Quantitative Structure−Activity Relationship of Peptides Binding to the Class II Major Histocompatibility Complex Molecule A^q Associated with Autoimmune Arthritis

Lotta Holm,[†] Kristina Frech,[†] Balik Dzhambazov,[‡] Rikard Holmdahl,[‡] Jan Kihlberg,[†,§] and Anna Linusson*[,†]

Department of Chemistry, Umeå University, SE-901 87 Umeå, Sweden, Section of Medical Inflammation Research, Lund University, Sölvegatan 19, I11 BMC, SE-221 84 Lund, Sweden, and AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden

Received October 17, 2006

Presentation of (glyco)peptides by the class II major histocompatibility complex molecule A^q to T cells plays a central role in collagen-induced arthritis, an animal model for the autoimmune disease rheumatoid arthritis. A peptide library was designed using statistical molecular design in amino acid space in which five positions in the minimal mouse collagen type II binding epitope CII260−267 were varied. A substantially reduced peptide library of 24 peptides with diverse and representative molecular characteristics was selected, synthesized, and evaluated for the binding strength to A^q. A multivariate QSAR model was established by correlating calculated descriptors, compressed to its principle properties, with the binding data using partial least-square regression. The model was successfully validated by an external test set. Interpretation of the model provided a molecular property binding motif for peptides interacting with A^q. The information may be useful in future research directed toward new treatments of rheumatoid arthritis.

Introduction

The immune system's ability to distinguish self from nonself is a crucial feature of the body's defenses against foreign antigens. This ability is linked to a ternary complex in which T cells play a key role in the activation or nonactivation of the immune system. Foreign and autoantigens are degraded into peptides by antigen presenting cells and subsequently presented by class II major histocompatibility complex (MHC[a]) molecules for recognition by T-cell receptors. This recognition constitutes an important step in a series of events that forms the immune response to foreign antigens. The self-peptide MHC complexes should not lead to any response, and when this tolerance is disrupted, an immune response toward endogenous tissue(s) is elicited and an autoimmune inflammatory disease, such as rheumatoid arthritis (RA), may develop.

RA is one of the most common autoimmune inflammatory diseases and is characterized by chronic inflammation of peripheral cartilaginous joints. The symptoms include swelling, stiffness, and pain in the joints and subsequent erosion of underlying bones, which might further lead to deformity and malfunction of the joints.[1] A large number of therapeutic agents against RA are commercially available, but there is no cure and most treatments do not inhibit the progression of the disease satisfactorily.[2] The major difficulties hindering the development of effective treatments for RA seem to be the complexity of the system, lack of knowledge of the mechanisms involved, and the need for reliable disease models.[3,4] RA has been linked to the MHC class II molecules DR1 and DR4,[5−7] and the activated immune systems of severely affected RA patients include both autoreactive T cells[8−10] and antibodies[11−13] directed against type II collagen (CII), which is the most abundant protein in cartilage.

Injection of CII in rats[14] or mice[15] provokes the development of collagen-induced arthritis (CIA) with symptoms and histopathology similar to those of RA. CIA is the most commonly used animal model for RA, and susceptibility to murine CIA is linked to the mouse MHC class II molecule A^q.[16,17] By use of synthetic peptides, the minimal CII peptide epitope required for both binding to the A^q molecule and inducing a T-cell response has been determined to be the octapeptide ranging from amino acid 260 to amino acid 267 (CII260−267, Figure 1).[18] The isoleucine at position 260 and phenylalanine 263 have been found to be essential anchor residues for binding to the A^q molecule, according to an alanine scan.[19,20] Furthermore, it has been shown that the T-cell response is often linked to specific recognition of a carbohydrate moiety resulting from post-translational modification of lysine at CII264 (β-D-galactopyranosyl modified hydroxylysine, i.e., a GalHyl moiety).[21−24] In some cases, the T-cell recognition is not linked solely to the sugar moiety but also to the glutamic acid located at position CII266.[25] Highly interesting results have shown that vaccination of mice with the mouse and rat galactosylated CII256−273 peptides can provide protection against development of CIA.[26] In particular, vaccination by the rat glycopeptide complexed with the A^q molecule significantly retarded progression of the disease and reduced its severity in mice with ongoing chronic relapsing arthritis.[27] These findings indicate that analyses with variants of the CII glycopeptide could be very useful for exploring the tricomponent A^q/glycopeptide/T-cell receptor interactions and could facilitate the development of potential drug or vaccine candidates for treating RA.

While the carbohydrate specificity has been extensively studied, less is known about the interactions between the A^q molecule and the peptide ligand in the ternary complex. Besides the proposed anchoring positions (Ile260 and Phe263), the introduction of a methylene ether amide bond mimetic between amino acids Ile260 and Ala261 in the minimal T-cell epitope CII260−267 resulted in a substantial drop (20-fold) in peptide affinity for A^q.[25] Moreover, Jane-wit et al. have postulated a common autoimmune-motif (KXXS) for peptides binding to

* To whom correspondence should be addressed. Phone: +46-90-7866890. Fax: +46-90-138885. E-mail: anna.linusson@chem.umu.se.
† Umeå University.
‡ Lund University.
§ AstraZeneca R&D Mölndal.

[a] Abbreviations: MHC, major histocompatibility complex; RA, rheumatoid arthritis; CII, type II collagen; CIA, collagen-induced arthritis; GalHyl, β-D-galactopyranosyl modified hydroxylysine; SMD, statistical molecular design; QSAR, quantitative structure−activity relationship; PCA, principle component analysis; PLS, partial least-squares projections to latent structures; DA, discriminant analysis.

10.1021/jm061209b CCC: $37.00
Published on Web 04/11/2007

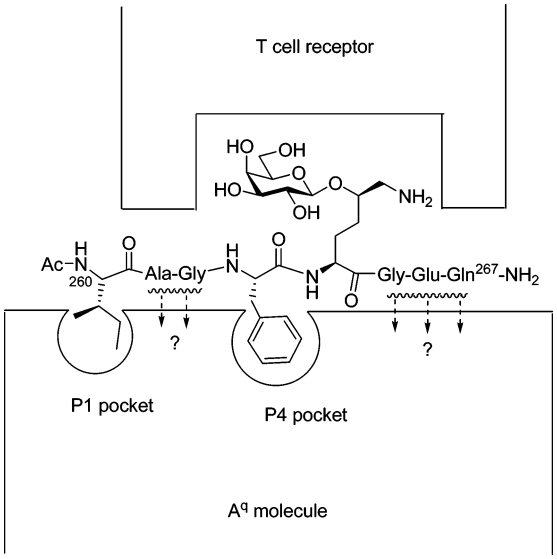


Figure 1. Schematic illustration of the ternary complex of the collagen glycopeptide CII260−267 presented by the mouse MHC class II molecule A^q to the T-cell receptor. The peptide is anchored to the A^q molecule with Ile260 in the P1 pocket and Phe263 in the P4 pocket. The arrows indicate the amino acid positions investigated in this study.

H-2q MHC class II molecule(s) based on comparisons of peptides responsible for eliciting experimental autoimmune myocarditis.[28]

Various complementary methods for designing and generating peptides to elucidate binding and/or subsequent T-cell responses are available, including scanning of single amino acids,[20,29] use of positional scanning combinatorial libraries,[30−32] phage-display libraries,[33,34] peptide isosteres,[25,35] and computational methods such as structure-based design[34,36] and statistical molecular design (SMD).[37,38] There has been great effort in predicting binding of short peptides to MHC molecules.[39,40] There are a wide range of motif-based prediction algorithms available online where SYFPEITHI[41] and RANKPEP[42,43] include models for mouse MHC class II molecules, but A^q is not included. In recent years, nonlinear pattern recognition methods like neural networks,[44] support vector machines,[45] and a kernel-based approach[46] have been reported for classification of binders versus nonbinders of MHC class II molecules. Only a few studies present quantitative structure−activity relationship (QSAR) models of peptides binding to MHC class II molecules with the objective of achieving quantitative predictions of binding strength and information about the influence of different amino acid positions. Doytchinova and Flower have presented an additive method for quantitative binding affinity prediction for peptides binding to molecule DRB1*0401.[47] The same methodology was also recently applied to six class II mouse alleles ($I\text{-}A^b$, $I\text{-}A^d$, $I\text{-}A^k$, $I\text{-}A^s$, $I\text{-}E^d$, and $I\text{-}E^k$).[48] In addition, a 3D-QSAR model for the human DR-4 molecule has been developed by Lin and co-workers.[49]

The advantage of using SMD for designing a set of peptides to investigate the importance and influence of different peptide positions by QSAR modeling is that it enables the effects of more than one molecular property at several positions to be investigated with a minimum number of peptides. It also yields information about potential interactive effects between properties at different positions and provides a sound basis for a subsequent QSAR modeling for prediction of binding.[50−52] The SMD approach is performed in building block space (i.e., amino acid space), where the molecular characteristics of the separate amino acids at specific positions used in the design can be directly correlated to the biological response (e.g., binding strength).[52] It has been shown that QSAR models based on properties of amino acids at each position (local models) were superior over models based on properties of whole peptides (global models) for peptides binding to the MHC class I HLA-A*0201.[53]

In the study reported here we designed a peptide library based on the minimal CII peptide binding to the mouse MHC class II molecule A^q, using the SMD approach. The design was performed in amino acid space, and peptides were selected using D-optimal design.[50] The chosen peptides were synthesized on solid phase, and their binding strength to A^q was tested in a cell-based competition assay. Multivariate methods were used to evaluate the binding data, and a QSAR model was subsequently established. An external peptide test set was used to further verify the binding model.

Results and Discussion

Peptide Scaffold. The minimal CII peptide epitope found to bind to the A^q molecule while retaining the ability to induce T-cell responses, i.e., the octapeptide CII260−267,[18] was decided to be the most suitable scaffold for this binding study (Figure 2). Although the even shorter peptide CII260−266 has been shown to bind (although very weakly),[18] it was believed that modifications were likely to cause severe losses of A^q binding.

Position 264 harbors the GalHyl moiety important for T-cell recognition, and it has been shown that the moiety has minimal influence on peptide binding to the A^q molecule.[18] This position was therefore kept constant and nonglycosylated in order to simplify the synthesis of the peptide library. However, it should be stressed that GalHyl plays an extremely important role in T-cell responses and thus warrants further study regarding autoimmune response. The known critical binding points, i.e., the anchor residues Ile260 and Phe263, were also left unchanged to prevent complete loss of binding among members of the peptide library. Consequently, five positions (Ala261, Gly262, Gly265, Glu266, Gln267) in CII260−267 were chosen to be systematically varied with SMD, referred to hereafter as positions 1−5 (Figure 2).

