifs, is an accurate prediction method although the independent contribution-to-binding assumption ignores the contribution of the overall peptide structure to binding.[14,15]

Artificial neural networks (ANNs) are based on models that can capture complex relationships in the data sets and are not based on assumptions such as quantitative matrices. ANNs are adaptive and can self-improve; they generalize well, and are effective with nonlinear problems. ANNSs are accurate predictors, and they require larger amounts of binding data than simpler prediction methods.[16,17]

Although beneficial in the case of pathogens, MHC molecules are also thought to play a crucial role in the development of autoimmune diseases, as many of them are found more frequently in individuals expressing particular MHC molecules. Well-known examples are rheumatoid arthritis, type I diabetes, celiac disease, and ankylosing spondylitis.[18–21] In particular, autoimmune rheumatic diseases as Sjogren's Syndrome (SS) and Systemic Lupus Erythematosus (SLE) are genetically linked to DQ2 (HLA-DQA1*0501/DQB1*0201) and DQ7 (HLA-DQA1*0501/DQB1*0301).[22–24] Several lines of evidence indicate also that DQ8 is a major risk factor for increased susceptibility to type I diabetes,[25,26] which is also an autoimmune disease.

The $\alpha_1\beta_1$ domain of the DQ molecules exhibit the basic fold of histocompatibility molecules, that is, a "floor" of eight $\beta$-pleated sheets bound on two sides by two antiparallel $\alpha$-helices.[6,8] There is a 55–60% identity in amino acid positions in the $\alpha_1\beta_1$ domain of any DQ molecule, and 75% identity in the same domain within the various DQ molecules.[27] Despite the polymorphism in both $\alpha$- and $\beta$-chains of DQ alleles, all of the invariant residues deemed crucial for forming hydrogen bonds to the foreign peptide backbone in the case of DR1 are found in the same positions in DQ.[3,28] Furthermore, the antigen-binding groove of the DQ molecules fulfills the "pocket" motif of this region for the class II MHC molecules. All the depression in the antigen-binding groove may not belong to pockets, so the number of pockets and the residues constituting each pocket may slightly vary depending on the nature of the antigenic peptide bound.[3,8]

The aim of our study is to predict with accuracy the T-cell epitope candidates of the La/SSB autoantigen, the main target of the autoimmune response in patients with SS and SLE,[29,30] in regard to the HLA-DQ2 and DQ7 molecules. It deserves to be mentioned that the chemical structures of the $\alpha$1 and $\beta$1 chains of DQ2 and DQ7 experience a great similarity with DQ8. HLA-DQ2/DQ7 were modeled on the basis of the crystal structure of HLA-DQ8,[31] and their compatibility was verified by the Procheck suite of programs,[32] the "O" program[33] and the TINKER molecular modeling software.[34] A general sequence pattern common to HLA-based T-cell epitopes, determined by examining the residues, which make up epitopes as members of characteristic sets based on their physical properties, was applied (Taylor) to assess epitopes of the La/SSB autoantigen.[11] The TEPITOPE[10,35,36] program that uses structural data from MHC–peptide complexes and incorporates multiple quantitative matrices, was next employed to predict the potential of the La/SSB epitopes to bind DQ7 and DQ2. A new Web-based computational system, MULTIPRED,[37] using ANN, was also applied to predict the DQ2 and DQ7 binding peptides of La/SSB. Common and/or similar T-cell epitopes, displayed by comparing Taylor with TEPITOPE and MULTIPRED, were superposed to the insulin-B peptide epitope. The best aligned epitopes were placed into the binding groove of the modeled DQ2 and DQ7 to perform energy minimization studies. Consistent with the obtained candidate T-cell epitopes of the La/SSB autoantigen for the DQ2 and DQ7 molecules are our findings concerning the solvent-accessible surface area (SASA) of the epitopes complexed to the HLA molecules using GETAREA program.[38]

To further confirm our findings, the potential interaction energy, $\Delta E_{int}$, of the MHC–epitope complexes was estimated, using the TINKER software,[34] to acquire a relative energy scale that reflects the degree of the preferred interactions. On the other hand, taking in consideration that only peptides that bind to MHC with affinity above a threshold function as T-cell epitopes, a quantitative prediction of the MHC–epitope binding affinities was also performed, using MHCPred, to obtain the $IC_{50}$ values that are inversely proportional to the binding affinities.[39–41]

## Computational Methods

### *Sequence similarity of HLA-DQ2 and HLA-DQ7 with HLA-DQ8*

The sequence similarity of the HLA-DQ2 and DQ7 molecules with HLA-DQ8 was performed using the Basic Local Alignment Search Tool (BLAST) (www.ncbi.nlm.nih.gov/BLAST/).[42] The

HLA-DQ8 (PDB code:1JK8) sequence was obtained from the PDB database.[43]

### Homology Modeling of HLA-DQ2 and HLA-DQ7 with the Crystal Structure of the HLA-DQ8

The initial models of DQ2 and DQ7 were obtained using the Swiss-PdbViewer program,[44] which allows the user to analyze several proteins at the same time and is tightly linked to Swiss-Model,[45] an automated homology modeling server. Using these two programs, it was possible to thread the DQ2 and DQ7 primary sequences onto the DQ8 3D template and get an immediate feedback of how well the threaded proteins would be accepted by the reference structure before submitting a request to build missing loops and refine side chain packing.

### Evaluation of the Homology Model

The Procheck suite of programs[32] was used for assessing the stereochemical quality of the modeled HLA-DQ2 and HLA-DQ7 structures at resolution 2Å. Finally, the compatibility of the modeled molecules was determined using the "O" program.[33] The TINKER molecular modeling software[34] was used to superimpose the two molecular structures (DQ8/DQ2 and DQ8/DQ7) in 3D. The superimposition was unit-weighted and the coordinates of the second structure were superimposed on the first structure. Energy minimizations were carried out using the CHARMM27 potential energy function[46,47] with the GB/SA continuum model for solvation.[48,49] The dielectric constant was set 1.0. The minimize procedure was performed with a limited memory L-BFGS minimization algorithm. The final RMS gradient was less than 0.01 kcal $mol^{-1}$ $Å^{-1}$ for all the minimized structures. VDW- and CHG-cutoffs were set 12 and 15 Å, respectively. The homology modeling procedure was performed on a Silicon Graphics (INDIGO2) workstation. TINKER calculations, superimpositions and energy minimizations, were performed on Linux machine with Dual processor running Mandrake 10.0.

### T-Cell Epitope Prediction

The sequence pattern approach (Taylor)[35] was applied to detect the potential candidate T-cell epitopes of the La/SSB autoantigen. The La/SSB sequence was provided in FASTA format from Swiss-Prot database (www.expasy.ch), and it was scanned using the PATTINPROT program for detecting the characteristic sequence patterns. The similarity level of the detected patterns during the scanning was 100%.

The TEPITOPE program,[10,35,36] a quantitative matrix method, was used to predict potential candidate T-cell epitopes in the La/SSB sequence. La/SSB sequence (408aa) was loaded into the TEPITOPE sequence editor and the prediction conditions such as HLA-DR alleles (25 HLA-II virtual matrices) and the % threshold value were selected and modified from the default setting. Fourteen HLA-DR alleles were selected (DRB1*0101, DRB1*0102, DRB1*0301, DRB1*0401, DRB1*0402, DRB1*0701, DRB1*0801, DRB1*0802, DRB1*1101, DRB1*1104, DRB1*1305, DRB1*1501, DRB1*1502, DRB5*0101), and the TEPITOPE prediction threshold was set at 4%.

The MULTIPRED computational system[37] was also used to predict the candidate T-cell epitopes of the La/SSB autoantigen. The artificial neural networks (ANNs) were selected as the prediction method for class II DR supertype. HLA class II DR hotspots were predicted by ANN method at threshold 20 for average value. MULTIPRED is available at http://antigen.i2r.a-star.edu.sg/multipred/.

