Molecular mechanism of the susceptibility difference between HLA-B*27:02/04/05 and HLA-B*27:06/09 to ankylosing spondylitis: substitution analysis, MD simulation, QSAR modelling, and in vitro assay

X. Cheng, Y. Mei, X. Ji, Q. Xue and D. Chen

ABSTRACT
The human leukocyte antigen HLA-B27 is directly involved in the disease pathogenesis of ankylosing spondylitis (AS). HLA-B27 has a high degree of genetic polymorphism, with 105 currently known subtypes; the presence of aspartic acid at residue 116 (Asp116) has been found to play an essential role in AS susceptibility. Here, we systematically investigated the molecular mechanism of the susceptibility difference between the AS-associated subtypes HLA-B*27:02/04/05 and AS-unassociated subtypes HLA-B*27:06/09 to AS at sequence, structure, energetic and dynamic levels. In total seven variable residues were identified among the five studied HLA-B27 subtypes, in which Asp116 can be largely stabilized by a spatially vicinal, positively charged His114 through a salt bridge, while five other variable residues seem to have only a marginal effect on AS susceptibility. We also employed a quantitative structure–activity relationship approach to model the statistical correlation between peptide structure and affinity to HLA-B*27:05, a genetic ancestor of all other HLA-B27 subtypes and associated strongly with AS. The built regression predictor was verified rigorously through both internal cross-validation and external blind validation, and was then employed to identify potential HLA-B*27:05 binders from >20,000 cartilage-derived self-peptides. Subsequently, the binding potency of the top five antigenic peptides to HLA-B*27:05 was assayed in vitro using a FACS-based MHC stabilization experiment. Consequently, two (QRVGSDEFK and LRGAGTNEK) out of the five peptides were determined to have high affinity (BL50 = 5.5 and 15.8 nM, respectively) and, as expected, both of them possess positively charged Lys at the C-terminus.

Introduction
Ankylosing spondylitis (AS) is a common inflammatory joint disorder affecting the axial skeleton, peripheral large joints, certain enthuses and extra-articular sites, characterized by...
inflammatory back pain and associated with considerable disability and diminished quality of life in affected individuals [1]. The nurse practitioner (NP) and physician assistant (PA) have been thought to play an important role in the management of this inflammatory disease; NPs and PAs are ideally positioned to educate patients regarding treatment options and other important aspects of the disease process, and to monitor disease activity using a number of validated instruments. AS has a strong genetic predisposition with the HLA-B27 molecule; the heritability of susceptibility to AS has been estimated at ~97% from twin recurrence risk analysis [2]. Although only 1% of patients with HLA-B27 develop AS, 90–95% of patients with AS are positive for HLA-B27 [3]. An increased level of HLA-B27 expression in AS patients as compared with healthy HLA-B27-positive subjects has also been reported [4].

HLA-B27 is a class I surface antigen encoded by the B locus in the major histocompatibility complex (MHC) on chromosome 6 and presents antigenic peptides derived from self and non-self antigens to T cells [5]. HLA-B27 itself is a serologic specificity, which encompasses 132 currently known alleles that encode 105 different products (subtypes) [6]; studies of the association of these allelic variants have suggested significant differences in susceptibility to AS [7]. Previous studies have demonstrated that correct nursing can help in the treatment of AS patients with different HLA-B2 subtypes in an appropriate manner. Epidemiological investigations revealed a strong association between AS and the three HLA-B2 subtypes B*27:02, B*27:04 and B*27:05, whereas there are two subtypes, B*27:06 and B*27:09, that apparently lack an association with AS [8]. AS has also been reported in individuals with some other subtypes, but most are extremely rare and their relationship to the disease has not been defined [9].

Studies on transgenic rodents indicated that HLA-B27 is directly involved in disease pathogenesis rather than merely being a marker for a linked disease-associated gene [10]. The ‘arthritogenic peptide’ hypothesis proposes that particular properties of the peptide-binding groove of some HLA-B27 subtypes enable these molecules to display microbial peptides that exhibit molecular mimicry with specific arthritogenic self-peptides. This would allow the response of HLA-B27-restricted cross-reactive cytotoxic T-lymphocytes to a foreign peptide to be directed against self-peptides as well. The autoimmune reaction might then lead to chronic inflammation [9]. Strikingly, the property difference between the peptide-binding grooves of most HLA-B27 subtypes is modest, commonly with only few residue changes. For example, in contrast to B*27:05, B*27:09 is weakly or not associated to AS. However, these subtypes differ by only a single D116H residue substitution and share 79% of their peptide repertoire [11]. Crystallographic evidence has revealed that the substitution is located in the floor of the binding groove where it contacts the peptide C-terminus [12]. This difference reshapes the conformation of the F pocket, and this may explain why the B*27:09 molecule binds prevalently peptides with a hydrophobic C-terminus, while the B*27:05 molecule has a broader specificity [13]. The presence of the Asp116 residue has been found as the prerequisite for a subtype susceptible to AS; it may influence the peptide-binding specificity of HLA by changing the charge and size of the groove [14].