In a first attempt to synthesize the library of selected peptides considerable problems with aqueous solubility were experienced. To circumvent these problems, lysine was introduced at the N and C terminals of the peptide (**N-Lys** and **C-Lys**, respectively) and tested for two commonly used CII-specific T-cell hybridoma cell lines (HCQ.4 and HDB2, Table 1). The peptide with lysine at the C terminal was recognized by both hybridoma. Consequently, the peptide scaffold was elongated with a C-terminal lysine to improve the solubility of the peptide library.[54] We also decided to N^{α}-acetylate the resulting peptide scaffold and to introduce a C-terminal amide to prevent loss of affinity to A^q due to the presence of charged functional groups in the peptide backbone (Figure 2).

Building Block Selection Using Multivariate Characterization. A representative set of peptides to synthesize and test was selected using SMD in building block space using amino acid descriptors. Several characterizations of amino acids have been presented previously,[38,55−59] and QSAR models of class I and class II MHC peptides described in the literature have been based on additive descriptors,[47,48,53,59−61] selected property descriptors,[62] and the *z*-scale descriptors[53,62] defined by Wold and Sandberg.[58] The *z*-scale descriptors are derived on the basis of experimentally determined characterization of the amino acids followed by a subsequent principle component analysis (PCA)[63] resulting in three to five principle properties (i.e., the *z*

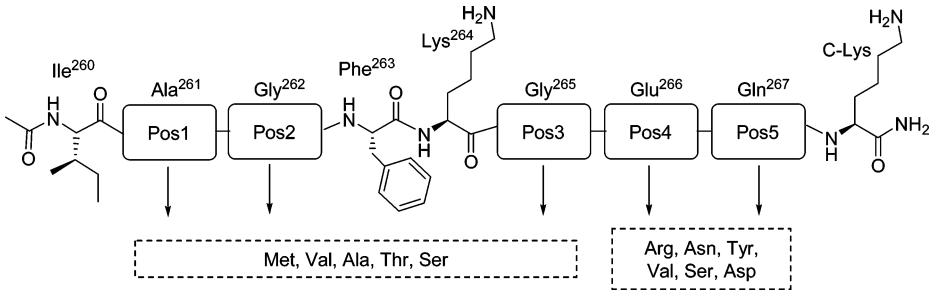


Figure 2. Peptide scaffold used to study molecular property preferences for peptide binding to the A^q MHC class II molecule based on the minimal CII glycopeptide with retained T-cell response (CII260−267). Five positions in the scaffold were varied by SMD, and the effects were studied. Major anchor residues (positions 260 and 263) were left unchanged. The nonmodified lysine residue was used in position 264, and an additional lysine was added to the C-terminal for solubility purposes. The arrows indicate the amino acid sets selected for incorporation at each position.

Table 1. Amino Acid Sequences of N- and C-Terminally Extended Peptides and Resulting T-Cell Responses

	position										hybridoma[a]	
peptide	259	260	261	262	263	264	265	266	267	268	HDB2	HCQ4
N-Lys	Ac-Lys	Ile	Ala	Gly	Phe	Lys	Gly	Glu	Gln-NH_2		+	−
C-Lys		Ac-Ile	Ala	Gly	Phe	Lys	Gly	Glu	Gln	Lys-NH_2	+	+
CII260−267[b]		Ac-Ile	Ala	Gly	Phe	Lys	Gly	Glu	Gln-NH_2		+	+

[a] + refers to an equally strong response to the corresponding longer peptide CII259−275, and − refers to no response at the tested concentrations. [b] Included as reference peptide.

scales).[56,58] Even though these descriptors have proven to be successful in several QSAR models of peptide binding, initial studies indicated that these descriptors were not sufficient for our problem, particularly in mapping the flexibility of the amino acids (data not shown). Furthermore, calculated molecular descriptors have been shown to contain similar information as experimentally derived ones.[64] Hence, a new set of 28 calculated molecular descriptors was selected to represent important molecular properties for binding such as size, flexibility, electronic description, lipophilicity, and hydrogen bond donor and acceptor capabilities. The characterization was made using semiempirical-based descriptors, indexes, and character counts, and the number of descriptors was compressed to its main principle properties using PCA in a manner similar to that for the z-scale descriptors. During the cause of this work a new set of amino acid descriptors was reported using a similar approach of using theoretical descriptors and a subsequent PCA.[65]

The PCA of the molecular descriptors resulted in four to five principal properties (four according to the Scree plot but five with eigenvalues larger than 1) for which the most important (**t1**−**t3**, describing 65% of the variation) were further used in the modeling procedure as design variables (Figure 3). The first component, **t1**, mainly separated amino acids based on size; small amino acids such as glycine and alanine had high score values, while large amino acids such as arginine and tryptophan had low **t1** score values (Figure 3a,b). For the second and third components (**t2** and **t3**, respectively), separation was mainly based on lipophilicity and flexibility, respectively. Three groups of amino acids could also be distinguished in the **t1** vs **t2** score plot (Figure 3a). Amino acids with hydrogen donor and acceptor capabilities, such as asparagine, arginine, and glutamic acid, are located in the upper-left quadrant. Aromatic amino acids are found in the lower-left quadrant with aliphatic ones in the lower-right quadrant. The two remaining principal properties, **t4** and **t5**, describe less generally applicable characteristics for binding, e.g., separation of amino acids containing hydroxyl or sulfur groups.

A comparative model of the A^q molecule[20] indicates that there is limited room for bulky side chains at positions 1−3 (corresponding to positions 261, 262, and 265), where small residues (glycine and alanine) occur in CII. This implies that the presence of large amino acids at any of these positions could severely reduce the peptides' ability to bind to the A^q molecule. Therefore, the principal property design space was reduced for these three positions to cover only amino acids of small to moderate sizes. Five amino acids representing this space were selected: Thr, Ser, Val, Met, and Ala (Figures 2 and 3a,c). For positions 4 and 5 (corresponding to CII266 and CII267) there was, according to the A^q model, no need for restrictions in the chemical space. Therefore, the total property space was used to cover as large an area as possible, and six amino acids were chosen: Arg, Asn, Asp, Tyr, Ser, and Val (Figure 2). These amino acids represented a maximum spread in characteristics such as aliphatic/aromatic, hydrogen bond donors and acceptors, size, and charged/neutral, as described by the three score values in the principal property space (Figure 3a,c).

Library Selection Using D-Optimal Design. Virtual combinations yielding all possible variations of the selected amino acids for the five positions in the scaffold resulted in a peptide library consisting of 4500 peptides ($5^3 \times 6^2$). D-optimal design[66] was applied to this library to reduce it further. Each amino acid at the five altered positions was represented by the three values of the scaled principal properties, **t1**−**t3**, which when combined into peptides resulting in 15 values representing a 15-dimensional principal property space. A library of 22 peptides with the most homogeneous distribution possible of selected amino acids at each position was chosen out of the suggested D-optimal designs while maximizing the volume spanned in the principal property space. Two peptides with molecular properties closest to the center of the principal property space were added as center points. The procedure used to select these 24 peptides ensured that the chosen library had high chemical diversity and information content (Table 2).

Synthesis, Biological Testing, and Preanalysis of Data. The peptides of the library (Table 2) were synthesized on solid phase as N- and C-terminal amides. Cleavage from solid support and purification by reverse-phase HPLC rendered the pure products, hereafter referred to as **1**−**22** and **CP1**−**CP2**. The binding

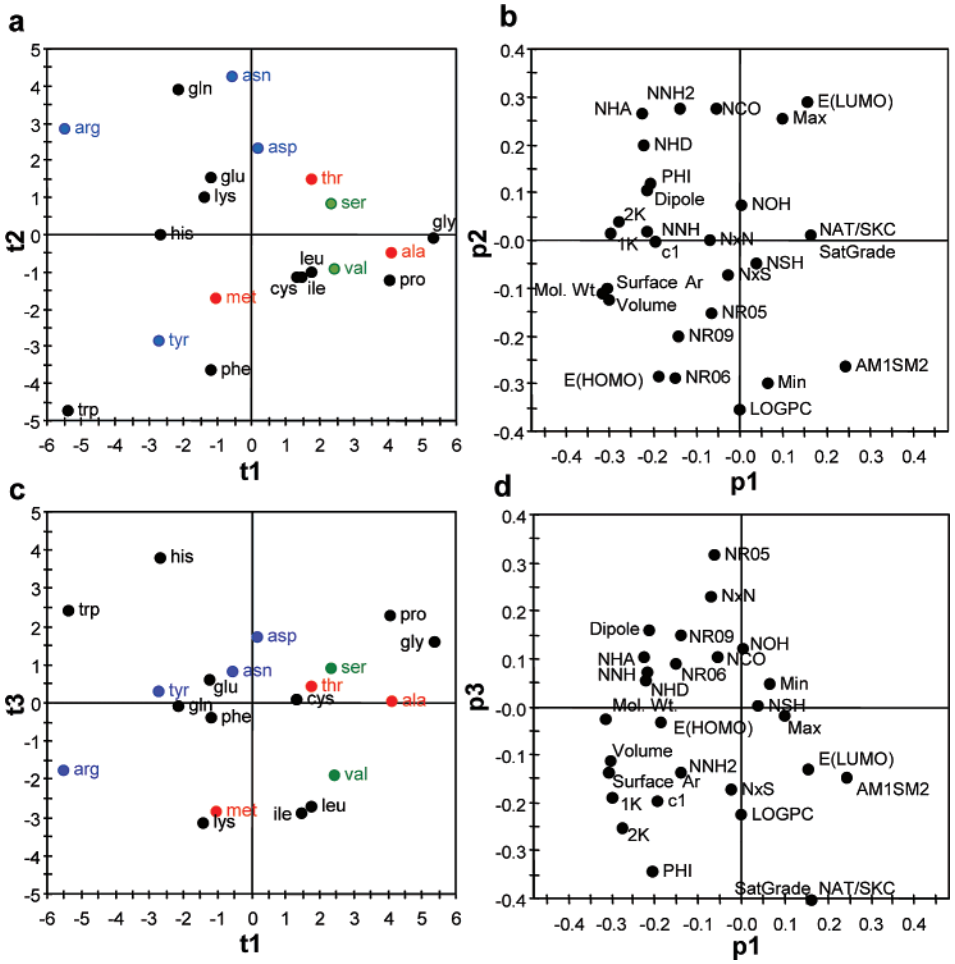


Figure 3. Score (a, c) and loading plots (b, d) resulting from PCA of the 20 coded amino acids described by 28 molecular descriptors. First versus second component plots are visualized in (a) and (b), while first versus third component plots can be seen in (c) and (d). Amino acids indicated in red (Met, Ala, Thr) and green (Val, Ser) were chosen as building blocks for the variations at positions 1−3. The larger principal property space covered by the building blocks in blue (Arg, Asn, Tyr, Asp) and green (Val, Ser) was used for positions 4 and 5. Explanations of the molecular descriptor abbreviations in (b) and (d) are given in the Experimental Section.

strength of the peptides to the MHCII A^q molecule was studied in a competitive assay in which the peptides were evaluated for their ability to prevent the binding of a biotinylated CLIP reference peptide to A^q-transfected cells. The test peptides were incubated at seven different concentrations (750, 250, 83, 28, 9, 3, and 1 μM; designated concentrations 1−7, respectively) in duplicate, and the experiments were repeated twice. Structure−activity relationships were evaluated using the % inhibition at different concentrations of the peptides as the biological response.