### Prediction of the SASA of the Proposed Epitopes within the HLA -DQ2 and HLA-DQ7 Molecules

GETAREA Web service[38] was used for the calculation of the SASA of the proposed epitopes of the La/SSB autoantigen. The Cartesian coordinates of the atoms of the proposed candidate epitopes, complexed with DQ2 and DQ7 models, were stored in PDB format and the radius of the water probe was set 1.4 Å.

### Prediction of the Epitope Binding Affinity to HLA-DQ2 and HLA-DQ7

MHCPred was used for the quantitative prediction of the peptide-HLA-DQ2/DQ7 binding affinity. MHCPred is composed of a variety of widely distributed human allele specific Quantitative Structure Activity Relationship (QSAR) models built using partial least squares (PLS). In this study the DRB0101 HLA allele was used, which has the highest sequence similarity with the DQ2 and DQ7 molecules. The binding affinity prediction was based on the contributions of the amino acids and their interactions, within the 9-mer core, of the proposed epitopes that correspond to P1, P4, and P9 binding pockets of DQ2 and DQ7. $IC_{50}$ values of insulin-B peptide and the predicted epitopes were calculated to measure their binding affinity to the corresponding HLA-DQ molecules. MHCPred is available from the URL: http://www.jenner.ac.uk/MHCPred.

## Results and Discussion

### HLA-DQ2 and HLA-DQ7 Homology Modeling with the Crystal Structure of HLA-DQ8

The homology modeling of the HLA-DQ2 and DQ7 molecules was based on the crystal structure of the HLA-DQ8 molecule. It was found that HLA-DQ2 experiences high-sequence similarity with HLA-DQ8: 91% identity for chain $\beta$ (DQB1*0201) and 90% identity for chain $\alpha$ (DQA1*0501). Furthermore, HLA-DQ7 is almost identical with HLA-DQ8: 96% identity for chain $\beta$ (DQB1*0301) and 90% identity for chain $\alpha$ (DQA1*0501). In these conditions, a homology modeling strategy can be applied with very good results, because the alignment between the new sequence and the reference model[50] can be obtained with few or no gaps (Fig. 1).

The initial models of DQ2 and DQ7 were obtained using the Swiss-PdbViewer program. DQ2 and DQ7 were superposed with DQ8 molecule to get the homology modeling requested, deduce the structural alignments, and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles, and distances between atoms were obtained thanks to the intuitive

**A**

```
α-chain DQ7 VADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYVDLGRKETVWCLPELRQFRGFDPQFAL  63
α-chain DQ2 VADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYVDLGRKETVWCLPELRQFRGFDPQFAL  63
α-chain DQ8 VADHVASYGVNLYQSYGPSGQYSHEFDGDEEFYVDLERKETVWQLPLFRRFRRFDPQFAL  60

α-chain DQ7 TNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNIFPPVVNIT  123
α-chain DQ2 TNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNIFPPVVNIT  123
α-chain DQ8 TNIAVLKHNLNIVIKRSNSTAATNEVPEVTVFSKSPVTLGQPNTLICLVDNIFPPVVNIT  120

α-chain DQ7 WLTNGHSVTEGVSETTFLSKSDHSFFKISYLTLLPSAEESYDCKVEHWGLDKPLLKHWEP  183
α-chain DQ2 WLTNGHSVTEGVSETTFLSKSDHSFFKISYLTLLPSAEESYDCKVEHWGLDKPLLKHWEP  183
α-chain DQ8 WLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFLPSDDEIYDCKVEHWGLDEPLLKHWEP  180

α-chain DQ7 E  184
α-chain DQ2 E  184
α-chain DQ8 E  181
```

**B**

```
β-chain DQ2 SPEDFVYQFKGMCYFTNGTERVRLVSRSIYNREEIVRFDSDVGEFRAVTLLGLPAAEYWN  62
β-chain DQ8 SPEDFVYQFKGMCYFTNGTERVRLVTRYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWN  60

β-chain DQ2 SQKDILERKRAAVDRVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDF  122
β-chain DQ8 SQKEVLERTRAELDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDF  120

β-chain DQ2 YPAQIKVRWFRNDQEETAGVVSTPLIRNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQS  182
β-chain DQ8 YPAQIKVRWFRNDQEETTGVVSTPLIRNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQN  180

β-chain DQ2 PITVEWRAQS  192
β-chain DQ8 PIIVEWRAQS  190
```

**C**

```
β-chain DQ7 SPEDFVYQFKAMCYFTNGTERVRYVTRYIYNREEYARFDSDVEVYRAVTPLGPPDAEYWN  62
β-chain DQ8 SPEDFVYQFKGMCYFTNGTERVRLVTRYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWN  60

β-chain DQ7 SQKEVLERTRAELDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDF  122
β-chain DQ8 SQKEVLERTRAELDTVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDF  120

β-chain DQ7 YPAQIKVRWFRNDQEETTGVVSTPLIRNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQN  182
β-chain DQ8 YPAQIKVRWFRNDQEETTGVVSTPLIRNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQN  180

β-chain DQ7 PITVEWRAQS  192
β-chain DQ8 PIIVEWRAQS  190
```

**Figure 1.** Comparison of the $\alpha$-chain of DQ2 and DQ7 ($\alpha$-chain of DQ2 and DQ7 share 100% identity) with DQ8 (A), and the $\beta$-chain of DQ with DQ8 (B) and the $\beta$-chain of DQ7 with DQ8 (C). Different amino acids between $\alpha$- and $\beta$-chains are shown in bold. Amino acids are in one-letter code.

graphic and menu interface. Automatic energy minimization was in place after every superposition of the two molecules (DQ8 with DQ2 and DQ8 with DQ7) to regularize deviations in the protein structure geometry.

The Procheck suite of programs was used for assessing the stereochemical quality of the modeled HLA-DQ2 and HLA-DQ7 structures and the compatibility of the modeled molecules was checked using the "O" program. The quality of the models is similar to the quality expected for structures determined at 2 Å. In the Ramachandran plot most of the residues were found in preferred regions (DQ8: 98.1%, DQ2: 97.8% and DQ7: 98.1% residues), and only a small percentage in disallowed region, which is also observed in the equivalent residues of the template used for the modeling. The RMSD was calculated to 0.034 and 0.033 for DQ8/DQ2 and DQ8/DQ7, respectively, taking in consideration the $C^{\alpha}$ atoms (Fig. 2).

### *T-Cell Epitope Prediction using Taylor's (1988) Sequence Pattern*

The analysis of known MHC-binding peptides as cytotoxic and helper T-cell epitopes revealed a common pattern in their primary structure: *C-H-H-C, G-H-H-C, G-H-H-P, P-H-H-P, C-H-H-P, C-H-H-H-C, P-H-H-C, C-H-P-C, G-P-H-H-P, G-P-H-P, C-H-P-P, C-H-P-H-P, C-P-H-C* (**C**: charged residues, **H**: hydrophobic residues, **P**: polar residues and **G**: glycine). To detect potential candidate T-cell epitopes, the occurrence of these patterns in the primary structure of the La/SSB autoantigen (408 aa) was investigated. Because the minimum size of T-cell epitopes is 9–12 amino acids, no pattern of four or five residues could be sufficient for recognition. Therefore, when a pattern of four amino acids was identified, six residues were added on either side. Similarly, when a five amino acid pattern was found, five residues were added on either side. Thirty-seven epitopes were selected according to the reported motifs (Table 1).