In order to elucidate the molecular mechanism and biological implications underlying the differential AS associations with HLA-B27 subtypes, we herein report a systematic analysis of representative subtypes as well as their intermolecular interactions with antigenic peptides. In the procedure, several subtypes of interest, including B*27:02, B*27:04, B*27:05, B*27:06 and B*27:09, were compared at sequence and structure levels, with attention paid to residue differences in their peptide-binding groove. The binding dynamics and interaction
affinity of these subtypes with peptide ligands were characterized in detail using molecular dynamics (MD) simulations and binding free energy analysis, respectively. Recently, we have successfully applied computational methods to study a specific molecular recognition in YAP WW1 domain–peptide interaction [15]. The correlation between the sequence pattern and peptide affinity to B*27:05 was also modelled via a quantitative structure–activity relationship (QSAR) approach, which was then used to infer potential self-peptide antigens involved in AS. Subsequently, the knowledge harvested from computational studies and predictions was applied to analysis and inference of the potent B*27:05 binders from cartilage-derived self-peptides, which was then substantiated with a quantitative T2 cell surface stabilization assay.

Materials and methods

**HLA-B27 subtypes and their association with AS**

HLA-B27 has a high degree of genetic polymorphism, with 105 currently known subtypes, named B*27:01 to B*27:106, encoded by 132 alleles (one designation “B*27:22” was withdrawn when subsequent studies revealed it to be identical to B*27:06) [16]. These variants are evolved from the most widespread subtype, B*27:05, and their distribution and association with AS risk differs across populations [17]. The most common subtypes are B*27:05 (Caucasians and American Indians), B*27:04 (Asians), and B*27:02 (Mediterranean populations), and these all show very strong association with AS. AS has also been reported to associate with other rare subtypes, including HLA-B*27:01, B*27:03, B*27:07, B*27:08, B*27:10, B*27:13, B*27:14, B*27:15, B*27:19 and B*27:25 [8]. On the other hand, the subtypes B*27:06 and B*27:09 have been the subject of interest because of their apparent lack of association with AS. Despite its high prevalence in Southeast Asia, B*27:06 has not been described in AS patients in this area. B*27:09 comprises a significant fraction of B27 alleles in Sardinia and Italy, but research has failed to identify a single AS patient with this subtype [18]. Here, the five representative subtypes, i.e. three AS-associated B*27:02, B*27:04 and B*27:05 as well as two AS-unassociated B*27:06 and B*27:09, were investigated in this study.

**The three-dimensional structures of HLA-B27 subtypes**

Four out of the five investigated HLA-B27 subtypes, namely B*27:04, B*27:05, B*27:06 and B*27:09, are currently available with their complex crystal structures and peptide ligands in the PDB database [19] under accession ID 5DEF, 3BP4, 5DEG and 3CZF, respectively. The crystal structure of B*27:02 has not been solved to date, and thus we herein employed a virtual mutagenesis strategy to model its atomic structure. Considering that B*27:05 is the molecular origin of all HLA-B27 subtypes and there are only three residue differences between the B*27:02 and B*27:05 (Asn77Asp, Ile80Thr and Ala81Leu), the B*27:05 crystal structure was used as a template to model the structure of B*27:02. First, the side chains of the three residues concerned were manually removed from the B*27:05 crystal structure, and then new side chains were added automatically to these residues by using a graph theory-based SCWRL program [20]. The SCWRL algorithm has been demonstrated to perform very well as compared with other methods including SCATD, IRECS, SPDBV and SCit in a HLA side-chain benchmark comparison [21]. In this way, the structure of B*27:02 can be readily modelled.
Quantitative structure–activity relationship

Peptide dataset
A dataset of experimental HLA-B*27:05 peptide-binding affinities was recently released by Walshe et al. [22]. Here, we only selected 9-mer peptides and, consequently, a total of 218 peptides were compiled. The binding affinities of these peptides to HLA-B*27:05 were determined by T2 cell stabilization assay and quantified as half-maximal binding level (BL_{50}), which is the peptide concentration yielding the half-maximal fluorescence index (FI) of the reference peptide. The affinities were converted to pBL_{50} (−logBL_{50}) for subsequent QSAR modelling. The 218 peptide samples are tabulated in Supplementary Materials, Table S1 (available via the Supplementary Content tab on the article’s online page).

Structural parameterization
Amino acid descriptors (AADs) have been widely used to characterize, model and predict the biological activity of diverse peptides [23]. Here, five kinds of AADs [24−28] listed in Table 1 were separately employed to carry out QSAR modelling of antigenic peptides. The sequence pattern of these peptides can be parameterized with each of the five AADs. For example, the z-scale consists of three components; when we use the z-scale to characterize a 9-mer antigenic peptide, in total 3 × 9 = 27 variables representing the peptide are generated.

Regression modelling
The correlation between the descriptor variables and affinity of peptides can be modelled using support vector regression (SVR) with RBF kernel function [29]. SVR is a machine learning algorithm based on statistical learning theory, which aims at structural risk minimization rather than the traditional empirical risk minimization and is especially suitable for small-sample, high-dimensional and strongly collinear problems. Here, SVR was carried out using the LibSVM program [30].

Model validation
The model validation was performed by both internal 5-fold cross-validation and external blind validation. In 5-fold cross-validation, a peptide set is evenly divided into five subsets, and this procedure is repeated five times. Each time, one of the five subsets is used as the test set and the other four subsets are put together to form a training set. Then the average statistics across all five trials are computed. In addition to internal validation, Golbraikh and Tropsha highlighted the importance of external blind validation in building predictive QSAR models [31]. Thus, we randomly split a peptide set into two subsets with ~3:1 ratio; the larger one (150 peptides) is a training set used to build QSAR models, while the smaller one (68 peptides) is a test set for blind validation of the built models.

