

# Structural Immunoinformatics: Understanding MHC-Peptide-TR Binding

Javed Mohammed Khan, Joo Chuan Tong, and Shoba Ranganathan

## Introduction

In higher jawed vertebrates, antigen presentation and recognition occurs in two steps, where, the peptide ligand first binds to the major histocompatibility complex (MHC) molecule followed by recognition of this peptide–MHC (pMHC) complex by the T cell receptor (TR). These two steps play a key role in the activation of normal adaptive immune responses. The first step in TR-mediated immune response is thus the binding and presentation of antigenic endogenous or exogenous peptide epitopes, which can be successfully predicted using sequence-based methods for alleles with large datasets of known binding peptides (Tong et al. 2007a). The second step, however, is an intricate theoretical problem that remains unsolved and is the next frontier in Immunoinformatics.

With the development of new structural modeling and docking techniques and an increase in the number of protein structures deposited in the Protein Data Bank (PDB) (Berman et al. 2000), the use of structure-based approaches to predict potential T-cell epitopes is increasingly successful (Ranganathan et al. 2008), often producing modeled structures accurate to within 2.00 Å RMSD from the experimental crystal structure, providing a wealth of information for structural analysis and prediction. With the development of a fast and accurate docking protocol followed by quantitative predictions for both MHC Class I and Class II alleles even with limited binding peptide data, we have been successful in unraveling the mystery behind the first step (Tong et al. 2004) in adaptive immune response. For an MHC molecule to recognize antigenic peptides and for pMHC to be subsequently recognized by TR, geometric and electrostatic complementarity between the receptor and ligand are essential, determining the stability of the complex. In this context, the introduction of structural information can greatly facilitate our understanding

---

S. Ranganathan (✉)

Department of Chemistry and Biomolecular Sciences and ARC Centre of Excellence in Bioinformatics, Macquarie University, 2109 Sydney, NSW Australia

Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Medical Drive, Singapore 117597

of how well a pMHC complex can associate with TR besides being able to predict epitopes, capable of eliciting TR response.

Here we introduce structural immunoinformatics concepts of TR/pMHC interaction based on their three-dimensional experimental and modeled structures, toward the development of a predictive model. To start with, we briefly describe the structural characteristics of pMHC complexes (Tong et al. 2006a), from which a novel supertype classification for class I MHC alleles has been developed (Tong et al. 2007b). Based on the analysis of these characteristics, we have developed a rapid and precise docking protocol to generate models of pMHC complexes which has been applied to antigenic epitope prediction for specific alleles implicated in different diseases (Tong et al. 2006b, c; Tong et al. 2007c), where sequence-based approaches are inapplicable due to limited data on antigenic peptides. We then summarize the available TR/pMHC structural resources from the Internet. Finally, we extend our pMHC interaction parameters and highlight the importance of TR/pMHC interactions to decipher how antigenic peptides elicit a T cell response.

## **MPID-T and Structurally Derived Interaction Parameters**

With the growth in the numbers of pMHC and TR/pMHC structures in PDB and some interaction parameters being reported as significant for peptide/MHC interactions (Kangueane et al. 2001), there was an increasing need for a database dedicated to these structures and their analysis. Hence a preliminary database called MPID (Govindarajan et al. 2003) was developed which is composed of 86 classical pMHC structures. This was later extended to the TR level when all the TR/pMHC structures along with the new pMHC structures were included. The new version was called MPID-T (Tong et al. 2006a) and it contained 187 pMHC and 16 TR/pMHC structures. MHC-peptide Interaction Database – TR (MPID-T) – is a manually curated database comprising all the X-ray crystallographic structures of pMHC and TR/pMHC complexes obtained from the PDB. It was developed in order to understand the structural determinants of TR/pMHC recognition and binding. It provides sequence-structure-function information governing TR/pMHC interactions besides a web-interface to carry out structural analysis of these complexes. The database (April 2008 update) contains over 294 pMHC complexes (Class I complexes: 235, Class II complexes: 59, TR/pMHC complexes: 42) from five MHC sources (human 187, murine 101, rat 3, chicken 2, and monkey 1), spanning 52 alleles. The analysis of these complexes revealed the significance of some protein-protein interaction parameters for the characterization of pMHC interface. These parameters include interface area or change in solvent accessible surface area (hydrophobic contacts), intermolecular hydrogen bonds, gap index (electrostatic interactions), and gap volume (geometric complementarity) between the MHC receptor and its corresponding peptide ligand. We will explain each of these parameters briefly with the view to extending them to characterize TR/pMHC

interactions and thus help us determine the potential of a pMHC complex to bind to TR and thereby elicit T cell response.