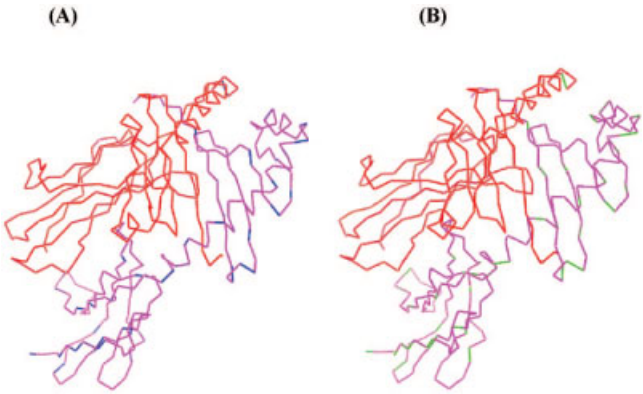


**Figure 2.** Superimposition of DQ7 onto DQ8 molecule (A) and DQ2 onto DQ8 molecule (B); $\alpha$- and $\beta$-chains of DQ8 are shown in black; $\alpha$- and $\beta$-chains of DQ7 and DQ2 are shown in grey. Structural differences derived from superimposition of DQ8/DQ7 and DQ8/DQ2 are indicated (figure was prepared with the Swiss-PdbViewer program[44]).

### *T-Cell Epitope Prediction Using TEPITOPE Software*

The TEPITOPE program uses HLA/peptide-binding data and structural data derived from the HLA/peptide complexes to predict potential candidate T-cell epitopes *in silico*. It assumes that the topography of a pocket is independent of neighbouring pockets, thus allowing the binding of similar residues to different HLA-II alleles. Extensive experimental work allowed the definition of the pocket profiles for 10 different polymorphic pockets that were combined to build >50 HLA-DR virtual matrices. The TEPITOPE software utilizes virtual matrix data in a linear prediction model, in which peptide conformation is entirely neglected.

The La/SSB sequence (408 aa) was loaded into the TEPITOPE sequence editor to predict promiscuous candidate T-cell epitopes. Fourteen HLA-DR alleles with the greatest sequence similarities to the corresponding HLA-DQ alleles were selected, and the TEPITOPE prediction threshold was set at 4%. This threshold predicts the 4% of the best-scoring peptides in La/SSB.

In a high stringent threshold (1%) less peptides are predicted for the HLA-DR alleles, while in a less stringent threshold as 4% more peptides are predicted.[10,15,36,50] Usually, it is suggested that the stringency level ranges from 1 to 4% for the first scan. Peptides predicted to bind to the HLA-DR alleles are shown in Table 2.

### *Quantitative Analysis of the Predicted Epitopes*

The quantitative analysis of selected sequences (any length >9 amino acids) indicates at which threshold levels the sequence would be predicted for each of the alleles included in the algorithm and provides the user with a better understanding of the promiscuity of the candidate epitope, which normally is not fully evident when performing the prediction with a single selected threshold. A quantitative analysis for all of the predicted candidate epitopes, given in Table 2, with a stringency "threshold" setting at 4% for all the HLA-II alleles was performed for La/SSB,[10,15,51] and the predicted epitopes with the best promiscuity are shown in Table 3.

In Figure 3 are shown, as an example, the threshold levels in the form of bar histograms for the epitope IFVVFDSIESA (154–164) and PLEIMIKFNRLNRLTTD (48–64) presented in Table 3.

### *T-Cell Epitope Prediction Using MULTIPRED Computational System*

MULTIPRED is a Web-based computational system for the prediction of peptides that bind multiple HLA alleles.[37] It uses hidden Markov models (HMMs) and artificial neural networks (ANNs) as predictive engines. A novel data representation method enables MULTIPRED to predict candidate T-cell epitopes that promiscuously bind multiple HLA class I and class II HLA-DR alleles.

**Table 1.** T-Cell Epitope Prediction of the La/SSB Autoantigen (Taylor, 1988).

| Pattern | Potential candidate T-cell epitopes of La/SSB | Region of La/SSB |
|---|---|---|
| | FNLPRD**KFLK**EQIKLD | (28–43) |
| | LSKSKA**ELMEI**SEDKT | (72–87) |
| | MRRTLH**KAFK**GSIFVV | (142–157) |
| **C—H—H—C** | SIESAK**KFVE**TPGQKY | (160–175) |
| | GEIKWI**DFVR**GAKEGI | (257–272) |
| | NKEVTW**ELVE**GEVEKE | (298–313) |
| | EGEVEK**EALKK**IIEDQ | (307–322) |
| | EKEAL**KKIIE**DQQESL | (311–326) |
| **G—H—H—C** | DENGATG**PVK**RAREET | (377–389) |
| **G—H—H—P** | RSVYIK**GFPT**DATLDD | (111–126) |
| **P—H—H—P** | WLEDKG**QVLN**IQMRRT | (130–145) |
| | DFNVIV**EALS**KSKAEL | (64–79) |
| **C—H—H—P** | GKGKGN**KAAQ**PGSGKG | (338–353) |
| | FQGKK**TKFAS**DDEHDE | (357–372) |
| | EETDKE**EPASK**QQKTE | (387–402) |
| **C—H—H—H—C** | GWVPLE**IMIK**FNRLN | (45–59) |
| | RRSPS**KPLPE**VTDEY | (89–104) |
| | YYFGDF**NLPR**DKFLKE | (23–38) |
| **P—H—H—C** | ILFKDD**YFAK**KNEERK | (182–197) |
| | LEEKIG**CLLKF**SGDLD | (226–241) |
| | PLEIMIK**FNRL**NRLTT | (48–63) |
| **C—H—P—C** | IMIKFN**RLNRL**TTDFN | (51–66) |
| | SKAELM**EISE**DKTKIR | (75–90) |
| | PSKPLP**EVTD**EYKNDV | (93–108) |
| **G—P—H—H—P** | WLEDKG**QVLN**IQMRR | (130–144) |
| **G—P—H—P** | AKDANN**GNLQ**LRNKEV | (286–301) |
| | MAALEA**KICH**QIEYYF | (10–25) |
| **C—H—P—P** | KFNRLN**RLTT**DFNVIV | (54–69) |
| | ALGKAK**DANN**GLNQLR | (282–297) |
| | LQLRNK**EVTW**ELVEGE | (294–309) |
| **C—H—P—H—P** | GSGKG**KVQFQ**GKKTK | (349–363) |
| | RDKFLK**EQIK**LDEGWV | (32–47) |
| | TLDDIK**EWLE**DKGQVL | (123–138) |
| **C—P—H—C** | SIFVVF**DSIE**SAKKFV | (153–168) |
| | VVFDS**IESAK**KFVETP | (156–171) |
| | EEDAEM**KSLEE**KIGCL | (218–233) |
| | SNHGEI**KWID**FVRGAK | (254–269) |

Residues in bold correspond to the indicated pattern.
**C:** charged residues, **H:** hydrophobic residues, **P:** polar residues and **G:** glycine.

HLA-DR (class II) proteins bind peptides through the binding groove, which is nine amino acids long. The training data comprise 2396 9-mer peptides (448 binders and 1948 nonbinders) related to eight HLA-DR alleles (DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*0801, DRB1*1101, DRB1*1301, DRB1*1501), which are included in the selected HLA-DR alleles for the TEPITOPE prediction, with the greatest sequence similarities to the corresponding HLA-DQ alleles except DRB1*1301. These data are mainly from the MHCPEP database,[52] published articles, and a set of HLA nonbinding peptides.