Table 1. The five kinds of amino acid descriptors.

| Amino acid descriptor | \(n\)\(^a\) | Class\(^b\)                      |
|-----------------------|-------|---------------------------------|
| z-scale [24]          | 3     | Physicochemical property        |
| MS-Whim [25]          | 3     | Three-dimensional property      |
| VhSE [26]             | 8     | Physicochemical property        |
| T-scale [27]          | 5     | Topological property            |
| DPPS [28]             | 10    | Physicochemical property        |

\(^a\)\(n\), the number of vector components in amino acid descriptors.

\(^b\)The class of original property parameters used to derive amino acid descriptor.
Model statistics
The performance of QSAR modelling and validation were characterized using the coefficients of determination of fitting on training set ($r^2$), of 5-fold cross-validation on training set ($q^2$) and of prediction on test set ($p^2$) [32]:

$$r^2 \text{ or } q^2 \text{ or } p^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

where $n$ is the number of peptides in training or test set, $y_i$ is the experimental affinity (pBL50) of peptide $i$, $\bar{y}$ is the mean value of $y_i$ over all peptides in training or test set, and $\hat{y}$ is the calculated affinity for peptide $i$ by fitting (fit), 5-fold cross-validation (cv) or prediction (prd).

Molecular dynamics simulation and binding free energy analysis
MD simulations of HLA–peptide complexes were carried out using AMBER10 force field [33]. A truncated octahedral box of TIP3P waters [34] was added with a 10 Å buffer around the complex. Counter-ions of Na+ were added to make the simulated system electroneutral. The simulated system was heated from 0 to 300 K over 300 ps, followed by a constant temperature equilibration at 300 K for 500 ps. Finally, 50-ns MD production simulations were performed in an isothermal isobaric ensemble with periodic boundary conditions. The time step of simulations was set to 2 fs. The particle mesh Ewald method [35] was employed to calculate long-range electrostatic interactions, and a cut-off distance of 10 Å was used to calculate the short-range electrostatics and van der Waals interactions. The SHAKE method [36] was employed to constrain all covalent bonds involving hydrogen atoms.

The HLA–peptide binding free energy was calculated using molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) method [37] based on 100 snapshots of their complex structure extracted evenly from MD equilibrium trajectory:

$$\Delta G = \Delta E_{\text{int}} + \Delta D_{\text{dilv}} - T \Delta S$$

where $\Delta E_{\text{int}}$ is the intermolecular interaction energy between the HLA and peptide, which was calculated using force field approach, and $\Delta D_{\text{dilv}}$ is the desolvation free energy due to peptide binding, which can be computed by numerical solution of the nonlinear Poisson–Boltzmann equation and surface area model. To consider entropy penalty upon peptide binding, normal mode analysis (NMA) was employed to estimate the vibrational component of the entropy using the nmode program in the AMBER11 package [38]. Due to the high computational demand, only 25 snapshots were used in the NMA analysis [39]. Frequencies of vibrational modes were computed at 300 K for the snapshots using a harmonic approximation of the energies.

FACS-based MHC stabilization assay
Peptides were synthesized using standard 9-fluorenyl methoxycarbonyl (Fmoc) solid phase chemistry, and then purified by RP-HPLC C18 columns and confirmed by mass spectrometry and amino acid analysis. The binding of peptides to B*27:05 was measured using a FACS-based MHC stabilization assay [22, 40]. Briefly, T2 cells were incubated in 96-well flat-bottom plates at $2 \times 10^5$ cells per well in a 200 μL volume of AIM V medium with human-microglobulin
at a final concentration of 100 nM with and without peptides at concentrations between 200 and 0.04 μM for 16 h at 37°C. Cells were then washed and surface levels of B*27:05 were assessed by staining with a mouse anti-human HLA-B27 FITC Ab. Cells were fixed at 4°C in 4% paraformaldehyde and analysed on a FACS Calibur. Results are expressed as FI values, which were converted to the test mean fluorescence intensity (MFI) minus the no peptide isotype control MFI divided by the no peptide HLA-B27-stained control MFI minus the no peptide isotype control MFI [40]. The half-maximal binding level (BL50), which is the peptide concentration yielding the half-maximal FI of the reference peptide in each assay, was calculated as pBL50 (log1/BL50). Here, the known B*27:05 binder GRLTKHHTFF [22] was used as a reference peptide and each sample was tested in duplicate.

Results and discussion

Comparison of HLA-B27 subtypes at sequence and structure levels

The primary sequences of B*27:04, B*27:04, B*27:05, B*27:06 and B*27:09 were retrieved from the IMGT/HLA database [41] (see Supplementary Materials, Table S2, available online). The five investigated subtypes are all 276 amino acids long and, according to multiple sequence alignment conducted by MView server [42], exhibit a very high conservation, with sequence identity >97% (Figure 1). In fact, only very few residues are different between two

Figure 1. Multiple sequence alignment of B*27:02, B*27:04, B*27:05, B*27:06 and B*27:09. The alignment was carried out using MView program [42].
or more of them, including those at 77, 80, 81, 114, 116, 152 and 211. As can be seen in Figure 2, there is no consistent difference between the AS-associated and AS-unassociated subtypes, except residue 116, which is always occupied by a negatively charged Asp in three AS-associated subtypes, whereas the positively charged His or bulky Tyr can be found in other two AS-unassociated subtypes. In addition, the His is present at residue 114 of the three AS-associated subtypes, but the same amino acid can also be found in the AS-unassociated subtype B*27:09, suggesting that His114 is not sufficient for an AS-associated subtype. Similarly, the Ser77 is shared by two AS-unassociated subtypes, but also presents in the AS-associated subtype B*27:04. Moreover, there is no significant amino acid preference observed in residues 80, 81, 152 and 211; all of them share overlapping amino acid types between AS-associated and AS-unassociated subtypes, indicating that these four residues may not be the key determinant of AS susceptibility.