The partial least-square projections to latent structures (PLS) method[67,68] has been the most commonly used regression method for developing QSAR models of peptides binding to class I and class II MHC molecules,[47,48,53,59−62] although a recent publication has shown promising QSAR results with support vector machine regression.[69] Here, the 15 principal property values used for the library selection, i.e., the combination of the PCA score values **t1**−**t3** for all the peptides (the **X** matrix), were correlated to the biological response (the **Y** matrix) by PLS regression.

The analysis of the **X** and **Y** data revealed that the responses at concentrations 2−4 (i.e., 250, 83, and 28 μM) contained the most information, since most peptides did not bind at the lower concentrations 5−7 and that solubility and/or toxicity problems were detected for some peptides at the highest concentration (750 μM, concentration 1). The response block (concentrations 2−4) had a linear relationship to the logarithm of the corresponding concentration, and nontransformed responses gave similar results as the logit-transformation (cf. pIC_{50}). The response values were scaled to unit variance, which gave results similar to those for the Pareto scaling. Single, duplicate, and quadruplet samples showed good reproducibility both in and between experiments except for run 2 of peptides **1**−**9** where the results deviated from the others and hence were excluded from further analysis. The most robust models were obtained when averages of duplicate samples were used as the biological responses. Therefore, in the final QSAR model the biological responses of the peptides were represented by the average percentage inhibition of the pairs of duplicate samples at concentrations 2−4 (referred to as Y2, Y3, and Y4) except that peptides **1**−**9** were each represented by a single duplicate sample average (Table 2). Peptides **7** and **9** and the two center peptides (**CP1** and **CP2**) yielded deviating data because of solubility problems and/or toxicity to the cells in the binding assay and were not included in the modeling.

QSAR models of peptides binding to class I and class II MHC molecules using PLS regression have been reported including only linear terms[47,48,53,60,62] but also linear *and* cross-terms that account for interactions between side chains at relative positions 1−2 and 1−3.[48,59,61] A comparison of the two models based on the data in this study revealed identical interpretation of the linear terms when considering the PLS weight vectors (regres-

Table 2. Amino Acid Sequences and A^q Binding Data for the Members of the Peptide Library Used for QSAR Modeling

	varied position									inhibition (%)[b]		
peptide[a]		Pos1	Pos2			Pos3	Pos4	Pos5		Y2	Y3	Y4
1	Ac-Ile	Met	Met	Phe	Lys	Met	Ser	Asp	Lys-NH_2	30/-	17/-	4/-
2	Ac-Ile	Met	Met	Phe	Lys	Val	Arg	Arg	Lys-NH_2	74/-	52/-	36/-
3	Ac-Ile	Met	Met	Phe	Lys	Thr	Asp	Val	Lys-NH_2	16/-	24/-	10/-
4	Ac-Ile	Met	Thr	Phe	Lys	Met	Tyr	Val	Lys-NH_2	48/-	40/-	13/-
5	Ac-Ile	Met	Thr	Phe	Lys	Ala	Val	Tyr	Lys-NH_2	72/-	59/-	47/-
6	Ac-Ile	Met	Ser	Phe	Lys	Thr	Arg	Asp	Lys-NH_2	54/-	27/-	15/-
7	Ac-Ile	Val	Val	Phe	Lys	Ser	Arg	Tyr	Lys-NH_2	- -	- -	- -
8	Ac-Ile	Val	Ala	Phe	Lys	Thr	Val	Val	Lys-NH_2	41/-	27/-	22/-
9	Ac-Ile	Val	Thr	Phe	Lys	Val	Asn	Asn	Lys-NH_2	- -	- -	- -
10	Ac-Ile	Val	Ser	Phe	Lys	Ala	Tyr	Arg	Lys-NH_2	51/39	60/61	49/47
11	Ac-Ile	Ala	Met	Phe	Lys	Met	Arg	Ser	Lys-NH_2	77/78	61/63	40/38
12	Ac-Ile	Ala	Met	Phe	Lys	Ser	Val	Asn	Lys-NH_2	49/50	32/29	18/19
13	Ac-Ile	Ala	Val	Phe	Lys	Met	Val	Arg	Lys-NH_2	68/70	54/59	40/42
14	Ac-Ile	Ala	Ala	Phe	Lys	Ala	Asp	Ser	Lys-NH_2	47/53	31/36	18/31
15	Ac-Ile	Ala	Thr	Phe	Lys	Thr	Tyr	Tyr	Lys-NH_2	70/67	58/54	38/21
16	Ac-Ile	Thr	Met	Phe	Lys	Met	Asp	Arg	Lys-NH_2	62/61	46/38	25/26
17	Ac-Ile	Thr	Met	Phe	Lys	Val	Val	Tyr	Lys-NH_2	59/46	50/42	35/32
18	Ac-Ile	Thr	Val	Phe	Lys	Val	Tyr	Asn	-NH_2	56/50	41/36	24/23
19	Ac-Ile	Thr	Thr	Phe	Lys	Ser	Ser	Ser	Lys-NH_2	20/13	6/2	6/-3
20	Ac-Ile	Thr	Ser	Phe	Lys	Met	Arg	Tyr	Lys-NH_2	75/80	61/69	48/50
21	Ac-Ile	Ser	Ala	Phe	Lys	Thr	Asn	Arg	Lys-NH_2	84/75	56/61	40/42
22	Ac-Ile	Ser	Thr	Phe	Lys	Ala	Arg	Val	Lys-NH_2	70/74	56/58	39/39
CP1	Ac-Ile	Val	Val	Phe	Lys	Val	Asp	Asp	Lys-NH_2	- -	- -	- -
CP2	Ac-Ile	Val	Ala	Phe	Lys	Val	Ser	Asp	Lys-NH_2	- -	- -	- -

[a] Peptides **1**−**22** were selected by D-optimal design while **CP1** and **CP2** represented center points. [b] % inhibition in which the average was obtained from duplicate samples in two separate runs for concentrations Y2, Y3, and Y4 corresponding to 250, 83, and 28 μM, respectively. Single dash (-) indicates that data from run 2 for peptides **1**−**9** were excluded because they were deviant. Double dashes (- -) indicates deviating data due to solubility and/or toxicological problems.

sion coefficient of 1.0) and a similar pattern of the normal probability plot of the residuals. The model statistics showed that the inclusion of the cross-terms gave a slightly higher R^2Y ($\Delta R^2 = +0.05$) than the model based on only the linear terms but a much worse Q^2 value ($\Delta Q^2 = -0.27$). Exclusions of cross-terms with low model coefficients improved the Q^2 value but not better than the linear model ($\Delta Q^2 = -0.03$). It appears that for our data inclusion of cross-terms did not improve the models significantly. Interestingly, the only other study, to our knowledge, for which interaction terms between side chains in a QSAR model for MHC class II molecules have been investigated resulted in the same conclusion.[48]

QSAR Model. The final linear PLS model showed a good correlation between the experimental and calculated inhibition values (Figure 4). This two-component model explained 66% of the variation in the biological response ($R^2_{Y2} = 0.59$; $R^2_{Y3} = 0.70$; $R^2_{Y4} = 0.70$) with a cross-validated Q^2 of 43% ($Q^2_{Y2} = 0.37$; $Q^2_{Y3} = 0.47$; $Q^2_{Y4} = 0.45$). The dModX plot and the normal probability plot of the residuals did not reveal any outliers.

The QSAR model was validated and tested for its predictability using an external test set. The affinity to A^q was predicted by the QSAR model for the 3 200 000 possible virtual peptides. Three clusters, each consisting of 5000 peptides predicted to have high, average, and low A^q affinity, can be seen in the PLS discriminant analysis (PLS-DA) plot (Figure 5). Most library peptides belonged to the class of peptides predicted to have average affinities, while several were predicted to have low affinity and only one was found in the high-affinity region (peptide **20**), indicating that there is scope for designing improved binders. Two peptides from each class (high-, medium-, and low-ranked binders) were selected as an external test set (Figure 5 and Table 3). These six peptides, referred to as **V1**−**V6**, were synthesized and tested for binding strength to A^q. These validation peptides were tested on a separate, later occasion with a different setup of biological material and modified protocol compared to the peptides used to build the

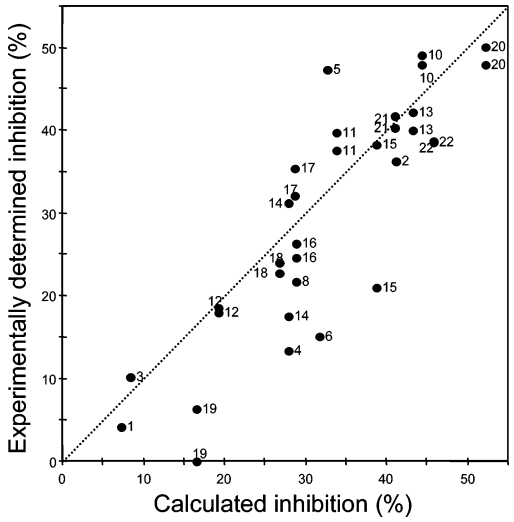


Figure 4. Calculated versus experimentally determined inhibition values for one of the three responses (Y4) used in the multi-**Y** PLS regression. The QSAR model was based on 15 principal property values (**t1**−**t3** at positions 1−5) for 20 peptides, and the three biological responses were represented as % inhibition at three different peptide concentrations (Y2, Y3, Y4).