**Table 2.** T-Cell Epitope Prediction of the La/SSB Autoantigen (TEPITOPE) for the Selected Matrices in a Stringency Level at 4% Threshold.

| HLA-II allele | Predicted T-cell epitopes of La/SSB |
|---|---|
| DRB1*0101 | **f**nrlnrtt, **I**RRSPSKPL,-**Y**IKGFPTDA, **l**niqmrrtl **F**VVFDSIES, **L**HILFSNHG, **f**vrgakegi |
| DRB1*0102 | **M**AALEAKIC, **ikf**nrlnrtt, **I**RRSPSKPL lniqmrrtl, FVVFDSIES,-LHILFSNHG |
| DRB1*0301 | **I**KLDEG**W**VP**L**EIMI**K**FNR, **YK**NDVKNRS **L**HKAFKGSI**f**vvfdsiesa, **L**LILFKDDY IGCLLKFSG, **I**LFSNHGEI, **F**VRGAKEGI |
| DRB1*0401 | **YF**GDFNLPR, **W**VPLEIMIK, **l**nrlttdfnvivealsks **i**rrspskpl, **yk**ndvknrs, **f**ptdatldd, **i**qmrr**t**lhk **F**VVFDSIESA, **l**hilfsnhg, **f**qgkktkfa |
| DRB1*0402 | **imi**kfnrln, **i**vealsksk, **i**rrspskpl, **V**KNRSVYIK **I**QMRRTLHK, **F**VVFDSIES, **il**fkddyfa, **l**hilfsnhg **l**rnkevtwe, **l**nkwkskgr, FQGKKTKFA |
| DRB1*0701 | ikfnrlnrl, **I**RRSPSKPL,-**M**RRTLHKA**F**KGSI**F**VVFDSIES **f**vetpgqkyketdllil, **I**LFSNHGEI,-**F**VRGAKEGI |
| DRB1*0801 | **w**vpleimik**F**NRLNRLTT, **V**KNRSvyikgfptd iqmrrtlhk, **F**VVFDSIES,-**V**EAKLRAKQ,-**I**LFKEKAKE LQLRNKEVTWE, **WK**SKGRR**F**KGKGKGNK, **F**QGKKTKFA |
| DRB1*0802 | **w**vpleimik**f**nrlnrltt, **f**nviveals **y**kndvknrsvyik, **I**QMRRTLHK,-**F**VVFDSIES **V**EAK**L**RAKQEQEA, **i**lfkekake, **lq**l**r**nkevtwe **WK**SKGRRFKGKGKGNK, **F**QGKKTKFA |
| DRB1*1101 | **wvpl**e**imikf**nrlnrltt, **f**nviveals **I**QMRRTLHK, **F**VVFDSIES, **F**QGKKTKFA |
| DRB1*1104 | **W**VPleimik**f**nrln**r**ltt, **I**QMRR**TLH**K FVVFDSIES, **V**QFQGKKTK |
| DRB1*1305 | **W**VPLEIMIM**KF**NRLNRLTT, **f**nvi**v**eals iq**mr**rtl**hkaf**, **F**VVFDSIES |
| DRB1*1501 | **mikf**nrlnrL**TT**, **IR**RSPSKPL, iqmrrtlhk **F**VVFDSIES, **L**ILFKDDYF, **L**HIL**F**SNHG |
| DRB1*1502 | **mikfnrl**nr**LTT**, **I**RRSP S**K**PL, **i**qmrrtlhk **F**VVFDSIES, **LI**LFKDDYF, **L**HILFSNHG |
| DRB5*0101 | **w**vp**le**imikfnr, **i**vealsksk, **IRRSP**SKPL **iqm**rrtlhkaf, **F**VV**f**dsiesa**kk**fve**tp**gqky LHILFSNHG, hlfkekak, wkskgrrfk, **vqf**qgkktk |

The bold letter indicates the starting point of each PF (Peptide Frames) at position P1. Promiscuous candidate epitopes with a nonameric peptide frame and multiple overlapping frames are predicted. The sequences in lowercase indicate the presence of a residue in a PF with more than 10× inhibitory effect.

**Table 3.** Predicted T-Cell Epitopes of the La/SSB Autoantigen with the Best Promiscuity (TEPITOPE) after a Quantitative Analysis with a Stringency Threshold at 4%.

| HLA-DR alleles | Predicted epitopes |
|---|---|
| DRB1*0101, DRB1*0102 DRB1*0301, DRB1*0401 DRB1*0402, DRB1*0701 DRB1*1501, DRB1*1502 DRB5*0101 | DLHILFSNHGE (248–258) |
| DRB1*0101, DRB1*0102 DRB1*0301, DRB1*0401 DRB1*0402, DRB1*0701 DRB1*0801, DRB1*0802 DRB1*1101, DRB1*1104 DRB1*1305, DRB1*1501 DRB1*1502, DRB5*0101 | MIKFNRLNRLTTD (52–64) |
| DRB1*0101, DRB1*0102 DRB1*0301, DRB1*0401 DRB1*0402, DRB1*0701 DRB1*0801, DRB1*0802 DRB1*1101, DRB1*1104 DRB1*1305, DRB1*1501 DRB1*1502, DRB5*0101 | KFNRLNRLTTD (54–64) |
| DRB1*0301, DRB1*0401 DRB1*0801, DRB1*0802 DRB1*1101, DRB1*1104 DRB1*1305, DRB5*0101 | PLEIMIKFNRLNRLTTD (48–64) |
| DRB1*0401, DRB1*0402 DRB1*0802, DRB1*1101 DRB1*1305, DRB5*0101 | RLNRLTTDFNVIVEALSKSK (57–76) |
| DRB1*0301, DRB1*0801 DRB1*0802, DRB1*1101 DRB1*1104, DRB1*1305 DRB5*0101 | EIMIKFNRLNR (50–60) |
| DRB1*0101, DRB1*0102 DRB1*0301, DRB1*0401 DRB1*0402, DRB1*0701 DRB1*0801, DRB1*0802 DRB1*1101, DRB1*1104 DRB1*1305, DRB1*1501 DRB1*1502, DRB5*0101 | IFVVFDSIESA (154–164) |
| DRB1*0101, DRB1*0102 DRB1*0401, DRB1*0402 DRB1*0701, DRB1*0801 DRB1*0802, DRB1*1101 DRB1*1104, DRB1*1305 DRB1*1501, DRB1*1502 DRB5*0101 | NIQMRRTLHKA (139–149) |
| DRB1*0301, DRB1*0401 DRB1*0801, DRB1*0802 DRB1*1101, DRB1*1104 DRB1*1305, DRB5*0101 | QIKLDEGWVPLEIMIKFNRL (39–58) |
| DRB1*0301, DRB1*0701 | TLHKAFKGSIFVVFDSIESA (145–164) |

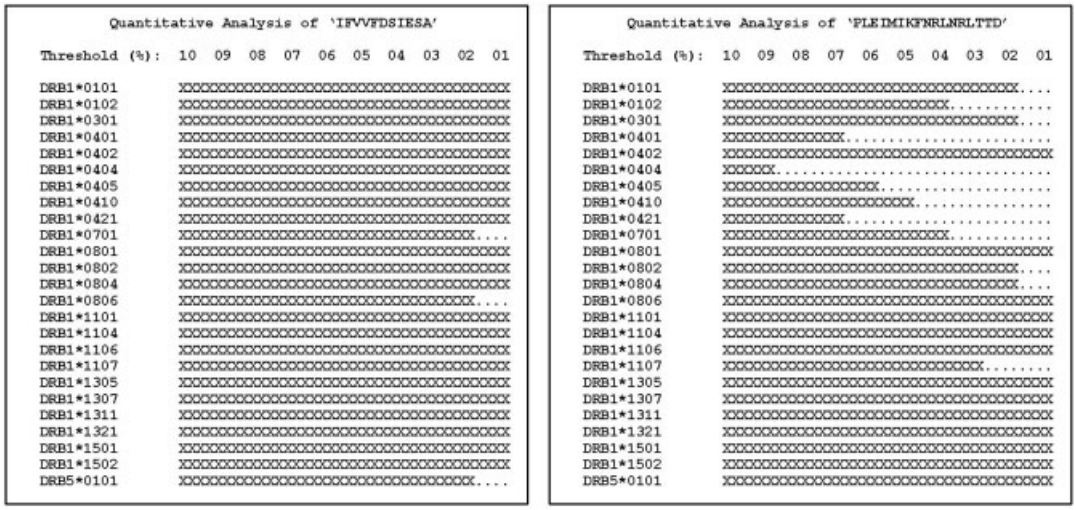


**Figure 3.** TEPITOPE'S quantitative analysis window. The "X" bars indicate the threshold levels at which the selected peptides were predicted for the HLA-II alleles.