### ***Interface Area Between Peptide and MHC***

One of the most significant forces that drive protein folding and protein–protein interactions are hydrophobic interactions. The hydrophobic free energy of a protein when it is transferred from polar to hydrophobic environment and the change in solvent accessible surface area ( $\Delta$ ASA) upon complex formation share a linear correlation (Chothia and Janin 1975). Thus, an indication of the binding strength of the interacting partners is provided by the knowledge of the surface area of a complex interface in direct contact with solvent. The measure of the maximum permitted van der Waals' contact area that is covered by the center of a water molecule as it rolls over the surface of the protein gives the accessible surface area. Interface area for pMHC complexes is defined as the mean  $\Delta$ ASA on complex formation when going from separated MHC and peptide molecules to a pMHC complex state and is calculated as half the sum of the total  $\Delta$ ASA of both molecules for each type of complex (Tong et al. 2006a). The interface areas of all the molecular systems are computed using the program NACCESS (Hubbard and Thornton 1993).

$$\Delta\text{ASA}(\text{pMHC}) = [\text{ASA}(\text{MHC}) + \text{ASA}(\text{peptide}) - \text{ASA}(\text{pMHC})]/2 \quad (1)$$

The mean  $\Delta$ ASA for class I pMHC complexes was found to be  $903.30 \pm 260.90 \text{ \AA}^2$ . Similarly, the corresponding  $\Delta$ ASA for class II pMHC complexes was  $894.40 \pm 364.00 \text{ \AA}^2$  (Ranganathan et al. 2008).

### ***Intermolecular Hydrogen Bonds***

The selectivity and stability of proteins and protein–protein complexes depend on various factors but one of the most important and major contribution comes from hydrogen bonds. Hydrogen bonds are the collective interaction of three atoms wherein a hydrogen atom is bound to donor electronegative atom and an acceptor electronegative atom in close proximity. The typical observed hydrogen bond distance is approximately 2.60–3.10 Å (1.00–1.20 Å between donor and hydrogen and 1.60–2.00 Å between hydrogen and acceptor). The significance of such bonding relies on both electronegative atoms being derived from the group: F, N, and O (Morrison and Boyd 1992). Only these elements are sufficiently negative and the hydrogen bound to any of these three elements is sufficiently positive for the required attraction to exist as their small atoms have a high concentration of negative charge on them. The interactions between side-chains are directed and their geometry is restricted due to the presence of hydrogen bonds. With decreasing bond length the strength of hydrogen bond usually increases.

## Gap Volume

Gap volume is a measure of the volume enclosed by the two interacting molecular subunits. The gap volume between the MHC and the peptide in each complex can be computed using the SURFNET program (Laskowski 1991). This is done by the algorithm by placing a series of spheres (maximum radius 5.00 Å) between the surfaces of each pair of the MHC and the peptide subunit atoms, such that its surface is in contact with the surfaces of the atoms on either side. Upon interception by other atoms, the size of each sphere is reduced accordingly and is subsequently discarded if it falls below a minimum allowed radius (1.00 Å). The gap volume between the two subunits is calculated by taking into account the sizes of all the remaining allowable gap-spheres.

## Gap Index

Electrostatic and geometric complementarity observed between associating molecules has always been an essential feature in receptor–ligand binding. Gap index (Jones and Thornton 1996) is a valuable method to evaluate complementarity of interacting interfaces:

$$\text{Gap index (Å)} = \frac{\text{Gap volume between peptide and MHC (Å}^3\text{)}}{\text{pMHC interface ASA (Å}^2\text{)}} \quad (2)$$

The results for the mean gap indices of class I ( $0.95 \pm 0.24$  Å) and class II ( $1.12 \pm 0.20$  Å) pMHC complexes (Kangueane et al. 2001) indicate that the interacting surfaces in pMHC complexes are significantly complementary. The gap index in class II complexes is, on an average, higher than in class I complexes. This implies that the interface area of class I complexes is greater than its corresponding gap volume. On the contrary, in class II complexes, the gap volume is greater than the interface area. The complexes of different alleles in both class I and class II structures did not show much difference in their gap indices.