Residue 116 is particularly interesting because it is the only difference between the AS-associated subtype B*27:05 and AS-unassociated subtype B*27:09 [17]. The His116Asp or Tyr116Asp substitution can be found in all the six pairs changing from two AS-unassociated subtypes to three AS-associated subtypes (Table 2), suggesting that the aspartic acid (Asp) present at residue 116 is essential for an AS-associated subtype, which would shift the

| Figure 2. The amino acid difference at seven variable residues among the five investigated HLA-B27 subtypes. |
| --- |
| Table 2. The residue substitution matrix between the five investigated HLA-B27 subtypes. |
| **Subtype** | **B*27:02** | **B*27:04** | **B*27:05** | **B*27:06** | **B*27:09** |
| **AS-associated subtype** | | | | | |
| B*27:02 | Ser77Asn | Thr80Leu | Leu81Ala | Glu152Val | Asp77Ser |
| B*27:04 | Thr80Leu | Leu81Ala | Asp114His | Tyr116Asp | Val152Glu |
| B*27:05 | Asp77Asn | Thr80Leu | Leu81Ala | Asp114His | Tyr116Asp |
| **AS-unassociated subtype** | | | | | |
| B*27:06 | Ser77Asn | Thr80Leu | Leu81Ala | Asp114His | Tyr116Asp |
| B*27:09 | Thr80Leu | Leu81Ala | Asp77Ser | His116Asp | Val152Glu |
| B*27:04 | Asp77Ser | His116Asp | Val152Glu | Ala211Gly |

*The blocks with gray background do not alter AS susceptibility.*
peptide repertoire between the two subtypes. However, previous research found that B*27:05 and B*27:09 shared ~80% of their peptide repertoire [11], implying that peptides within a relatively small fraction of the shared repertoire influence susceptibility to AS [43]. In addition, the Asp114His substitution from AS-unassociated to AS-associated subtypes is also frequently observed from the residue substitution matrix (Table 2), which is thus thought to be an assistant factor in determining the AS susceptibility of a subtype.

The location of seven variable residues were then mapped on the complex crystal structure of B*27:05 with an antigenic peptide (PDB: 3BP4) (Figure 3). It is seen that six out of the seven residues are located in the peptide-binding groove of B*27:05, while only one (residue 211) is at the α3 domain of HLA molecule, far away from the groove. Therefore, the variable residue 211 is considered not to be involved in AS susceptibility. More specifically, the six groove-locating residues are all concentrated in a small region close to the C-terminus of the antigenic peptide, and thus changes in these residues may primarily influence binding preference among the C-terminal amino acid types of peptide ligands. However, it is evident that not all of the six variable residues can contribute substantially to AS susceptibility between different HLA-B27 subtypes. In fact, residues Thr80, Leu81 and Val152 do not contact the antigenic peptide directly, indicating that they can only have modest effects on peptide binding. In addition, although residues Asp77, His114 and Asp116 can interact directly with the peptide ligand, only Asp116 touches the anchor position at the peptide C-terminus. However, from the crystal structure it seems that the positively charged His114 is able to stabilize the negatively charged Asp116 through a salt bridge between them. Consequently, residue 116 is thought to influence peptide binding directly and significantly, in which residue 114 may play an indirect role to stabilize residue 116. Thus, both are suggested to co-determine the AS susceptibility of HLA-B27 subtypes.

Figure 3. The location of seven variable residues in the complex crystal structure of B*27:05 with an antigenic peptide (PDB: 3BP4).
Figure 4. The single-point mutation energy profiles from three AS-associated subtypes B*27:02, B*27:04, and B*27:05 to two AS-unassociated subtypes B*27:06 and B*27:09.
The effect of variable residues on peptide binding

Next, we examined the role of variable residues in determining AS susceptibility. There are three AS-associated subtypes, B*27:02, B*27:04 and B*27:05, and two AS-unassociated subtypes, B*27:06 and B*27:09. The residue differences between any two of the five subtypes are tabulated in Table 1. Here, we only considered the mutation of three AS-associated to two AS-unassociated subtypes, resulting in $3 \times 2 = 6$ subtype mutation pairs, namely, B*27:02→B*27:06, B*27:02→B*27:09, B*27:04→B*27:06, B*27:04→B*27:09, B*27:05→B*27:06 and B*27:05→B*27:09. A 9-mer antigenic peptide GRIDKPILK from the 60S ribosomal protein L8 was adopted, which has been shown to bind specifically for AS-associated over AS-unassociated subtypes [11, 44]. The single-point mutation profiles of the peptide response to the six pairs are shown in Figure 4. For example, the B*27:02→B*27:09 pair includes four single-point residue mutations, i.e. Asn77→Asp, Ile80→Thr, Ala81→Leu and His116→Asp. The binding free energies $\Delta G$ of peptide to wide-type B*27:02 as well as to four single-point mutants B*27:02(Asn77Asp), B*27:02(Ile80Thr), B*27:02(Ala81Leu) and B*27:02(His116Asp) were separately calculated using MD simulation and MM/PBSA analysis, and thus the single-point mutation energy $\Delta \Delta G$ upon each mutation can be obtained by $\Delta \Delta G = \Delta G - \Delta G_{\text{wild type}}$. As can be seen in Figure 4, the six pairs exhibit a very similar profile of single-point mutation energies $\Delta \Delta G$, that is, the two mutations at residues 114 and 116 are significantly unfavourable to peptide binding ($\Delta \Delta G > 1.5$ kcal/mol), while the other five mutations have only a moderate or modest effect on the binding ($-1$ kcal/mol $< \Delta \Delta G < 1$ kcal/mol).