To predict the candidate T-cell epitopes of La/SSB autoantigen (408 aa) the class II DR supertype and the ANN method were selected. The 9-mer binding scores range from 1 to 9, with scores 4–9 referring to predicted binders (8–9 high binders, 7–6 moderate binders, 5–4 low binders). Scores 1–3 refer to predicted nonbinders. The first 20 predicted 9-mer peptides of La/SSB autoantigen, with the highest sum, indicating the total sum of the individual binding scores, in descending order of the binding scores to multiple HLA class II alleles were selected. Subsequently, the hotspots at threshold 20 for average value were obtained and the predicted hotspot regions of La/SSB autoantigen sorted by average are given in Table 4.

### *Comparison of the Candidate Epitopes Predicted by Taylor's Sequence Pattern Approach with the TEPITOPE Software and the MULTIPRED Computational System*

T-cell epitopes, predicted by TEPITOPE and MULTIPRED, comprising common and/or similar patterns with Taylor's approach are indicated in Table 5. In this study, the sequence pattern method of Taylor, which is a binding motif approach, was used as starting point because it is simple to implement and easy to understand. Subsequently, TEPITOPE, a quantitative matrix method, which is an extension of binding motifs, was applied, and 20 candidate T-cell epitopes incorporating Taylor's motifs were identified. Finally, by applying MULTIPRED, based on HMMs and ANNs, which are adaptive, can self-improve, generalize well, and are effective with nonlinear problems, nine candidate T-cell epitopes comprising, common, and or similar motifs with Taylor and TEPITOPE approaches, were found. One may assume that common candidate T-cell epitopes deriving from the combined application of more than one prediction methods may proved more accurate and reliable.

### *Homology Modeling of the Potential Candidate Epitopes with the Insulin-B Peptide and Energy Minimization Calculations of the DQ2/DQ7 Epitope Complexes*

The insulin-B peptide bound to DQ8 revealed three major contact points with DQ8, glutamic acid, tyrosine, and glutamic acid side chains (E, Y, E), within the polymorphic pockets designated **P1, P4**, and **P9**. The same pockets were observed in the modeled DQ2 and DQ7.[31] T-cell epitopes predicted by Taylor's approach with common and/or similar patterns with TEPITOPE and MULTIPRED were superposed with the insulin-B peptide. A variety of combinations were performed to find the best superposed epitopes to the insulin-B peptide: *LVEALYLVCGERGG*. As an example, the SIFVVFDSIESAKKFV (153–168) candidate epitope of La/SSB,

**Table 4.** HLA Class II DR Predicted Hotspot Regions of the La/SSB Autoantigen (MULTIPRED) Sorted by Average.

| Position | Sequences | Average | Length |
|---|---|---|---|
| 45–67 | GWVPLEIMIKFNRLNRLTTDFNV | 26.75 | 23 |
| 134–169 | KGQVLNIQMRRTLHKAFKGSIFVVFDSIESAKKFVE | 23.86 | 36 |
| 18–37 | CHQIEYYFGDFNLPRDKFLK | 21.39 | 20 |
| 178–194 | TDLLILFKDDYFAKKNE | 20.73 | 17 |
| 102–127 | DEYKNDVKNRSVYIKGFPTDATLDDI | 20.43 | 26 |

**Table 5.** Comparison of the Candidate T-Cell Epitopes Identified by the Taylor's Sequence Pattern Approach, the TEPITOPE Program, and the MULTIPRED Computational System.

| Taylor's approach | Tepitope | Multipred |
|---|---|---|
| MRRTLH**KAFK**GSIFVV (142–157) | TLHKA**FK**GSIFVVFDSIESA (145–164) | KGQVLNIQMRRTLH**KAFK**GSIFVVF**DSIESAKKFVE** (134–169) |
| DFNVIVE**ALSK**SKAEL (64–79) | RLNRLTTDFNVIV**EALSKS**K (57–76) | — |
| GWVPL**EIMIK**FNRLN (45–59) | **PLEIMIK**FNRLNRLTTD (48–64) | GWVPL**EIMIKFNRLNRLTT**DFNV (45–67) |
| PLEIM**IKFNR**LNRLTT (48–63) | QIKLDEGWVPLEIM**IKFNRL** (39–58) | GWVPL**EIMIKFNRLNRLTT**DFNV (45–67) |
| RRSPS**KPLPE**VTDEY (89–104) | KIRRSPS**KPLPE** (87–97) | — |
| | **KFNR**LNRLTTD (54–64) | |
| PLEIMI**KFNR**LNRLTT (48–63) | **MIKFNR**LNRLTTD (52–64) | GWVPL**EIMIKFNRLNRLTT**DFNV (45–67) |
| | PLEIMI**KFNR**LNRLTTD (48–64) | |
| | QIKLDEGWVPLEIM**IKFNRL** (39–58) | |
| | **KFNR**LNRLTTD (54–64) | |
| | MI**KFNR**LNRLTTD (52–64) | GWVPL**EIMIKFNRLNRLTT**DFNV (45–67) |
| IMIKFNRLNRLTTDFN (51–66) | PLE**IMIKFNRL**NRLTTD (48–64) | |
| | QIKLDEGWVPLEIMI**KFNRL** (39–58) | |
| | **KFNR**LNRLTTD (54–64) | |
| KFNRLNRLTTDFNVIV (54–69) | MI**KFNR**LNRLTTD (52–64) | GWVPLEIMIKFNRLNRLTTDFNV (45–67) |
| | PLEIMI**KF**NRLNRLTTD (48–64) | |
| SIFVVF**DSIE**SAKKFV (153–168) | IFVVF**DSIE**SA (154–164) | KG**QVLN**IQMRRTLH**KAFK**GSIFVVFDS**IESAKKFVE** (134–169) |
| VVFDSI**ESAK**KFVETP (156–171) | TLHKAFKGSIFVVF**DSIE**SA (145–164) | **KGQVLNI**QMRRTLH**KAFK**GSIFVV**FDSIESAKKFVE** (134–169) |
| SNHGEI**KWID**FVRGAK (254–269) | DLHILFSNHGEI**KWIDF** (248–265) | — |

Common and/or similar sequence patterns are shown in bold.

obtained by the three prediction approaches, was superposed with the insulin-B peptide. The Swiss-Pdb Viewer program was used for the alignment of the two sequences and the eight possible candidate epitopes (Fig. 4), derived from the superimposition of the SIFVVFDSIESAKKFV onto the insulin-B peptide LVEALYLVCGERGG, were subjected to energy minimization calculations with the modeled DQ7. It was found that the modeled HLA-DQ7/epitope (154–167) complex did not experience any significant structural differences. Subsequently, the Procheck suite of programs was utilized for assessing the stereochemical quality of the DQ7/epitope (154–167) complex, and finally, the compatibility of the complex was confirmed using the "O" program (Fig. 5). The described approach of the step by step alignment and the energy minimization calculations were performed with all the common candidate epitopes, and the final proposed epitopes of the La/SSB autoantigen for the HLA-DQ2/DQ7 are given in Table 6.

By applying homology modeling and energy minimization experiments of the selected epitopes, after comparing three different computational methods, the best superposed epitopes to the insulin-B peptide with the lowest structural differences to the reference peptide were identified (Table 6). It is concluded that these peptides possess the highest probability to be binders of the HLA-DQ2/DQ7 molecules.

### *SASA of the Proposed Epitopes within the HLA -DQ2 and HLA-DQ7 Molecules*

GETAREA, a Web service, is one of the most efficient programs for calculating the SASA, which allowed us to submit the Cartesian coordinates of atoms of the proposed candidate epitopes, complexed with DQ2 and DQ7, stored in the PDB format, to obtain the SASA of the epitopes[24] within the grooves.