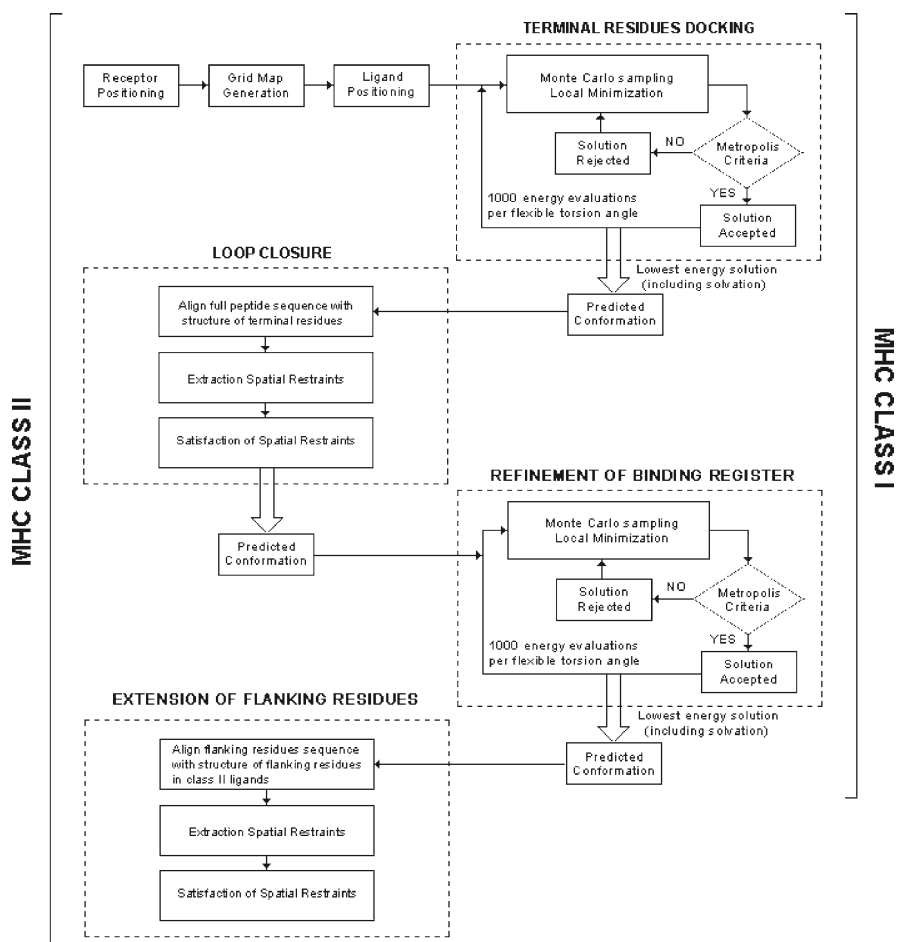
## Supertype Classification Based on Structural Characteristics

Human leukocyte antigens (HLA) were until now classified based on the common structural features of HLA proteins (Doytchinova et al. 2004; Doytchinova and Flower 2005) and/or their functional binding specificities (Lund et al. 2004; Kangueane et al. 2005). These approaches leave the structural interaction characteristics among different HLA supertypes with antigenic peptides unexplored. We therefore classified 68 HLA class I molecules using the number of intermolecular hydrogen bonds between each HLA protein and its corresponding bound peptide, solvent accessibility of each pMHC complex, gap volume, and gap index as