QSAR model. Hence, the binding data could not be directly compared in terms of percentage inhibition, so we determined which of the three classes (i.e., high, medium, or low predicted affinity) each of the peptides belonged to and also ranked them relative to each other on the basis of the biological binding data.

The model successfully distinguished binders from nonbinders (Table 3). All the peptides that were predicted to bind with either high or average affinity to A^q (peptides **V1**−**V4**) did bind to the A^q molecule, while the peptides predicted to have the lowest affinity (**V5** and **V6**) showed no response at all. The highly scored peptide **V1** was found to give the strongest measured response. The other highly scored peptide **V2** and the peptides predicted to have average affinity (**V3** and **V4**) displayed similar

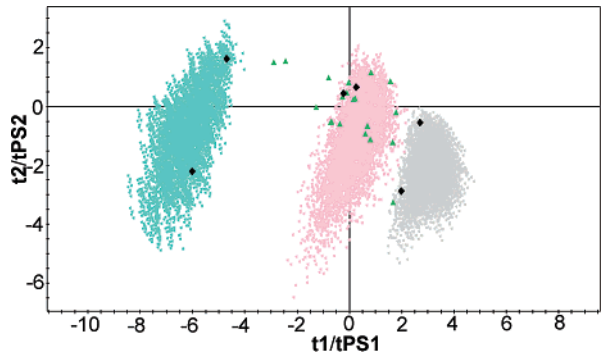


Figure 5. PLS-DA score plot (**t1** versus **t2**) based on three classes of virtual peptides predicted by the QSAR model to be strong, medium, and weak binders to the A^q molecule. The 5000 highest ranked peptides are located to the right (gray). The 5000 closest to the average are located in the center (pink), while the 5000 lowest ranked are located to the left (blue) in the plot. The six peptides constituting the external test set are marked with black diamonds. The distribution of the peptides building up the QSAR model is illustrated by green triangles, which show their predicted PLS-DA score values (**tPS1** and **tPS2**).

binding preferences and had lower binding affinity to A^q than **V1**. The results of the external validation confirmed the predictability of the QSAR model obtained.

A comparison of our model with the QSAR models developed for other mouse MHC class II molecules (I-A^b, I-A^d, I-A^k, I-A^s, I-E^d, and I-E^k)[48] reveals that the R^2 and Q^2 are lower for our model (0.66 and 0.43 compared to 0.99 and 0.83). However, R^2 and Q^2 are internal validation criteria and external test sets are a superior alternative for evaluating the model quality. The predictive power of the models for the six class II molecules presented by Hattotuwagama et al. showed great variation but had similar high R^2 and Q^2 values of the models, where the A^b model had a very high predictivity while that of E^k was very poor.[48] Our model successfully managed to predict the binding strength of six new peptides despite the relatively low internal statistical terms.

The interpretation of the influence of the different amino acid properties at the different positions is presented in detail below. The regression coefficients of the different peptide positions showed the same pattern for all three investigated concentrations, and the regression weight values ($w \times c$ plot) can be seen in Figure 6. A design guide of preferred directions in the score plots of the 20 amino acids is provided in the Supporting Information.

1. Interpretation of the QSAR Model. Positions 4 and 5 had the strongest influence on the peptide binding to A^q according to the regression model, as shown by their dominating regression weight values (Figure 6). The variables describing size and flexibility were the main contributors for position 4, as both **t1** and **t3** were strongly negatively correlated with the response. For a good binder the amino acid at position 4 should preferably be large and flexible, e.g., arginine. The original molecular descriptors responsible for these features were revealed, based on their PCA loading values, to be volume, surface area, and the Kier flexibility indexes 1K, 2K, and PHI (Figure 3). In addition, the positively correlated regression weight for **t2** indicated that hydrogen bond donors/acceptors could be preferred over more lipophilic amino acids. Although the side chain of this residue has been shown to be very important for T-cell stimulations,[25] our model shows that it also influences binding to the A^q. By comparison of the relative position in the principal property space of glutamic acid found in rat CII with that of aspartic acid found in mouse CII, it can be seen that glutamic acid is preferred to aspartic acid in position 4. This correlates well to experimental results suggesting that rat CII binds more strongly to the A^q molecule than mouse CII[70] and correlates with the finding that exchanging glutamic acid by aspartic acid at position 266 in CII256−270 resulted in a more than 10-fold reduction in binding strength to A^q.[20]

In position 5, just as in position 4, large and to some extent flexible amino acids were strongly preferred because the **t1** and **t3** variables were negatively correlated with the response (Figure 3 and 6). Interpretation of the PCA loading vectors **p1** and **p3** showed that the dominating molecular descriptors, as for position 4, were volume, surface area, and the Keir flexibility indexes. The lipophilicity, as explained by **t2**, had a moderate impact on binding, and the regression weight value indicates a binding preference for hydrophobic and aromatic amino acids, e.g., phenylalanine and methionine.

In position 3, the regression weight values of **t2** and **t3** variables were of moderate sizes and negatively correlated with the response. The dominating original molecular descriptors were LOGPC, AM1SM2, and the Keir indexes, indicating preferences for hydrophobic and flexible amino acids (Figure 3). However, the size of the side chains does not appear to be an important feature for binding within the limited studied chemical space (cf. size of Pos3:**t1** in Figure 6). The most suitable coded amino acids for this position belong to any of the two adjacent clusters containing isoleucine, leucine, valine, methionine, or maybe even the aromatic phenylalanine.

The PLS regression weight values for the side chain properties at positions 1 and 2 described by **t1**−**t3** were low to medium, indicating that the variations made at these positions had no major effect on the binding of the peptide to the A^q molecule. All amino acids within the investigated area in the principal property space were tolerated. However, the moderate positive weight values of **t1** and **t3** indicate that small rigid amino acids were preferred (Figure 3), while the polarity of the amino acids did not seem to matter (cf. Pos1:**t2** and Pos2:**t2** in Figure 6). The preferred amino acids at these positions correspond well with those naturally occurring in the CII peptide, i.e., Ala261 and Gly262.

2. Summarizing the Model. Preferred amino acids and characteristics of the studied positions are summarized in Figure 7. A molecular property binding motif was discernible and easily transferred to sequence binding motifs. Besides the previously identified CII260 and CII263 anchor positions, positions 4 and 5 (corresponding to Glu266 and Gln267 in CII) were the most important for peptide binding to the A^q molecule. Residues at these two positions should preferably be large and flexible. In addition, the model indicates that residues at position 4 should contain hydrogen bond donors and acceptors to promote binding, while residues at position 5 should be hydrophobic. At positions 1 and 2 (corresponding to Ala261 and Gly262 in CII) small and rigid amino acids are favored, even though moderately sized amino acids were tolerated. Finally, position 3 (corresponding to Gly265 in CII) should harbor aliphatic, hydrophobic residues for optimal binding. These proposed preferences were supported by the strong A^q binding of peptide **V1** from the external test set. For **V1** the amino acids at all positions fulfilled the suggested requirements: Thr at position 1, Ala at position 2, Ile at position 3, Arg at position 4, and Trp at position 5. It should be noted that this designed peptide has no amino acids in common at any of the five varied amino acid positions with the octamer originating from CII. On the other hand, peptide **V5**, which was predicted to be a poor binder, had unfavorable amino acids at all of the investigated positions, effectively preventing its binding to the A^q molecule. The A^q binding motif

Table 3. Amino Acid Sequences and A^q Binding Data for the Six Peptides Used for External Validation of the QSAR Model

	varied positions								A^q binding[a]		
peptide		Pos1	Pos2			Pos3	Pos4	Pos5		pred[b]	found[c]	rank[d]
V1	Ac-Ile	Thr	Ala	Phe	Lys	Ile	Arg	Trp	Lys-NH_2	+	+	1
V2	Ac-Ile	Ile	Gly	Phe	Lys	Ala	Arg	Met	Lys-NH_2	+	+, m	2
V3	Ac-Ile	Ile	Gln	Phe	Lys	Thr	Leu	Arg	Lys-NH_2	m	+, m	2
V4	Ac-Ile	Thr	Asn	Phe	Lys	Gly	Ser	Arg	Lys-NH_2	m	−, m	4
V5	Ac-Ile	Tyr	Met	Phe	Lys	Glu	Pro	Gly	Lys-NH_2	−	−	5
V6	Ac-Ile	Gln	Met	Phe	Lys	Arg	Ala	Gly	Lys-NH_2	−	−	6

[a] Definitions: +, strong binders; m, medium binders; −, poor binders. [b] Predicted class membership by the PLS-DA model. [c] Class membership as determined from binding data. [d] Relative ranking of peptides based on binding data.

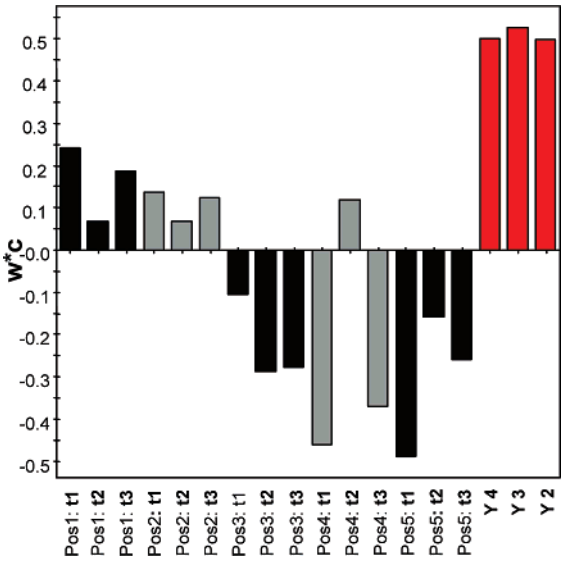


Figure 6. PLS weight values ($w \times c$) for the QSAR model based on 15 principal property values (**t1**−**t3** at positions 1−5) and three biological responses represented as % inhibition at three different peptide concentrations (Y2, Y3, Y4).

indicated by the QSAR model could be used in future studies to predict the binding propensity of other self-peptides presented by the disease-associated A^q molecule. The results of this study do not support the hypothesis that the KXXS motif, which has been postulated to be associated with A^q-restricted antigenicity,[28] is important for binding. Similar results were also seen in a recent study reporting significant features for T-cell recognition, where the GalHyl-X-E motif was identified to be extremely important for the T-cell response.[25] The QSAR models of peptides binding the six mouse alleles presented by Hattotuwagama et al. did not show a common theme for favored and disfavored amino acid residues.[48] A comparison with the interpretation of our model did not reveal any clear common favored peptide binding pattern between A^q and any of the six other modeled mouse MHC class II molecules. These results are not surprising because MHC molecules are known to have diverse characteristics in presenting peptide antigens.