Residue 116, as expected, plays a critical role in the selective interaction of the antigenic peptide between AS-associated over AS-unassociated subtypes; mutation of the negatively charged Asp in AS-associated subtypes to the positively charged His or aromatic, bulky Tyr in AS-unassociated subtypes should cause a large affinity loss for the peptide ($\Delta \Delta G > 2$ kcal/mol). As can be seen in Figure 3, residue 116 is very close to the positively charged Lys at the peptide C-terminus, thus forming a geometrically satisfactory salt bridge between them. Mutation at this residue would break this strong electrostatic interaction and thus greatly reduce the peptide affinity. Thus, it is readily concluded that those self-peptides with high immunogenicity in AS would generally prefer a positively charged residue at their C-terminus, which could confer strong electrostatic potential to the negatively charged Asp116 common in AS-associated subtypes. This finding is consistent with the B*27:05-bound repertoire eluted from HLA-B27 transfectant C1R cells, where most B*27:05 specific peptides possess a positively charged C-terminal residue such as Lys and Arg [11].

The His114Asp mutation appears also to cause a significantly unfavourable effect on peptide binding, with $\Delta \Delta G > 1.5$ kcal/mol. However, structural analysis of the B*27:05–peptide complex revealed that the His114 residue can only interact moderately with the hydrophobic side chain of the peptide Pro6 residue, but does not form substantial contacts with other peptide residues, suggesting that the His114Asp mutation may not influence peptide binding directly. In fact, the His114 is present in all three AS-associated subtypes, but can also be found in an AS-unassociated subtype B*27:09. In AS-associated subtypes, the positively charged His114 is spatially vicinal to, and thus electrostatically stabilizes, the negatively charged Asp116 through a well-defined salt bridge between them. Decomposition of B*27:05–peptide binding free energy supported this supposition; the interaction energies of His114 and Asp116 with the peptide ligand were calculated as $-2.78$ and $-8.42$ kcal/mol, respectively. Evidently, the peptide can bind much more tightly to Asp116 as compared with
His114. However, the Asp116–peptide interaction would be disrupted considerably by Asp114 when substituting His114 by a positively charged Asp, with interaction energy reducing from –8.42 to –3.60 kcal/mol. Thus, the His114Asp mutation impairs B*27:05–peptide binding by primarily destabilizing Asp116 rather than breaking its direct interaction with the peptide.

From the single-point mutation energy profiles (Figure 4) it seems that the mutations at residues 77, 80 and 81 do not influence peptide binding substantially. This is expected because the three residues cannot form direct interaction with the peptide and/or the mutations cannot alter the intermolecular interaction between HLA and peptide. In fact, clinical mutations at the three sites are mostly conserved and homogeneous, and do not have a significant effect on peptide binding. In this respect, the mutations at residues 77, 80 and 81 are not essential in AS susceptibility, although they are also common in the clinic.

Here, we also compared the MD-equilibrated conformations of antigenic peptide GRIDKPILK in complex with different subtypes. As can be seen in Figure 5(a), the three peptide conformations bound to AS-associated subtypes B*27:02, B*27:04 and B*27:05 are well consistent with the standard HLA–peptide complex structure, where the two ends of the peptide bind tightly to the HLA groove, while pointing the central region out of the groove to form a protruding shape. However, the peptide-bound conformations with AS-unassociated subtypes B*27:06 and B*27:09 seem differ to the standard mode, that is, the C-terminus of the two peptide conformations exhibits a slight motion from the standard mode (Figure 5(b)), thus largely impairing the peptide affinity to AS-unassociated subtypes (the C-terminal 9th position is a anchor residue that dominates HLA–peptide binding [45]).

**QSAR modelling and in vitro assay of B*27:05 binding peptides**

HLA–B*27:05 is deemed to be the genetic ancestor from which the other B27 alleles have evolved, and has been found to strongly associate with AS [9]. Here, a total of 218 9-mer
antigenic peptides with experimentally measured affinity (pBL\textsubscript{50}) to the subtype were retrieved from a recent report [22] (see Supporting Information Table S1, available online). These peptides were split into a training set and a test set consisting of 150 and 68 peptides, respectively. The peptide sequences were separately characterized using the five kinds of AADs listed in Table 1. Here, five QSAR models were developed separately using five AADs based on the training set, and they were then used to perform extrapolation on the test set; the resulting statistics are tabulated in Table 3. In the procedure the RBF kernel parameters of SVR models such as $\varepsilon$-insensitive, penalty factor and radial width were optimized through a grid search [46].

It is seen that all five models can perform well on the training set with internal fitting $r^2$ and 5-fold cross-validation $q^2$ larger than 0.65, suggesting a good fitting ability and internal stability of these built QSAR models. However, the predictive powers of the five models differ considerably, with $p^2$ ranging between 0.538 and 0.701, although all of them satisfy the criterion $p^2 > 0.5$ for a predictive QSAR model as recommended by Golbraikh and Tropsha [31]. On both training set and test set the divided physicochemical property scores (DPPS) exhibits optimal performance in all the five investigated descriptors with $r^2$, $q^2$ and $p^2$ of 0.829, 0.746 and 0.701, respectively. The DPPS was derived by principal component analysis.