described above. This type of classification of the HLA proteins into supertypes helps in the identification of promiscuous T cell epitopes that bind multiple alleles and is thus the underlying reason for the development of successful epitope-based vaccines covering a wide number in the world population and all the ethnicities (Sette et al. 2001; Sette et al. 2002). The interaction parameters investigated in this study tend to vary among different alleles and were thus grouped in a supertype dependant manner. Our analysis resulted in successful classification of the HLA class I molecules into eight supertypes based on their crystallographic structures (Tong et al. 2007b). The HLA-A supertypes had three main clusters AI, AII, and AIII whereas HLA-B supertypes had five clusters namely BI, BII, BIII, BIV, and BV. Our data largely overlap the definition of binding motifs. The proposed methodology of classification which considers conformational information of both peptide and HLA proteins provides an alternative to the characterization of supertypes using either peptide or HLA protein information alone. A hierarchical clustering technique using the agglomerative algorithm (Doytchinova et al. 2004; Doytchinova and Flower 2005) was applied in this approach. The distance between the structures was computed by the single-linkage method based on the separation between the each pair of data points (Barnard and Downs 1992). The nearest neighbors were merged into clusters. Smaller clusters were then merged into larger clusters based on inter-cluster distances, until all structures are combined. We have considered the last three levels for defining HLA class I supertypes.

## The MHC–Peptide Docking Protocol

Of all the techniques used for investigating intermolecular interactions, computer-simulated ligand binding or docking is the most powerful and widely used. The general purpose of docking simulation is: (i) to find the most probable translational, rotational, and conformational juxtaposition of a given ligand–receptor pair, and (ii) to evaluate the relative goodness-of-fit or how well a ligand can bind to the receptor. We now introduce a rapid and highly accurate docking protocol for the modeling of bound peptide ligands to the MHC receptor. We begin with the sequence of the ligand for which the structure is to be generated (peptide) and the availability of the target MHC receptor structure. Our docking protocol consists of four steps in all, out of which three essential steps are common for both class I and class II MHC complexes: (i) rigid docking of terminal residues of the peptide nonamer; (ii) loop closure of central residues by satisfaction of spatial constraints; (iii) followed by ab initio refinements of backbone and ligand interacting side-chain. However, for MHC class II complexes, the peptide length is usually between 12 and 15 residues and therefore there was a necessity to carry out addition of extra residues at both ends of the peptide nonamers as these residues usually extend out of the binding groove of the MHC molecule. Thus, step (iv) extension of flanking residues was included in the docking procedure for class II complexes. The flow diagram of the docking protocol is illustrated in Fig. 1.



**Fig. 1** Flowchart of the four-step docking procedure used in this chapter

### *Step 1: Rigid Docking of Nonamer Termini*

To decide the number of combinations for two molecules within an enclosed sampling space is the key issue in docking simulation methodologies and an important factor that determines the best fit for the final output. There are six degrees of global-rotational and translational freedom of one molecule relative to the other, as well as one internal dihedral rotation per rotational bond. An increase in molecule size and sampling space increases search on the conformational space exponentially. Minimization of the conformational search space of ligand within the large sampling space enclosed by the MHC binding groove is a challenge in pMHC docking simulation. A possible approach to initiate docking simulations is to identify suitable anchor residues or nonamer termini (probes) for rigid docking. A probe

must satisfy two criteria: (i) the anchor must have sufficient contact with the receptor, and (ii) the structure of the anchor must be highly conserved.

Peptide residues at the N and C termini of the nonamer are almost in invariant positions at the end of the binding groove of the MHC with mean backbone C $\alpha$  RMSD within  $0.15 \pm 0.14 \text{ \AA}$  (Tong et al. 2004) and are ideal for such purpose. To model each probe to the receptor, a fast soft-interaction energy function (Fernández-Recio et al. 2002) is adopted. This is performed using an internal coordinate mechanics (ICM) (Abagyan and Maxim 1999) global optimization algorithm, with flexible ligand interface side-chains and a grid map representation of the receptor energy localized to small cubic regions of  $1.00 \text{ \AA}$  radius from the backbone of each probe. Within their respective grid map, each anchor residue performs a random walk. Using a Biased Monte Carlo procedure, which begins by pseudo-randomly selecting a set of torsion angles in the probe and subsequently finding the local energy minimum about those angles, the side-chain torsions were changed at random for each step. Upon satisfaction of the Metropolis criteria with probability  $\min(1, \exp[-\Delta G/RT])$ , where  $R$  is the universal gas constant and  $T$  is the absolute temperature of the simulation, new conformations are adopted. To keep the positional variables of the ligand molecule close to the starting conformation, loose restraints were imposed on it. The stimulation temperature was set to 300 K. The internal energy of the probe and the intermolecular energy based on the same optimized potential maps used in the docking step together comprised the optimal energy function used during simulations:

$$E = E_{\text{Hvw}} + E_{\text{Cvw}} + 216E_{\text{el}}^{\text{solv}} + 253E_{\text{hb}} + 435E_{\text{hp}} + 0.20E_{\text{solv}} \quad (3)$$