The longer immunodominant part of CII, i.e., CII256−270, binds to both of the MHC class II molecules A^q (mouse) and DR4 (human).[20,71,72] It has been proposed that the DR4-binding motif of CII256−CII270 is shifted by three amino acids compared to that of A^q, resulting in Phe263 and Glu266 being at the anchor positions instead of Ile260 and Phe263.[71,73] A comparison of the suggested DR4 peptide-binding motifs[33,34] with our results regarding the A^q-binding motif supports this hypothesis, but too few amino acids have been altered within the DR4 epitope to warrant further comparative conclusions.

No experimentally determined 3D structure of the A^q molecule is available, but indications about the properties of the A^q binding site can be obtained from the postulated preferred molecular properties of the peptides (Figure 7). In addition to earlier indications of two large hydrophobic binding pockets (P1 and P4),[20] the preferences for small, rigid amino acids in the P2 and P3 binding pockets suggest that they are of limited size. One could also speculate that the presence of large amino acids in P2 and/or P3 could prevent the anchor residues from reaching their binding pockets. In P6 the QSAR model implies that there is a hydrophobic area in the A^q molecule. The P7 pocket appears to be rather large with some polar surface area, while the P8 pocket is indicated to be a large, hydrophobic binding area. When these preferences were compared with the 3D comparative model of the A^q molecule,[20] the models matched remarkably well.

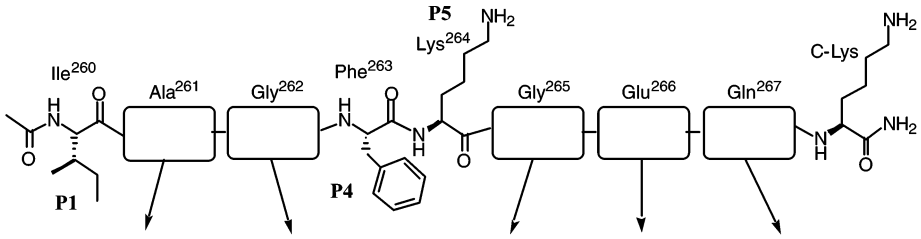
In the preferred design directions, as indicated by the QSAR model, there are few amino acids with required characteristics among the coded amino acids. More suitable non-natural amino acids could possibly be identified to further enhance the binding to the A^q molecule, which could also increase the metabolic stability of the peptides. In addition, the information presented in Figure 7 could be used to develop novel peptide mimetics based on the proposed preferred molecular properties. New modified peptides with variations in physicochemical properties and binding strength would be highly valuable in the development of effective immunization procedures for use in future vaccination studies.

Conclusions

A peptide library was designed, synthesized, and evaluated for binding to the mouse MHC class II molecule A^q. The SMD approach made it possible to select a chemically diverse and informative library of 22 peptides and two center points out of 3 200 000 possible peptide combinations. This highly reduced set of peptides, together with inhibition data from a cell-based competitive assay at three different concentrations of the peptides, resulted in a high-quality QSAR model based on PLS modeling that was successfully validated with an external test set of six peptides.

A molecular property binding motif for peptides binding to the mouse MHC class II molecule A^q was established on the basis of interpretation of the QSAR model of the five varied positions. The C-terminal positions of the peptide scaffold (corresponding to CII266 and CII267, respectively) appeared to have the strongest influence on the A^q/peptide interaction, while the positions corresponding to CII261 and CII262 are the ones that have the least influence according to the QSAR model. In addition, the model provided indications of the characteristics of the binding site of the A^q molecule and the findings in the present investigation correlated well with a comparative 3D model of the protein.

The QSAR model provides novel information and insight regarding the A^q molecule/peptide component of the ternary A^q, glycopeptide, and T-cell receptor complex. This information



Peptide pos. / A^q pos.	Pos1 / P2	Pos2 / P3	Pos3 / P6	Pos4 / P7	Pos5 / P8
Preferred binders					
Molecular property	small rigid	small rigid	hydrophobic flexible	large flexible (H-donors & acceptors)	large flexible (aromatic, hydrophobic)
Suggested amino acids	G, P, A (S, T)	G, P, A (S, T)	I, L, M, V (F)	R, Q (M, K, N)	F, M (R, K)
Poor binders					
Molecular property	large flexible	large flexible	polar H-donors & acceptors	small rigid (aliphatic)	small rigid (polar)
Suggested amino acids	R, K, M	R, K, M	D, N, E, Q S, T	P, G	P, G (S)
Binding site properties					
A^q molecule	small polar	small polar	hydrophobic	big pocket polar	big pocket hydrophobic

Figure 7. Summary of the interpretation of the QSAR model illustrating the preferred and rejected amino acids and of molecular properties to promote A^q binding. Pos1−Pos5 represent the five varied positions in the peptide scaffold, and suggested characteristics at corresponding positions in the binding site of the A^q molecule (P2, P3, P6−P8) are also shown. Amino acids are represented by their conventional one-letter code.

could facilitate attempts to develop new treatments of autoimmune diseases such as RA.

Experimental Section

Theoretical Characterization of Amino Acids. The structures of the 20 naturally occurring amino acids were generated using Spartan software[74] and subsequently characterized by 11 molecular descriptors including properties related to size (surface area, molecular weight, volume), electronic features (dipole, HOMO, LUMO, maximum charge, minimum charge, partial charge of c-α-c1), and lipophilicity (log *P*, SM2) extracted from semiempirical AM1 calculations within Spartan software.[74] In addition to these 3D-based descriptors, 17 descriptors including functional group counts (numbers of OH, NH, NH_2, SH, CO, other N $\{N \times N\}$ and other S $\{N \times S\}$ groups), numbers of five-, six-, and nine-membered rings, number of hydrogen donors {NHD} and acceptors {NHA}, indexes (path 1 Keir shape index-1K, path 2 Keir shape index {2K}, Kier flexibility index {PHI}), and saturation ratios (SatGrade, NAT/SKC) were computed using the Dragon software.[75] A table of the descriptors is given in the Supporting Information. These molecular descriptors were compressed by PCA[63,76,77] using SIMCA software.[78] The number of significant principal components was decided using their eigenvalues, a Scree plot, and chemical interpretation of the loadings for the corresponding components.

Statistical Molecular Design and Data Analytical Methods. 1. Selection of Peptides. Each amino acid at the altered positions was represented by the corresponding values of the scaled principal properties score vectors. Peptides resulting from different combinations of amino acids yielded the structure descriptor matrix (**X**). The score values were then scaled to unit variance for each dimension to avoid bias in the weighting of the varied positions or molecular properties due to differences in variance. D-optimal design[50,66,79] was performed using MODDE software[80] to generate 15 libraries, each with 22 peptides. The selection by D-optimal design maximized the volume spanned in the principal property space through maximization of the determinant (Det) of the **X′X** matrix. The final library to synthesize was chosen from the 15 generated libraries based on G efficiency, Log(Det of **X′X**), Norm.log(Det of **X′X**), condition number, and the most homogeneous distribution of the selected amino acids at each position. The 50 peptides with the minimal Euclidean distance to the calculated center point in the principal property space were calculated, and two peptides were chosen as center points based on their synthetic feasibility and added to the library of 22 peptides.

2. Projections to Latent Structures by Means of Partial Least-Squares. The structure descriptor matrix (**X**), based on the combination of PCA score values representing the molecular properties of the amino acids, was related to the biological activity response matrix (**Y**) using the PLS regression method.[67,68] The PLS method maximizes the covariance between the latent variables of the **X** and **Y** matrixes (multi-**Y**) and correlates these latent variables through linear combinations to a regression model. Even though it is a linear method, a nonlinear relationship can be handled to some extent through transformation of **Y**, inclusion of nonlinear terms, and extraction of additional PLS components. The % inhibition at several different concentrations (multi-**Y**), scaled to unit variance, was used as the response. The use of correlated response variables as a multi-**Y** matrix enhanced the stability and reliability of the models because the biological data contained noise and, in some cases, missing and deviating data. The quality of the model was investigated by estimating the amounts of explained variation (R^2) and cross-validated predicted variation (Q^2)[81] using seven cross-validation rounds. The validation of the final model was made by using an external test set. The interpretation of the influence of the different amino acid positions was based on the first PLS component weight vector, which provides the best estimate of the variable importance when only one response or highly correlated multiresponses are used.[82] All PLS modeling was performed using SIMCA software.[78]

Preanalysis of Biological Data. The **X** and **Y** data were preanalyzed to determine relevant concentrations, scaling preferences, transformations, included model terms, and influence of the experimental layout to assess the reproducibility of the assays and to detect peptide outliers. Relevant concentrations as multi-**Y** and

model terms to include in the final PLS-QSAR model and whether the data should be logit-transformed and scaled to unit variance, pareto-scaled or nonscaled, were assessed by considering the amounts of variation explained by the resulting PLS models (R^2), their cross-validation values (Q^2), and the normal probability plots of the residuals. Three sets of model terms were investigated: only linear terms, linear and all cross terms for relative positions 1−2 and 1−3, and linear and selected cross terms (1−2 and 1−3) with PLS coefficients larger than 0.05 for all investigated concentrations. In order to test the biological effects of the entire peptide library including center points and reference peptides as duplicated samples, the experiment was divided and tested in three different experiments (experiment 1, peptides **1**−**9**; experiment 2, peptides **10**−**18**; experiment 3, peptides **19 CP**s). The experiments were then repeated (runs 1 and 2), resulting in four samples/response curves for each peptide. PCA of the **Y** variables, for all four samples (duplicated samples and duplicated runs), was used to assess the reproducibility both in and between experiments. Standard deviations were calculated, and a PLS with indicator variables was performed to identify deviating experiments. Thus, the **X** matrix was extended by six columns with either 1 or 0 as indicator variable, representing belonging or not belonging to a certain experiment number and run. Deviating data due to solubility problems and/or toxicity to cells used in the binding assay were visually detected when running the FACS analysis and/or as outliers in the PCA.