The residues of the proposed epitopes (Table 6) corresponding to the P1, P4, and P9 pockets of the antigen-binding groove of the modeled DQ2 and DQ7 should have very low solvent accessibility compared to the neighboring and in particular to the terminal residues, which are extended out of the ends of the peptide-binding groove. In Figure 6 is depicted the total surface area for the La/SSB epitope 311–324 in its complexed form with HLA-DQ2 and for epitope 149–161 complexed with HLA-DQ7.

The SASA experiments confirm that the residues corresponding to the P1, P4, and P9 pockets (Table 6) of the binding grooves exhibit low solvent accessibility because they are less exposed to the microenvironment.

### *Potential Interaction Energy, Binding Affinity, and $IC_{50}$ Values of the Proposed Epitopes*

TINKER software[34] was used for the calculation of the potential interaction energy. The potential interaction energy, $\Delta E_{int}$, of the DQ2/DQ7–epitope complexes, was calculated from the equation:

$$\Delta E_{int} = E_{pMHC} - E_{MHC} - E_p$$

where $\Delta E_{pMHC}$ is the total potential energy of the MHC–epitope complex in the bound state, $E_{MHC}$ and $E_p$ are the potential energies of the MHC molecule and the predicted epitope in the unbound state, respectively. In Table 7 are listed the proposed MHC–epitope complexes starting from the most preferred to the less favorable.

Furthermore, MHCPred was applied for the quantitative prediction of the peptide–HLA-DQ2/DQ7 binding affinity. There is increasing evidence suggesting that as the MHC binding affinity of


**Figure 4.** Superimposition of the candidate epitope SIFVVFDS-IESAKKFV (153–168) to the insulin-B peptide LVEALYL-VCGERGG. The same residues are indicated with the symbol (*), while residues with similar characteristics with the symbols (..) and (.). Amino acid residues corresponding to the P1, P4, and P9 pockets are shown in light gray. Note that the length of the candidate epitopes ranges from 12 to 14 amino acid residues.

a peptide rises, the greater the probability to be recognized by a TCR of the T-cell repertoire so that the immune response is better initiated. One of the ways to measure binding affinity is to calculate the $IC_{50}$ value, which indicates the concentration required for 50% inhibition in a competitive binding assay. A peptide binding to an MHC may be recognized, by a TCR, if it binds with a $pIC_{50} > 6.3$ ($pIC_{50} = -\log IC_{50}$) or $IC_{50}$ values between 0.01 to 5000 nM.

In Table 8 are quoted the proposed epitopes according to their $IC_{50}$ values, which are inversely proportional to their binding affinity. Epitopes for HLA-DQ7 are all good binders except the epitope 78–90, while in the case of HLA-DQ2, epitopes spanning the sequences 131–144 and 161–174 are less effective.

## Conclusions

In this study, the candidate T-cell epitopes of the La/SSB autoantigen, main target of the autoimmune response in patients with SS and SLE, were predicted using as templates the HLA-DQ2 and DQ7 molecules, which are genetically linked to patients with autoimmune rheumatic diseases, and in particular, with Sjogren's Syndrome (SS) and Systemic Lupus Erythematosus (SLE). Modeling of DQ2 and DQ7 was based on the crystal structure of DQ8, an HLA molecule of high-risk factor of type I diabetes, which is also an autoimmune disease. The quality of the modeled HLA-DQ2/DQ7 is similar to the quality expected for structures determined at 2 Å. Almost all residues of the modeled DQ2 and DQ7 were found in the preferred regions of the Ramachandran plot. Superimposition of modeled DQ2 and DQ7 with DQ8 using TINKER molecular modeling software gave satisfactory RMSD values (0.034 for DQ8/DQ2 and 0.033 for DQ8/DQ7), confirming the compatibility and reliability of the modeled HLA molecules.

Three different approaches were applied and compared in our study to identify T-cell epitope candidates of the La/SSB autoantigen: Taylor's sequence pattern method, TEPITOPE quantitative matrices, and MULTIPRED artificial neural networks. The obtained common and/or similar candidate T-cell epitopes were subjected to homology modeling with the insulin-B peptide, and the best superposed epitopes were placed into the modeled HLA-DQ2 and DQ7 binding grooves to perform energy minimization

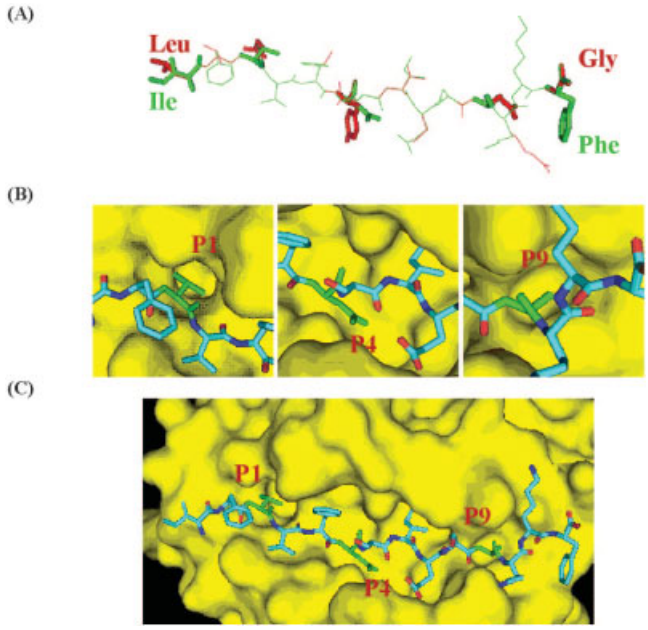


**Figure 5.** The best aligned epitope IFVVFDSIESAKKF (154–167) (shown in green) obtained after superimposition with the insulin-B peptide LVEALYLVCGERG (shown in red). The first and the last amino acid residues of the candidate epitope (Ile and Phe) and the insulin-B peptide (Leu and Gly) are indicated as sticks (A). The contact residues, V156, D159, and A164, of the candidate epitope (154–167) (shown in green) placed into the binding pockets P1, P4, and P9 of the modelled HLA-DQ7 (B). Molecular surface representation of the modeled HLA-DQ7/epitope (154–167) complex. The surface of the binding site of the modeled HLA-DQ7 molecule is shown in yellow, and the contact residues in pockets P1, P4, and P9 are shown in green (C). (Figure prepared with PyMol.[53])

**Table 6.** The Proposed T-Cell Epitopes of the La/SSB Autoantigen for the HLA-DQ7 and HLA-DQ2 Molecules.

| Sequence | Region of La/SSB |
|---|---|
| **DQ7 molecule** | |
| IV**EALS**KSKA**E**L | 68–79 |
| EL**M**EI**S**EDKT**K**IR | 78–90 |
| AF**K**GS**I**FVVF**D**SI | 149–161 |
| GS**I**FV**VF**DSI**E**SAK | 152–165 |
| IF**VV**F**D**SIES**A**KKF | 154–167 |
| AKDAN**N**GNLQ**L**RNK | 286–299 |
| **DQ2 molecule** | |
| LK**E**QI**K**LDEG**W**V | 36–47 |
| IV**EALSK**SKA**E**L | 68–79 |
| SK**A**EL**M**EISE**D**KT | 75–87 |
| LE**D**KG**QVL**NI**Q**MRR | 131–144 |
| AF**K**GS**I**FVVF**D**SIE | 149–162 |
| I**E**SA**KK**FVE**T**PGQK | 161–174 |
| **AKD**ANNGNLQ**L**R | 286–297 |
| **AK**DANNGNLQ**L**RNK | 286–299 |
| E**KE**AL**KK**IIE**D**QES | 311–324 |

The amino acid residues (in bold) correspond to the P1, P4, and P9 pockets of the antigen binding groove.