Table 3. Statistics of five QSAR models built separately with five kinds of amino acid descriptors.

| Model   | Amino acid descriptor | Training set (150 peptides) | Training set (68 peptides) |
|---------|-----------------------|-----------------------------|-----------------------------|
|         |                       | $r^2$ | $q^2$ | $p^2$ | $r^2$ | $q^2$ | $p^2$ |
| Model 1 | z-scale               | 0.764 | 0.692 | 0.621 |
| Model 2 | MS-Whim               | 0.739 | 0.652 | 0.538 |
| Model 3 | VHSE                  | 0.803 | 0.734 | 0.660 |
| Model 4 | T-scale               | 0.746 | 0.662 | 0.589 |
| Model 5 | DPPS                  | 0.829 | 0.746 | 0.701 |

Figure 6. Scatter plot of calculated against experimental affinities for the 150 training peptides and 68 test peptides using the DPPS-based QSAR model built with DPPS descriptor.
of 23 electronic properties, 37 steric properties, 54 hydrophobic properties and five hydrogen bond properties of 20 amino acids, and has been successfully applied to QSAR modelling of A*02:01–peptide interaction [27]. The scatter plot of calculated against experimental affinities for the 150 training peptides and 68 test peptides using the DPPS-based QSAR model built with DPPS descriptor is shown in Figure 6. As can be seen, most peptide scatters distribute close to the expected slope and only very few slight outliers can be observed in
the plot. In particular, the B*27:05 nonbinders (experimental pBL50 = 0) are concentrated into a small region separated from other binders, indicating that the model is able to reliably classify active and inactive peptides.

The intercorrelation among the 90 variables generated by using the DPPS to characterize the 218 B*27:05 binding 9-mer peptides was examined. As can be seen in Figure 7, the intercorrelation of most variable pairs is modest or moderate with coefficient $R < 0.5$, and only very few pairs exhibit a relatively strong intercorrelation with $R > 0.5$, indicating a low multi-collinearity involved in the DPPS-based QSAR model. Next, the variable importance of the DPPS-based QSAR model was also investigated using a variable leave-one-out strategy [47]. Based on the training set we first adopted all 90 DPPS variables to build a QSAR model and then carried out extrapolation on the test set, resulting in a coefficient of determination of prediction ($p_{all}^2 = 0.701$). The procedure was then repeated 90 times, but each time only 89 variables were employed to develop the model and to perform extrapolation, resulting in a coefficient of determination of prediction ($p_{all-i}^2$), where $i$ is the missing variable. The relative importance of $i$th variable can be expressed as $\Delta p_i^2 = p_{all}^2 - p_{all-i}^2$. $\Delta p_i^2 < 0$ and $\Delta p_i^2 > 0$ represent $i$th variable significant and insignificant, respectively. As can be seen in Figure 8 (the intercorrelation coefficients $R$ are tabulated in Supplementary Materials, Table S3, available online), most of these variables are significant; removal of any one would reduce the predictive power of the DPPS-based QSAR model, although the importance of each single variable seems to be very modest, with the $\Delta p_i^2$ less than −0.05. Specifically, there are two variable blocks that appear to be relatively important as compared with others, which correspond separately to anchor residues 2 and 9 of HLA-binding peptides. This well reflects the immunobiological fact that the two anchor residues play a crucial role in HLA–peptide binding [45].

Cartilage antigens have been discussed as one of the most promising potential candidates for the immune response in AS. Previously, Atagunduz et al. derived more than 20,000 9-mer self-peptides from 18 human cartilage proteins as potential antigen targets of HLA-B27 molecules [48]. Here, we used the DPPS descriptor to characterize these peptide candidates and then employed the built QSAR model to predict their affinity to B*27:05. Consequently, most peptides were predicted as nonbinders or weak binders (predicted pBL50 < 5), while only 125 peptides exhibited high theoretical affinity (predicted pBL50 > 8) towards B*27:05. Here, the binding potency of the top five peptides was measured using a FACS-based MHC stabilization assay. As can be seen in Table 4, the calculation seems to overestimate peptide-binding affinity as compared with experimental results. This is acceptable when considering that some other factors, such as peptide flexibility and interactive effect between peptide residues, are not considered in the DPPS descriptor, which may cause systematic bias to the calculation. Consequently, four out of the five tested peptides exhibited high or

| Peptide   | Cartilage protein | Uniprot | Experimental BL50 (nM)a | Experimental pBL50 | Predicted pBL50 |
|-----------|-------------------|---------|--------------------------|--------------------|-----------------|
| VRPEFELVK | matrix protein    | P21941  | 72.4                     | 7.14               | 9.67            |
| QRVGSDEFK | type V collagen   | A8TX70  | 5.5                      | 8.26               | 9.55            |
| TKMGSPLPO | fibromodulin      | Q06828  | 147.9                    | 6.83               | 9.52            |
| GGLTGTGAR | type II collagen  | P08123  | 3548.1                   | 7.80               | 9.46            |
| LRGAGTNEK | anchorin          | P08758  | 15.8                     | 5.45               | 9.45            |

*aAverage values from duplicate.
moderate affinity to B*27:05 with experimental $BL_{50} < 1000$ nM, while only one peptide (GANGLTGAR) from type II collagen was determined as a modest binder of B*27:05 (experimental $BL_{50} > 3000$ nM). Two peptides (QRVGSDEFK and LRGAGTNEK) were measured to have high affinity (experimental $BL_{50} < 20$ nM), both of which possess positively charged Lys at residue 9; this is well consistent with the basic binding motif of the AS-associated subtypes proposed above. Another two (VRPEFELVK and TKMPGPLPQ) also showed moderate potency (experimental $BL_{50} < 200$ nM). These self-peptides could be the potential candidates for B*27:05 antigens implicated in AS.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Affiliated Sixth People’s Hospital Shanghai Jiao Tong University.