External Validation. A representative test set for external validation was chosen from all possible 3 200 000 in silico generated peptides, based on the coded amino acids. The virtually combined peptides were represented by their values of the scaled principal properties score vectors, and the established QSAR model equation was used to predict their binding strength (% inhibition). Three ranking algorithms were then used to sort and select three sets of peptides: the 5000 predicted to be the strongest binders, the 5000 predicted to have the closest to average binding strength, and the 5000 predicted to be the weakest binders.[83] Separation of the three classes of predicted peptides was enhanced using PLS-DA, and selections were made from each class. The test set was synthesized and biologically evaluated for A^q binding.

Solid-Phase Peptide Synthesis. The peptides were synthesized in a manually operated reactor or a Pioneer peptide synthesis system (Applied Biosystems, The Netherlands), using standard solid-phase peptide synthesis methodology on a Tentagel-S NH_2 resin (Rapp Polymere, Germany) in which the linker Fmoc-2,4-dimethoxy-4′-(carboxymethyloxy)benzhydrylamine (Rink) was first coupled to the resin. This rendered peptides as C-terminal amides after cleavage from the resin. N^{α}-Fmoc amino acids carrying standard side chain protective groups (Bachem, Switzerland and Neosystem S.A., France, 4 equiv), as well as the Rink linker (4 equiv), were coupled to the resin in dimethylformamide (DMF), which was predistilled and used immediately or stored for a short time over 3 Å molecular sieves. In the manually operated reactor diisopropyl carbodiimide (DIC, 3.9 equiv) was used as a coupling reagent in the presence of 1-hydroxybenzotriazole (HOBt, 6 equiv). The progress of the reaction was monitored by the naked eye using bromophenol blue as an indicator.[84] Alternatively, coupling reactions were performed in the Pioneer peptide synthesis system utilizing 0.5 M HBTU and 0.5 M DIPEA as coupling reagents with UV monitoring, all according to the manufacturer's instructions. Fmoc protective groups were removed after each coupling cycle using 20% piperidine in DMF. After completion of the synthetic sequence, the N-terminal of the peptides was acetylated by incubation with Ac_2O/DMF (1:2) for 1 h. The peptides were cleaved and deprotected by incubation with TFA/H_2O/thioanisole/ethanedithiol (35:2:2:1) for 3 h at 40 °C. Following repeated concentration from HOAc, the peptides were precipitated from Et_2O and the crude products were freeze-dried. Purification by reversed-phase HPLC and freeze-drying gave the homogeneous compounds (**N-Lys**, **C-Lys**, **1**−**22**, **CP1, CP2**, **V1**−**V6**). Analytical reversed-phase HPLC was performed using a Kromasil C-8 column (250 mm × 4.6 mm, 5 μm, 100 Å), elution with a linear gradient of MeCN (0 → 100% or 0 → 80% over 60 min), balance H_2O, both containing 0.1% TFA, and flow rate of 1.5 mL/min. Preparative reversed-phase HPLC was performed using a larger Kromasil C-8 column (250 mm × 20 mm, 5 μm, 100 Å) with the same eluents but a flow rate of 11 mL/min. In both cases the eluate was monitored by a UV detector at 214 nm. The identity of the peptides was confirmed by MS and their purity (≥95%) by analytical HPLC.

A^q Binding Assay. The binding of peptides to A^q MHC class II molecules was preformed in a competitive assay using flow cytometry analysis. Briefly, the test peptides (**1**−**22**, **CP1**, **CP2**, **V1**−**V6**) and a reference peptide were incubated in 96-well plates at seven different concentrations (750, 250, 83, 28, 9, 3, and 1 μM, which are concentrations 1, 2, 3, 4, 5, 6, 7, respectively) for 2.5 h at 37 °C with a fixed concentration of biotinylated CLIP peptide (5 μM) and M12Q 14-7 cells transfected with H-2A^q. After being washed to remove excess peptide, the cells were stained with 0.2 μL of streptavidin−phycoerythrin (SAPE), which binds to the biotinylated CLIP peptide. The phycoerythrin (PE) dye was detected by flow cytometry analysis using FACSort (Becton Dickingson, San Jose, CA) and Becton Dickingson software. The "% inhibition" for each peptide was calculated from the gated mean fluorescence by comparison with the fluorescence from the positive control (no inhibiting peptide) after subtracting the signal from the negative control (no biotinylated CLIP peptide). The experiment was performed in duplicate and was repeated once. The biological evaluation of the validation peptides was performed in a similar way in duplicate but at a separate occasion and later than for the designed peptide library. Since the reference peptide induced a weak response, the validation peptides were subjected to an additional binding study with a modified protocol.

Recombinant Aq molecules were captured by incubation at 4 °C overnight in a 96-well microtiter plate precoated with the mAb Y3-P and blocked with PBS containing 2% low fat milk. After washing, increasing concentrations of glycopeptides were added and incubated for 48 h at room temperature together with a fixed concentration of biotinylated CLIP peptide ("CLIPbio", 2.5 μM). CLIPbio-MHC class II complexes were quantified using the dissociation-enhanced lanthanide fluoroimmunoassay (DELFIA) kit system based on the time-resolved fluoroimmunoassay technique with europium-labeled streptavidin (Wallac, Turku) according to the manufacturer's instructions. The six validation peptides (**V1**−**V6**) and a reference peptide were tested at 0.8, 4, 20, 100, and 500 μM.

The full dose response curves for the second experiment were used to classify the validation peptides and to rank them in terms of relative binding strength.

Determination of T-Cell Hybridoma Responses. The response of each T-cell hybridoma line, i.e., the amount of IL-2 secreted following incubation with antigen-presenting spleen cells in the various concentrations of glycopeptides (**N-Lys** and **C-Lys**) was determined in a standard assay using the T-cell clone CTLL.[85] Briefly, 5×10^4 hybridoma T cells were cocultured with 5×10^5 syngeneic spleen cells and antigen in a volume of 200 μL in 96-well flat-bottom microtiter plates. After 24 h, 100 μL aliquots of the supernatants were removed and frozen to kill any transferred hybridoma T cells. IL-2 sensitive CTLL T cells (1×10^5/mL, 100 μL/well) were added to the thawed supernatant. The CTLL cultures were incubated for 24 h, after which they were pulsed with 1 μCi of [^{3}H]thymidine and incubated for an additional 15−18 h. The cells were harvested on glass fiber sheets in a Filtermate cell harvester (Packard Instruments, Meriden, CT), and the amount of radioactivity in them was determined using a Matrix 96 direct β counter (Packard). All experiments were performed in duplicate.

Acknowledgment. This work was funded by grants from the Swedish Research Council, the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, and the program "Glycoconjugates in Biological Systems" (GLIBS) sponsored by the Swedish Foundation for Strategic Research. We thank Dr. Fredrik Pettersson for the ranking algorithms.

Supporting Information Available: A complete list of the used molecular descriptors, method details of the preanalysis, detailed statistics of the PLS models, purity data of the peptides, and biological data of the validation peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Arnett, F. C.; Edworthy, S. M.; Bloch, D. A.; McShane, D. J.; Fries, J. F.; Cooper, N. S.; Healey, L. A.; Kaplan, S. R.; Liang, M. H.; Luthra, H. S.; Medsger, T. A.; Mitchell, D. M.; Neustadt, D. H.; Pinals, R. S.; Schaller, J. G.; Sharp, J. T.; Wilder, R. L.; Hunder, G. G. The American-Rheumatism-Association 1987 Revised Criteria for the Classification of Rheumatoid-Arthritis. *Arthritis Rheum.* **1988**, *31* (3), 315−324.

(2) Smolen, J. S.; Steiner, G. Therapeutic Strategies for Rheumatoid Arthritis. *Nat. Rev. Drug Discovery* **2003**, *2*, 473−488.

(3) de Haan, E. C.; Moret, E. E.; Wagenaar-Hilbers, J. M. J.; Liskamp, R. M. J.; Wauben, M. H. M. Possibilities and Limitations in the Rational Design of Modified Peptides for T cell Mediated Immunoterapy. *Mol. Immunol.* **2005**, *42*, 365−373.

(4) Albani, S.; Prakken, B. T Cell Epitope-Specific Immune Therapy for Rheumatic Diseases. *Arthritis Rheum.* **2006**, *54*, 19−25.

(5) Astorga, G. P.; Williams, R. C., Jr. Altered Reactivity in Mixed Lymphocyte Culture of Lymphocytes from Patients with Rheumatoid Arthritis. *Arthritis Rheum.* **1969**, *12* (6), 547−554.

(6) Stastny, P. Mixed Lymphocyte Cultures in Rheumatoid Arthritis. *J. Clin. Invest.* **1976**, *57* (5), 1148−1157.

(7) Gregersen, P.; Silver, J.; Winchester, R. The Shared Epitope Hypothesis. An Approach To Understanding the Molecular Genetics of Susceptibility to Rheumatoid Arthritis. *Arthritis Rheum.* **1987**, *30*, 1205−1213.

(8) Londei, M.; Savill, C. M.; Verhoef, A.; Brennan, F.; Leech, Z. A.; Duance, V.; Maini, R. N.; Feldmann, M. Persistence of Collagen Type-II-Specific T-Cell Clones in the Synovial-Membrane of a Patient with Rheumatoid-Arthritis. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (2), 636−640.

(9) Kim, H. Y.; Kim, W. U.; Cho, M. L.; Lee, S. K.; Youn, J.; Kim, S. I.; Yoo, W. H.; Park, J. H.; Min, J. K.; Lee, S. H.; Park, S. H.; Cho, C. S. Enhanced T Cell Proliferative Response to Type II Collagen and Synthetic Peptide CII (255−274) in Patients with Rheumatoid Arthritis. *Arthritis Rheum.* **1999**, *42* (10), 2085−2093.