**Table 7.** Potential Interaction Energy ($\Delta E_{int}$) of the MHC–Epitope Complexes (DQ8-p, DQ7-p, and DQ2-p), in kcal/mol.

| Complex | $\Delta E_{int}$ |
|---|---|
| **DQ8** | |
| DQ8-p | −74.778 |
| **DQ7** | |
| DQ7-$p_{(149–161)}$ | −122.63 |
| DQ7-$p_{(78–90)}$ | −72.785 |
| DQ7-$p_{(286–299)}$ | −69.358 |
| DQ7-$p_{(154–167)}$ | −65.879 |
| DQ7-$p_{(68–79)}$ | −51.545 |
| DQ7-$p_{(152–165)}$ | −28.388 |
| **DQ2** | |
| DQ2-$p_{(286–297)}$ | −88.272 |
| DQ2-$p_{(149–162)}$ | −86.230 |
| DQ2-$p_{(131–144)}$ | −76.134 |
| DQ2-$p_{(286–299)}$ | −74.111 |
| DQ2-$p_{(161–174)}$ | −71.061 |
| DQ2-$p_{(311–324)}$ | −65.917 |
| DQ2-$p_{(68–79)}$ | −49.201 |
| DQ2-$p_{(36–47)}$ | −47.912 |
| DQ2-$p_{(75–87)}$ | −47.781 |

calculations and epitopes with the lowest structural differences were obtained. Six epitopes were attributed to the HLA-DQ7 and nine to HLA-DQ2. It was also found that residues corresponding to the P1, P4, and P9 pockets of the HLA-DQ2 and DQ7 binding grooves experience very low solvent accessibility in contrast with the neighboring and the terminal residues because they are less exposed to the microenvironment of the groove.

The proposed T-cell epitopes were classified according to their binding efficiency. It is concluded that five epitopes, out of six, for HLA-DQ7 and seven epitopes, out of nine, for HLA-DQ2 are good

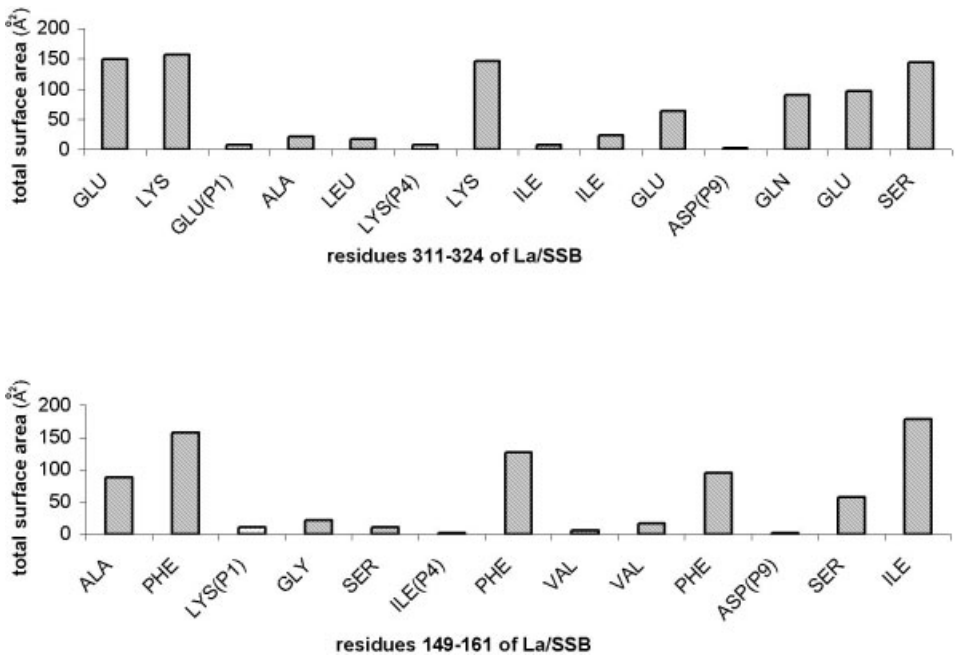


**Figure 6.** The total surface areas for the La/SSB epitopes 311–324 and 149–161 complexed with HLA-DQ2 and HLA-DQ7, respectively.

**Table 8.** Binding Affinity Prediction of the Proposed Epitopes for HLA-DQ2 and HLA-DQ7 (MHCPred).

| Epitope sequence | 9-mer peptides | Predicted $-\log IC_{50}$ (M) | Predicted $IC_{50}$ Value (nM) |
|---|---|---|---|
| **DQ8** | **DQ8** | | |
| 35–47 | $\mathbf{E}_{37}$ A L **Y** L V C G $\mathbf{E}_{45}$ | 6.988 | 102.80(0.29) |
| **DQ7** | **DQ7** | | |
| 152–165 | $\mathbf{I}_{154}$ F V **V** F D S I $\mathbf{E}_{162}$ | 7.714 | 19.32(0.33) |
| 68–79 | $\mathbf{E}_{70}$ A L S K **S** K A $\mathbf{E}_{78}$ | 7.348 | 44.87(0.29) |
| 286–299 | $\mathbf{D}_{288}$ A N **N** G N L Q $\mathbf{L}_{296}$ | 6.816 | 152.76(0.33) |
| 149–161 | $\mathbf{K}_{151}$ G S **I** F V V F $\mathbf{D}_{159}$ | 6.685 | 206.54(0.38) |
| 154–167 | $\mathbf{V}_{156}$ V F **D** S I E S $\mathbf{A}_{164}$ | 6.455 | 350.75(0.38) |
| 78–90 | $\mathbf{M}_{80}$ E I **S** E D K T $\mathbf{K}_{88}$ | 6.004 | 990.83(0.29) |
| **DQ2** | **DQ2** | | |
| 68–79 | $\mathbf{E}_{70}$ A L **S** K S K A $\mathbf{E}_{78}$ | 7.348 | 44.87(0.29) |
| 311–324 | $\mathbf{E}_{313}$ A L **K** K I I E $\mathbf{D}_{321}$ | 7.17 | 67.61(0.25) |
| 75–87 | $\mathbf{A}_{77}$ E L **M** E I S E $\mathbf{D}_{85}$ | 7.151 | 70.63(0.25) |
| 288–297 | $\mathbf{D}_{288}$ A N **N** G N L Q $\mathbf{L}_{296}$ | 6.816 | 152.76(0.33) |
| 288–299 | $\mathbf{D}_{288}$ A N **N** G N L Q $\mathbf{L}_{296}$ | 6.816 | 152.76(0.33) |
| 149–162 | $\mathbf{K}_{151}$ G S **I** F V V F $\mathbf{D}_{159}$ | 6.685 | 206.54(0.38) |
| 36–47 | $\mathbf{E}_{38}$ Q I **K** L D **E** G $\mathbf{W}_{46}$ | 6.348 | 448.75(0.21) |
| 131–144 | $\mathbf{D}_{133}$ K G **Q** V L N I $\mathbf{Q}_{141}$ | 6.196 | 636.80(0.38) |
| 161–174 | $\mathbf{E}_{162}$ S A **K** K F V E $\mathbf{T}_{170}$ | 6.137 | 729.46(0.29) |

The first column shows the epitope sequence, the second column shows the 9-mer core of the-epitope sequence, the third column shows the predicted-$\log IC_{50}$(M) ($pIC_{50}$), and the forth column shows the $IC_{50}$ values (nM). The confidence of the prediction (max = 1) is shown in parantheses. P1, P4, and P9 residues are in bold. For comparison, values of the insulin-B peptide for DQ8 are also indicated.

binders of MHC, and they may be sufficiently recognized by the immune system.