References

[1] J. Sieper, J. Braun, M. Rudwaleit, A. Boonen, and A. Zink, Ankylosing spondylitis: An overview, Ann. Rheum. Dis. 61 (2002), pp. iii8–iii18.
[2] M.A. Brown, L.G. Kennedy, A.J. MacGregor, C. Darke, E. Duncan, J.L. Shatford, A. Taylor, A. Calin, and P. Wordsworth, Susceptibility to ankylosing spondylitis in twins: The role of genes, HLA, and the environment, Arthritis Rheum. 40 (1997), pp. 1823–1828.
[3] A.P. Malaviya and A.J. Ostor, Early diagnosis crucial in ankylosing spondylitis, Practitioner 255 (2011), pp. 21–24.
[4] A. Cauli, G. Dessole, M.T. Fiorillo, A. Vacca, A. Mameli, P. Bitti, G. Passiu, R. Sorrentino, and A. Mathieu, Increased level of HLA-B27 expression in ankylosing spondylitis patients compared with healthy HLA-B27-positive subjects: A possible further susceptibility factor for the development of disease, Rheumatology 41 (2002), pp. 1375–1379.
[5] P. Bowness, HLA-B27, Annu. Rev. Immunol. 33 (2015), pp. 29–48.
[6] J.D. Reveille, Major histocompatibility genes and ankylosing spondylitis, Best Pract. Res. Clin. Rheumatol. 20 (2006), pp. 601–609.
[7] M.A. Brown, B.P. Wordsworth, and J.D. Reveille, Genetics of ankylosing spondylitis, Clin. Exp. Rheumatol. 20 (2002), pp. 543–549.
[8] J.D. Taurog, The mystery of HLA-B27: If it isn’t one thing, it’s another, Arthritis Rheum. 56 (2007), pp. 2478–2481.
[9] N.J. Sheehan, HLA-B27: What’s new?, Rheumatology 49 (2010), pp. 621–631.
[10] R.E. Hammer, S.D. Maika, J.A. Richardson, J.P. Tang, and J.D. Taurog, Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: An animal model of HLA-B27-associated human disorders, Cell 63 (1990), pp. 1099–1112.
[11] M. Ramos, A. Paradela, M. Vazquez, A. Marina, J. Vazquez, and J.A. Lopez de Castro, Differential association of HLA-B*2705 and B*2709 to ankylosing spondylitis correlates with limited peptide subsets but not with altered cell surface stability, J. Biol. Chem. 277 (2002), pp. 28749–28756.
[12] M. Hülsmeyer, M.T. Fiorillo, F. Bettosini, R. Sorrentino, W. Saenger, A. Ziegler, and B. Uchanska-Ziegler, Dual, HLA-B27 subtype-dependent conformation of a self-peptide, J. Exp. Med. 199 (2004), pp. 271–281.
[13] M.T. Fiorillo, G. Greco, M. Maragno, I. Potolicchio, A. Monizio, M.L. Dupuis, and R. Sorrentino, The naturally occurring polymorphism Asp116→His116, differentiating the ankylosing spondylitis-associated HLA-B*2705 from the non-associated HLA-B*2709 subtype, influences peptide-specific CD8 T cell recognition, Eur. J. Immunol. 28 (1998), pp. 2508–2516.
[14] M. D’Amato, M.T. Fiorillo, C. Carcassi, A. Mathieu, A. Zuccarelli, P.P. Bitti, R. Tosi, and R. Sorrentino, *Relevance of residue 116 of HLA-B27 in determining susceptibility to ankylosing spondylitis*, Eur. J. Immunol. 25 (1995), pp. 3199–3201.

[15] D. Chen, S. Liu, W. Zhang, and L. Sun, *Rational design of YAP WW1 domain-binding peptides to target TGFβ/BMP/Smad-YAP interaction in heterotopic ossification*, J. Pept. Sci. 21 (2015), pp. 826–832.

[16] M.A. Khan, *Polymorphism of HLA-B27: 105 subtypes currently known*, Curr. Rheumatol. Rep. 15 (2013), p. 362.

[17] B. Galocha and J.A. de Castro, *Folding of HLA-B27 subtypes is determined by the global effect of polymorphic residues and shows incomplete correspondence to ankylosing spondylitis*, Arthritis Rheum. 58 (2008), pp. 401–412.

[18] M. D’Amato, M.T. Fiorillo, C. Carcassi, A. Mathieu, A. Zuccarelli, P.P. Bitti, R. Tosi, and R. Sorrentino, *Relevance of residue 116 of HLA-B27 in determining susceptibility to ankylosing spondylitis*, Eur. J. Immunol. 25 (1995), pp. 3199–3201.

[19] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The protein data bank*, Nucl. Acids Res. 28 (2000), pp. 235–242.

[20] A.A. Canutescu, A.A. Shelenkov, and R.L. Dunbrack Jr, *A graph-theory algorithm for rapid protein side-chain prediction*, Protein Sci. 12 (2003), pp. 2001–2014.

[21] B. Knapp, U. Omasits, and W. Schreiner, *Side chain substitution benchmark for peptide/MHC interaction*, Protein Sci. 17 (2008), pp. 977–982.

[22] V.A. Walshe, C.K. Hattotuwagama, I.A. Doytchinova, and D.R. Flower, *A dataset of experimental HLA-B*2705 peptide binding affinities*, Dataset Pap. Sci. 2014 (2014), p. 914684.