(10) Bäcklund, J.; Carlsen, S.; Hoger, T.; Holm, B.; Fugger, L.; Kihlberg, J.; Burkhardt, H.; Holmdahl, R. Predominant Selection of T Cells Specific for the Glycosylated Collagen Type II Epitope (263−270) in Humanized Transgenic Mice and in Rheumatoid Arthritis. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (15), 9960−9965.

(11) Menzel, J.; Steffen, C.; Kolarz, G.; Eberl, R.; Frank, O.; Thumb, N. Demonstration of Antibodies to Collagen and Collagen−Anticollagen Immune Complexes in Rheumatoid Arthritis Synovial Fluids. *Ann. Rheum. Dis.* **1976**, *35*, 446−450.

(12) Cook, A.; Stockman, A.; Brand, C.; Tait, B.; Mackay, I.; Muirden, K.; Bernard, C.; Rowley, M. Antibodies to Type II Collagen and HLA Disease Susceptibility Markers in Rheumatoid Arthritis. *Arthritis Rheum.* **1999**, *42* (12), 2569−2576.

(13) Burkhardt, H.; Koller, T.; Engstrom, A.; Nandakumar, K.; Turnay, J.; Kraetsch, H.; Kalden, J.; Holmdahl, R. Epitope-Specific Recognition of Type II Collagen by Rheumatoid Arthritis Antibodies Is Shared with Recognition by Antibodies That Are Arthritogenic in Collagen-Induced Arthritis in the Mouse. *Arthritis Rheum.* **2002**, *46* (9), 2339−2348.

(14) Trentham, D. E.; Townes, A. S.; Kang, A. H. Autoimmunity to Type II Collagen an Experimental Model for Arthritis. *J. Exp. Med.* **1977**, *146*, 857−868.

(15) Courtenay, J. S.; Dallman, M. J.; Dayan, A. D.; Martin, A.; Mosedale, B. Immunization against Heterologous Type II Collagen Induces Arthritis in Mice. *Nature* **1980**, *283* (5748), 666−668.

(16) Wooley, P. H.; Luthra, H. S.; Stuart, J. M.; David, C. S. Type II Collagen-Induced Arthritis in Mice. I. Major Histocompatibility Complex (I Region) Linkage and Antibody Correlates. *J. Exp. Med.* **1981**, *154* (3), 688−700.

(17) Brunsberg, U.; Gustafsson, K.; Jansson, L.; Michaëlsson, E.; Ährlund-Richter, L.; Pettersson, S.; Mattsson, R.; Holmdahl, R. Expression of Transgenic Class II Ab Gene Confers Susceptibility to Collagen-Induced Arthritis. *Eur. J. Immunol.* **1994**, *24*, 1698−1702.

(18) Holm, L.; Kjellén, P.; Holmdahl, R.; Kihlberg, J. Identification of the Minimal Glycopeptide Core Recognized by T Cells in a Model for Rheumatoid Arthritis. *Bioorg. Med. Chem.* **2005**, *13*, 473−482.

(19) Rosloniec, E. F.; Whittington, K. B.; Brand, D. D.; Myers, L. K.; Stuart, J. M. Identification of MHC Class II and TCR Binding Residues in the Type II Collagen Immunodominant Determinant Mediating Collagen-Induced Arthritis. *Cell. Immunol.* **1996**, *172* (1), 21−28.

(20) Kjellen, P.; Brunsberg, U.; Broddefalk, J.; Hansen, B.; Vestberg, M.; Ivarsson, I.; Engstroem, A.; Svejgaard, A.; Kihlberg, J.; Fugger, L.; Holmdahl, R. The Structural Basis of MHC Control of Collagen-Induced Arthritis; Binding of the Immunodominant Type II Collagen 256−270 Glycopeptide to H-2Aq and H-2Ap Molecules. *Eur. J. Immunol.* **1998**, *28* (2), 755−767.

(21) Michaëlsson, E.; Malmström, V.; Reis, S.; Engström, Å.; Burkhardt, H.; Holmdahl, R. T Cell Recognition of Carbohydrates on Type II Collagen. *J. Exp. Med.* **1994**, *180*, 745−749.

(22) Michaëlsson, E.; Broddefalk, J.; Engström, Å.; Kihlberg, J.; Holmdahl, R. Antigen Processing and Presentation of a Naturally Glycosylated Protein Elicits Major Histocompatibility Complex Class II-Restricted, Carbohydrate-Specific T Cells. *Eur. J. Immunol.* **1996**, *26*, 1906−1910.

(23) Broddefalk, J.; Bäcklund, J.; Almqvist, F.; Johansson, M.; Holmdahl, R.; Kihlberg, J. T Cells Recognize a Glycopeptide Derived from Type II Collagen in a Model for Rheumatoid Arthritis. *J. Am. Chem. Soc.* **1998**, *120*, 7676−7683.

(24) Holm, B.; Backlund, J.; Recio, M. A. F.; Holmdahl, R.; Kihlberg, J. Glycopeptide Specificity of Helper T Cells Obtained in Mouse Models for Rheumatoid Arthritis. *ChemBioChem* **2002**, *3* (12), 1209−1222.

(25) Holm, L.; Bockermann, R.; Wellner, E.; Backlund, J.; Holmdahl, R.; Kihlberg, J. Side-Chain and Backbone Amide Bond Requirements for Glycopeptide Stimulation of T-Cells Obtained in a Mouse Model for Rheumatoid Arthritis. *Bioorg. Med. Chem.* **2006**, *14*, 5921−5932.

(26) Bäcklund, J.; Treschow, A.; Bockermann, R.; Holm, B.; Holm, L.; Issazadeh-Navikas, S.; Kihlberg, J.; Holmdahl, R. Glycosylation of Type II Collagen Is of Major Importance for T Cell Tolerance and Pathology in Collagen-Induced Arthritis. *Eur. J. Immunol.* **2002**, *32* (12), 3776−3784.

(27) Dzhambazov, B.; Nandakumar, K.; Kihlberg, J.; Fugger, L.; Holmdahl, R.; Vestberg, M. Therapeutic Vaccination of Active Arthritis with a Glycosylated Collagen Type II in Complex with MHC Class II Molecules. *J. Immunol.* **2006**, *176*, 1525−1533.

(28) Jane-wit, D.; Yu, M.; Edling, A. E.; Kataoka, S.; Johnson, J. M.; Stull, L. B.; Moravec, C. S.; Tuohy, V. K. A Novel Class II-Binding Motif Selects Peptides That Mediate Organ-Specific Autoimmune Disease in SWXJ, SJL/J, and SWR/J Mice. *J. Immunol.* **2002**, *169*, 6507−6514.

(29) Hammer, J.; Bono, E.; Gallazzi, F.; Belunis, C.; Nagy, Z.; Sinigaglia, F. Precise Prediction of Major Histocompatibility Complex Class II-Peptide Interaction Based on Peptide Side Chain Scanning. *J. Exp. Med.* **1994**, *180*, 2553−2358.

(30) Pinilla, C.; Appel, J. R.; Houghten, R. A. Rapid Identification of High Affinity Peptide Ligands Using Positional Scanning Synthetic Peptide Combinatorial Libraries. *BioTechniques* **1992**, *13*, 901−905.

(31) Sospedra, M.; Pinilla, C.; Martin, R. Use of Combinatorial Peptide Libraries for T-Cell Epitope Mapping. *Methods* **2003**, *29* (3), 236−247.

(32) Nino-Vasquez, J. J.; Allicorri, G.; Borras, E.; Wilson, D. B.; Valmori, D.; Simon, R.; Martin, R.; Pinilla, C. A Powerful Combination: The Use of Positional Scanning Libraries and Biometrical Analysis To Identify Cross-Reactive T-Cell Epitopes. *Mol. Immunol.* **2004**, *40* (14−15), 1063−1074.

(33) Hammer, J.; Takacs, B.; Sinigaglia, F. Identification of a Motif for HLA-DR1 Binding Peptides Using M13 Display Libraries. *J. Exp. Med.* **1992**, *176*, 1007−1013.

(34) Bolin, D. R.; Swain, A. L.; Sarabu, R.; Berthel, S. J.; Gillespie, P.; Huby, N. J. S.; Makofske, R.; Orzechowski, L.; Perrotta, A.; Toth, K.; Cooper, J. P.; Jiang, N.; Falcioni, F.; Campbell, R.; Cox, D.; Gaizband, D.; Belunis, C. J.; Vidovic, D.; Ito, K.; Crowther, R.; Kammlott, U.; Zhang, X.; Palermo, R.; Weber, D.; Guenot, J.; Nagy, Z.; Olson, G. L. Peptide and Peptide Mimetic Inhibitors of Antigen Presentation by HLA-DR Class II MHC Molecules. Design, Structure−Activity Relationships, and X-ray Crystal Structures. *J. Med. Chem.* **2000**, *43*, 2135−2148.

(35) Ettouati, L.; Salvi, J. P.; Trescol-Biemont, M. C.; Walchshofer, N.; Gerlier, D.; Rabourdin-Combe, C.; Paris, J. Substitution of Peptide Bond 53−54 of HEL(52−61) with an Ethylene Bond Rather Than Reduced Peptide Bond Is Tolerated by an MHC-II Restricted T Cell. *Pept. Res.* **1996**, *9* (5), 248−253.

(36) de Haan, E.; Wauben, M.; Grosfeld-Stulemeyer, M.; Moret, E. Structure-Based Design and Evaluation of MHC Class II Binding Peptides. *Biologicals* **2001**, *29* (3−4), 289−292.

(37) Linusson, A.; Wold, S.; Norden, B. Statistical Molecular Design of Peptoid Libraries. *Mol. Diversity* **1998**, *4* (2), 103−114.

(38) Larsson, A.; Johansson, S. M. C.; Pinkner, J. S.; Hultgren, S. J.; Almqvist, F.; Kihlberg, J.; Linusson, A. Multivariate Design Synthesis and Biological Evaluation of Peptide Inhibitors of FimC/FimH Protein−Protein Interactions in Uropathogenic Escherichia Coli. *J. Med. Chem.* **2005**, *48*, 935−945.