Our study constitutes the ground work for a rapid and reliable experimentation concerning the T-cell epitope mapping of the La/SSB autoantigen, and could lead to the development of immunotherapeutic drug-like molecules as inhibitors of the T-cell inappropriate immune response in autoimmune diseases.

## References


1. Jensen, P. E. Biopolymers 1997, 43, 303.
2. Madden, D. R., et al. Cell 1993, 75, 693.
3. Stern, L. J.; Brown, J. H.; Jardetzky, T. S.; Gorga, J. C.; Urban, R. G.; Strominger, J. L.; Wiley, D. C. Nature 1994, 368, 215.
4. De Groot, A. S.; Sbai, H.; Aubin, C. S.; McMurray, J.; Martin, W. Immunol Cell Biol 2002, 80, 255.
5. Robinson, J.; Malik, A.; Parham, P.; Bodmer, J. G.; Marsh, S. G. Tissue Antigens 2000, 55, 280.
6. Bjorkman, P. J.; Saper, M. A.; Samroui, B.; Bennett, W. S.; Strominger, J. L.; Wiley, D. C. Nature 1987, 329, 506.
7. Fremont, D. H.; Matsumura, M.; Strura, E. A.; Peterson, P. A.; Wilson, I. A. Science 1992, 257, 919.
8. Brown, J. H.; Jardetzky, T. S.; Gorga, J. C., et al. Nature 1993, 364, 33.
9. Brusic, V.; Bajic, V. B.; Petrovsky, N. Methods 2004, 34, 436.
10. Bian, H.; Reidhaar–Olson, J. F.; Hammer, J. Methods 2003, 29, 299.
11. Rothbard, J. B.; Taylor, W. R. EMBO J 1988, 7, 93.
12. Taylor, W. R. J Mol Biol 1986, 188, 233.
13. Celis, E.; Larson, J.; Otvos, L., Jr.; Wunner, W. H. J Immunol 1990, 145, 305.
14. Brusic, V.; Rudy, G.; Honeyman, M.; Hammer, J.; Harrison, L. Bioinformatics 1998, 14, 121.
15. Sturniolo, T.; Bono, E.; Ding, J.; Raddrizzani, L.; Tuereci, O.; Sahin, M.; Braxenthaler, M.; Gallazzi, F.; Protti, M. P.; Sinigaglia, F.; Hammer, J. Nat Biotechnol 1999, 17, 555.
16. Brusic, V.; Flower, D. R. Biosilico 2004, 2, 18.
17. Honeyman, M. C.; Brusic, V.; Stone, N. L.; Harrison, L. C. Nat Biotechnol 1998, 16, 966.
18. Altman, D. M.; Sansom, D.; Marsh, S. G. E. Immunol Today 1991, 12, 267.
19. Ollerup, J.; Hillert, J.; Frederickson, J., et al. Proc Natl Acad Sci USA 1989, 86, 7113.
20. Bergseng, E.; Xia, J.; Kim, C-Y.; Khosla, C.; Sollid, L. M. J Biol Chem 2005, 280, 21791.
21. Sollid, L. M.; Marcussen, G. E. K., et al. J Exp Med 1989, 169, 345.
22. Harley, J. B.; Reichlin, M.; Arnett, F. C.; Alexander, E. L.; Bias, W. B.; Provost, T. T. Science 1986, 232, 1145.
23. Scofield, R. H.; Frank, M. B.; Neas, B. R.; Horowitz, R. M.; Handgrave, K. L.; Fugisake, A.; McArthreur, R.; Harley, J. B. Clin Immunol Immunopathol 1994, 72, 335.
24. Todd, J. A.; Acha–Orbea, H.; Bell, J. I. Science 1988, 240, 1003.
25. Wong, F. S.; Moustakas, A. K.; Wen, L.; Papadopoulos, G. K.; Janeway, C. A., Jr. Proc Natl Acad Sci USA 2002, 99, 5551.
26. Martinez, N. R.; Augstein, P.; Moustakas, A. K.; Papadopoulos, G. K.; Gregori, S.; Adorini, L.; Jackson, D. C.; Harrison, L. C. J Clin Invest 2003, 111, 1365.
27. Brown, J. H.; Jardetzky, T.; Saper, M. A.; Samraoui, B.; Bjorkam, P. J.; Wiley, D. C. Nature 1988, 332, 845.
28. Paliakakis, K.; Routsias, J.; Petratos, K.; Ouzounis, C.; Kokkinidis, M.; Papadopoulos, G. K. J Struct Biol 1996, 117, 145.
29. Routsias, J. G.; Kosmopoulou, A.; Makri, A.; Panou–Pomonis, E.;

Sakarellos, C.; Sakarellos–Daitsiotis, M.; Moutsopoulos, H. M.; Tzioufas, A. G. J Med Chem 2004, 47, 4327.
30. Terzoglou, A. G.; Routsias, J. C.; Sakarellos, C.; Sakarellos–Daitsiotis, M.; Moutsopoulos, H. M.; Tzioufas, A. G. J Autoimmun 2003, 21, 47.
31. Lee, K. H.; Wucherpfennig, K. W.; Wiley, D. C. Nat Immunol 2001, 2, 501.
32. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. J Appl Crystallogr 1993, 26, 283.
33. Jones, T. A.; Bergdoll, M.; Kjeldgaard, M. Crystallography and Modelling Methods in Molecular Design; Springer Verlag: Berlin, 1990.
34. Ponder, J. W. TINKER: Software Tools for Molecular Design, 4.1; Washington University School of Medicine: St. Louis, MO, 2003.
35. Kwok, W. W.; Gebe, J. A.; Liu, A.; Agar, S.; Ptacek, N.; Hammer, J.; Koelle, D. M.; Nepom, G. T. Trends Immunol 2001, 22, 583.
36. Bian, H.; Hammer, J. Methods 2004, 34, 468.
37. Zhang, G. L.; Kham, A. M.; Srinivasan, K. N.; August, T.; Brusic, V. Nucleic Acids Res 2005, 33, W173.
38. Fraczkiewicz, R.; Braun, W. J Comp Chem 1988, 19, 319.
39. Guan, P.; Doytchinova, I. A.; Zygouri, C.; Flower, D. R. Appl Bioinformatics 2003, 2, 63.
40. Guan, P.; Doytchinova, I. A.; Zygouri, C.; Flower, D. R. Nucleic Acids Res 2003, 31, 3621.
41. Hattotuwagama, C. K.; Guan, P.; Doytchinova, I. A.; Zygouri, C.; Flower, D. R. J Mol Graph Model 2004, 22, 195.
42. Altschul, S. F.; Madden, T. L.; Scaffer, A. A.; Zhang, Z.; Miller, W.; Lipman, D. J. Nucleic Acids Res 1997, 25, 3389.
43. Westbrook, J.; Feng, Z.; Chen, L.; Yang, H.; Berman, H. M. Nucleic Acids Res 2003, 31, 489.
44. Guex, N.; Peitch, M. C. Electrophoresis 1997, 18, 2714.
45. Schwede, T.; Kopp, J.; Guex, N.; Peitch, M. C. Nucleic Acids Res 2003, 31, 3381.
46. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comp Chem 1983, 4, 187.
47. MacKerell, A. D.; Brooks, B., Jr.; Brooks, C. L.; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. Encyclopedia Comp Chem 1998, 1, 271.
48. Still, W. C.; Tempczyk, A.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
49. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. J Phys Chem A 1997, 101, 3005.
50. Westhead, D. R.; Thornton, J. M. Curr Opin Biotechnol 1998, 9, 383.
51. Hammer, J.; Bono, E.; Gallazzi, F.; Belunis, C.; Nagy, Z.; Sinigaglia, F. J Exp Med 1994, 180, 2353.
52. Brusic, V.; Rudy, G.; Harrison, L. C. Nucleic Acids Res 1998, 26, 368.
53. DeLano, W. L. The PyMol Molecular Graphics System, 2002 (on World Wide Web http://www.pymol.org).