[23] P. Zhou, F. Tian, Y. Wu, Z. Li, and Z. Shang, *Quantitative sequence-activity model (QSAM): Applying QSAR strategy to model and predict bioactivity and function of peptides, proteins and nucleic acids*, Curr. Comput. Aided Drug Des. 2014 (2014), p. 914684.

[24] S. Hellberg, M. Sjoestroem, B. Skagerberg, and S. Wold, *Peptide quantitative structure-activity relationships, a multivariate approach*, J. Med. Chem. 30 (1987), pp. 1126–1135.

[25] A. Zaliani and E. Gancia, *MS-WHIM scores for amino acids, a new 3D-description for peptide QSAR and QSPR studies*, J. Chem. Inf. Comput. Sci. 39 (1999), pp. 525–533.

[26] H. Mei, Z.H. Liao, Y. Zhou, and S.Z. Li, *A new set of amino acid descriptors and its application in peptide QSARs*, Biopolymers 80 (2005), pp. 775–786.

[27] F. Tian, P. Zhou, and Z. Li, *T-scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides*, J. Mol. Struct. 830 (2007), pp. 106–115.

[28] F. Tian, L. Yang, F. Lv, Q. Yang, and P. Zhou, *In silico quantitative prediction of peptides binding affinity to human MHC molecule: An intuitive quantitative structure-activity relationship approach*, Amino Acids 36 (2009), pp. 535–554.

[29] C. Cortes and V. Vapnik, *Support-vector networks*, Mach. Learn. 20 (1995), pp. 273–297.

[30] C.C. Chang and C.J. Lin, *LIBSVM: A library for support vector machines*, ACM Transact. Intel. Syst. Technol. 2 (2011), pp. 1–27.

[31] A. Golbraikh and A. Tropsha, *Beware of q2!, J. Mol. Graph. Model. 20 (2002), pp. 269–276.

[32] J.H. Wang, Y.L. Liu, J.H. Ning, J. Yu, X.H. Li, and F.X. Wang, *Is the structural diversity of tripeptides sufficient for developing functional food additives with satisfactory multiple bioactivities?, J. Mol. Struct. 2013 (1040), pp. 164–170.

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

[34] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water*, J. Chem. Phys. 79 (1983), pp. 926–935.

[35] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems*, Chem. Phys. 98 (1993), pp. 10089–10092.

[36] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes*, J. Comput. Phys. 23 (1977), pp. 327–341.
[37] N. Homeyer and H. Gohlke, *Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method*, Mol. Inf. 31 (2012), pp. 114–122.

[38] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R.J. Woods, *The Amber biomolecular simulation programs*, J. Comput. Chem. 26 (2005), pp. 1668–1688.

[39] T. Hou, K. Chen, W.A. McLaughlin, B. Lu, and W. Wang, *Computational analysis and prediction of the binding motif and protein interacting partners of the Abl SH3 domain*, PLoS Comput. Biol. 2 (2006), p. e1.

[40] I.A. Doytchinova, V.A. Walshe, N.A. Jones, S.E. Gloster, P. Borrow, and D.R. Flower, *Coupling in silico and in vitro analysis of peptide-MHC binding: A bioinformatic approach enabling prediction of superbinding peptides and anchorless epitopes*, J. Immunol. 172 (2004), pp. 7495–7502.

[41] J. Robinson, J.A. Halliwell, H. McWilliam, R. Lopez, P. Parham, and S.G. Marsh, *The IMGT/HLA database*, Nucleic Acids Res. 41 (2013), pp. D1222–D1227.

[42] N.P. Brown, C. Leroy, and C. Sander, *MView: A web-compatible database search or multiple alignment viewer*, Bioinformatics 14 (1998), pp. 380–381.

[43] M. García-Peydró, M. Martí, and J.A. López de Castro, *High T cell epitope sharing between two HLA-B27 subtypes (B*2705 and B*2709) differentially associated to ankylosing spondylitis*, J. Immunol. 163 (1999), pp. 2299–2305.

[44] B. Galocha, J.R. Lamas, J.A. Villadangos, J.P. Albar, and J.A. López de Castro, *Binding of peptides naturally presented by HLA-B27 to the differentially disease-associated B*2704 and B*2706 subtypes, and to mutants mimicking their polymorphism*, Tissue Antigens 48 (1996), pp. 509–518.

[45] D. Fruci, P. Rovero, G. Falasca, A. Chersi, R. Sorrentino, R. Butler, N. Tanigaki, and R. Tosi, *Anchor residue motifs of HLA class-I-binding peptides analyzed by the direct binding of synthetic peptides to HLA class I alpha chains*, Hum. Immunol. 38 (1993), pp. 187–192.

[46] A. Ben-Hur and J. Weston, *A user’s guide to support vector machines*, Meth. Mol. Biol. 609 (2010), pp. 223–239.

[47] P. Zhou, F. Tian, F. Lv, and Z. Shang, *Comprehensive comparison of eight statistical modelling methods used in quantitative structure-retention relationship studies for liquid chromatographic retention times of peptides generated by protease digestion of the Escherichia coli proteome*, J. Chromatogr. A 1216 (2009), pp. 3107–3116.

[48] P. Atagunduz, H. Appel, W. Kuon, P. Wu, A. Thiel, P.M. Kloetzel, and J. Sieper, *HLA-B27-restricted CD8+ T cell response to cartilage-derived self peptides in ankylosing spondylitis*, Arthritis Rheum. 52 (2005), pp. 892–901.