(39) Flower, D. R. Towards in Silico Prediction of Immunogenic Epitopes. *Trends Immunol.* **2003**, *24* (12), 667−674.
(40) Sette, A.; Buus, S.; Appella, A.; Smith, J. A.; Chesnut, R.; Miles, C.; Colon, S. M.; Grey, H. M. Prediction of Major Histocompatibility Complex Binding Regions of Protein Antigens by Sequence Pattern Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86* (9), 3296−3300.
(41) Rammensee, H.-G.; Bachmann, J.; Emmerich, N. P. N.; Bachor, O. A.; Stevanovic, S. SYFPEITHI: Database for MHC Ligands and Peptide Motifs. *Immunogenetics* **1999**, *50*, 213−219.
(42) Reche, P. A.; Glutting, J.-P.; Reinherz, E. L. Prediction of MHC Class I Binding Peptides Using Profile Motifs. *Hum. Immunol.* **2002**, *63*, 701−709.
(43) Reche, P. A.; Glutting, J.-P.; Zhang, H.; Reinherz, E. L. Enhancement of the RANKPEP Resource for the Prediction of Peptide Binding to MHC Molecules Using Profiles. *Immunogenetics* **2004**, *56*, 405−419.
(44) Burden, F. R.; Winkler, D. A. Predicitve Bayesian Neural Network Models of MHC Class II Peptide Binding. *J. Mol. Graphics Modell.* **2005**, *23*, 481−489.
(45) Bhasin, M.; Raghava, G. P. S. SVM Based Method for Predicting HLA-BRB1*0401 Binding Peptides in an Antigen Sequence. *Bioinformatics* **2004**, *20* (3), 421−423.
(46) Salomon, J.; Flower, D. R. Predicting Class II MHC-Peptides Binding: A Kernel Based Approach Using Similarity Scores. *BMC Bioinf.* **2006**, *7*, 501.
(47) Doytchinova, I. A.; Flower, D. R. Towards the in Silico Identification of Class II Restricted T-Cell Epitopes: A Partial Least Square Iterative Self-Consistent Algorithm for Affinity Prediction. *Bioinformatics* **2003**, *19* (17), 2263−2270.
(48) Hattotuwagama, C. K.; Toseland, C. P.; Guan, P.; Taylor, D. J.; Hemsley, S. L.; Doytchinova, I. A.; Flower, D. R. Toward Prediction of Class II Mouse Major Histocompatibility Complex Peptide Binding Affinity: In Silico Bioinformatic Evaluation Using Partial Least Squares, a Robust Multivariate Statistical Technique. *J. Chem. Inf. Model.* **2006**, *46*, 1491−1502.
(49) Wei, H.-Y.; Tsai, K.-C.; Lin, T.-H. Modeling Ligand−Receptor Interaction for Some MHC Class II HLA-DR4 Peptide Mimic Inhibitors Using Several Molecular Docking and 3D QSAR Techniques. *J. Chem. Inf. Model.* **2005**, *45*, 1343−1351.
(50) Baroni, M.; Clementi, S.; Cruciani, G.; Kettaneh-Wold, N.; Wold, S. D-Optimal Design in QSAR. *Quant. Struct.−Act. Relat.* **1993**, *12*, 225−231.
(51) Sjöström, M.; Eriksson, L. Applications of Statistical Experimental Design and PLS Modeling in QSAR. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, 1995; Vol. 2, pp 63−90.
(52) Linusson, A.; Gottfries, J.; Lindgren, F.; Wold, S. Statistical Molecular Design of Building Blocks for Combinatorial Chemistry. *J. Med. Chem.* **2000**, *43* (7), 1320−1328.
(53) Doytchinova, I. A.; Walshe, V.; Borrow, P.; Flower, D. R. Towards the Chemometric Dissection of Peptide-HLA-A*0201 Binding Affinity: Comparison of Local and Global QSAR Models. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 203−212.
(54) Peptide 18 was synthesised without the C-terminal lysine, but no solubility problems were experienced at the tested concentrations because of the peptide's high content of polar amino acids in the sequence.
(55) Sneath, P. H. A. Relations between Chemical Structure and Biological Activity in Peptides. *J. Theor. Biol.* **1966**, *12*, 157−195.
(56) Hellberg, S.; Sjöström, M.; Skagerberg, B.; Wold, S. Peptide Quantitative Structure−Activity Relationships, a Multivariate Approach. *J. Med. Chem.* **1987**, *30*, 1126−1135.
(57) Jonsson, J.; Eriksson, L.; Hellberg, S.; Sjöström, M.; Wold, S. Multivariate Parametrization of 55 Coded and Non-Coded Amino Acids. *Quant. Struct.−Act. Relat.* **1989**, *8*, 204−209.
(58) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J. Med. Chem.* **1998**, *41*, 2481−2491.
(59) Doytchinova, I. A.; Blythe, M. J.; Flower, D. R. Additive Method for the Prediction of Protein−Peptide Binding Affinity. Application to the MHC Class I Molecule HLA-A*0201. *J. Proteome Res.* **2002**, *1*, 263−272.
(60) Doytchinova, I. A.; Walshe, V. A.; Jones, N. A.; Gloster, S. E.; Borrow, P.; Flower, D. R. Coupling in Silico and in Vitro Analysis of Peptide−MHC Binding: A Bioinformatic Approach Enabling Prediction of Superbinding Peptides and Anchorless Epitopes. *J. Immunol.* **2004**, *172*, 7495−7502.
(61) Doytchinova, I. A.; Guan, P.; Flower, D. R. Quantitative Structure−Activity Relationship and the Prediction of MHC Supermotifs. *Methods* **2004**, *34*, 444−453.
(62) Guan, P.; Doytchinova, I. A.; Walshe, V.; Borrow, P.; Flower, D. R. Analysis of Peptide−Peptide Binding Using Amino Acid Descriptors: Prediction and Experimental Verification for Human Histocompatibility Complex HLA-A*0201. *J. Med. Chem.* **2005**, *48*, 7418−7425.
(63) Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2* (1−3), 37−52.
(64) Andersson, P. M.; Sjöström, M.; Wold, S.; Lundstedt, T. Comparison between Physicochemical and Calculated Molecular Descriptors. *J. Chemom.* **2000**, *14*, 629−642.
(65) Mei, H.; Liao, Z. H.; Zhou, Y.; Li, S. Z. A New Set of Amino Acid Descriptors and Its Application in Peptide QSARs. *Biopolymers* **2005**, *80* (6), 775−786.
(66) Mitchell, T. J. An Algorithm for the Construction of "D-Optimal" Experimental Designs. *Technometrics* **1974**, *16*, 203−210.
(67) Wold, H. Soft Modeling. The Basic Design and Some Extensions. In *Systems under Indirect Observation*; Jöreskog, K. G., Wold, H., Eds.; North-Holland Publishing Company: Amsterdam, 1982; Vol. II, pp 1−53.
(68) Wold, S. PLS for Multivariate Linear Modeling. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, 1995; Vol. 2, pp 195−218.
(69) Liu, W.; Meng, X.; Xu, Q.; Flower, D. R.; Li, T. Quantitative Prediction of Class I MHC Peptide Binding Affinity Using Support Vector Machine Regression (SVR) Models. *BMC Bioinf.* **2006**, *7*, 182.
(70) Michaëlsson, E.; Andersson, M.; Engström, Å.; Holmdahl, R. Identification of an Immunodominant Type-II Collagen Peptide Recognized by T Cells in H-2q Mice: Self Tolerance at the Level of Determinant Selection. *Eur. J. Immunol.* **1992**, *22*, 1819−1825.
(71) Andersson, E. C.; Hansen, B. E.; Jacobsen, H.; Madsen, L. S.; Andersen, C. B.; Engberg, J.; Rothbard, J. B.; McDevitt, G. S.; Malmström, V.; Holmdahl, R.; Svejgaard, A.; Fugger, L. Definition of MHC and T Cell Receptor Contacts in the HLA-DR4-Restricted Immunodominant Epitope in Type II Collagen and Characterization of Collagen-Induced Arthritis in HLA-DR4 and Human CD4 Transgenic Mice. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7574−7579.
(72) Rosloniec, E. F.; Brand, D. D.; Myers, L. K.; Esaki, Y.; Whittington, K. B.; Zaller, D. M.; Woods, A.; Stuart, J. M.; Kang, A. H. Induction of Autoimmune Arthritis in HLS-DR4 (DRB1*0401) Transgenic Mice by Immunization with Human and Bovine Type II Collagen. *J. Immunol.* **1998**, *160*, 2573−2578.
(73) Rosloniec, E. F.; Whittington, K. B.; Zaller, D. M.; Kang, A. H. HLA-DR1 (DRB1*0101) and DR4 (DRB1*0401) Use the Same Anchor Residues for Binding an Immunodominant Peptide Derived from Human Type II Collagen. *J. Immunol.* **2002**, *168*, 253−259.
(74) *UNIX Spartan 4.0*; Wavefunction, Inc. (18401 Von Karman Ave, Suite 370, Irvine, CA 92612).
(75) *Dragon*; Talete Srl (Via V. Pisani, 13-20124 Milano, Italy).
(76) Wold, H. Estimation of Principal Components and Related Models by Iterative Least Squares. In *Multivariate Analysis*; Krishnaiah, P. R., Ed.; Academic Press: New York, 1966; pp 391−420.
(77) Jackson, J. E. *A User's Guide to Principal Components*; Wiley: New York, 1991.
(78) *Simca-P 10.5*; Umetrics (Box 7960, S-907 19 Umeå, Sweden).
(79) DuMouchel, W.; Jones, B. A Simple Bayesian Modification of D-Optimal Designs To Reduce Dependence on an Assumed Model. *Technometrics* **1994**, *36* (1), 37−47.
(80) *Modde 6.0*; Umetrics (Box 7960, S-907 19 Umeå, Sweden).
(81) Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J. R. Stat. Soc.* **1974**, *B36*, 111−133.
(82) Trygg, J.; Wold, S. Orthogonal Projections to Latent Structures (O-PLS). *J. Chemom.* **2002**, *16*, 119−128.
(83) Petterson, F. Department of Bioinformatics, Wellcome Trust Centre, Oxford, U.K.; fredrikp@well.ox.ac.uk.
(84) Krchnak, V.; Vagner, J.; Lebl, M. Noninvasive Continous Monitoring of Solid-Phase Peptide Synthesis by Acid−Base Indicator. *Int. J. Pept. Protein Res.* **1988**, *32*, 415−416.
(85) Gillis, S.; Smith, K. A. Long Term Culture of Tumour-Specific Cytotoxic T Cells. *Nature* **1977**, *268*, 154−156.

JM061209B