The internal energy included internal van der Waals interactions, hydrogen bonding, and torsion energy calculated with ECEPP/3 parameters, and the Coulomb electrostatic energy with a distance-dependent dielectric constant ( $\epsilon = 4r$ ). In order to select the best-refined solutions the surface-based solvation energy and the configurational entropy of side-chains were included in the final energy.

## ***Step 2: Loop Closure of Middle Residues***

By satisfaction of spatial constraints (Sali and Blundell 1993) based on the allowed subspace for backbone dihedrals in accordance with the conformations of nonamer termini docked into the ends of the binding groove, an initial conformation of the central loop is generated at this stage. The three steps that are used to perform this are: (i) The alignment of the entire peptide sequence and the sequences of probes docked into the binding groove gives the distance and dihedral angle restraints on the peptide sequence. (ii) By extrapolation from the known 3D structures of probes in the alignment, expressed as probability density functions, the restraints on spatial features of the unknown central residues are derived. Stereo-chemical restraints include bond distances, bond angles, planarity of peptide groups and side-chain rings, chiralities of C $\alpha$  atoms and side-chains, van der Waals contact distances and the bond lengths, bond angles, and dihedral angles of cysteine disulfide bridges.

(iii) Optimizing the molecular probability density function using variable target function technique that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms satisfies spatial restraints on the unknown central residues.

### ***Step 3: Refinement of Binding Register***

Partial refinement was performed for both the ligand backbone and side-chain to improve the accuracy of the initial model using ICM Biased Monte Carlo procedure (Abagyan and Maxim 1999). By introducing partial flexibility to the ligand backbone, preliminary stages of refinements attempt to nullify the penalty derived from the initial rigid docking of terminal residues thus making this an effective and flexible docking procedure. Restraints were imposed upon the positional variables of the C $\alpha$  atoms of probes to keep it close to the starting conformation. This refinement step is performed using the energy function:

$$E = E_{\text{vw}} + E_{\text{hbonds}} + E_{\text{torsions}} + E_{\text{electr}} + E_{\text{solv}} + E_{\text{entropy}} \quad (4)$$

Ligand and receptor side-chain torsions within 4.00 Å from the receptor were refined upon the final backbone structure.

### ***Step 4: Extension of Flanking Residues***

By now, MHC class I ligand models have already been fully constructed and this step is applicable only to MHC class II ligands. The only construction remaining is of the flanking residues that extend out of the MHC class II binding groove. The conformations of the flanking peptide residues are generated by satisfying the spatial constraints in the allowed subspace for backbone dihedrals defined by the conformation of the bound core nonameric peptide docked into the binding groove. This is again performed in three stages: (i) distance and dihedral angle restraints on the entire peptide sequence are derived from its alignment with the nonamer sequence in the binding groove; (ii) the restraints on spatial features of the flanking residues are derived by extrapolation from the known 3D structure of flanking residues in the alignment, expressed as probability density functions; and (iii) the spatial restraints on the flanking residues are then satisfied by optimization of the molecular probability density function using a variable target function technique that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms.

## **Epitope Prediction**

We will now compare the applications of this protocol for the discrimination of binders/non-binders from MHC class I and class II alleles. First, we discuss the docking of 68 peptides with known IC<sub>50</sub> values and 12 peptides with experimental T cell



proliferation values on to the binding groove of DQ3.2 $\beta$  MHC class II allele associated with several allergies and autoimmune diseases. Our model predicts DQ3.2 $\beta$  binding peptides with high accuracy [area under the receiver operating characteristic (ROC) curve  $A_{\text{ROC}} > 0.88$ ] (Tong et al. 2006b), compared with experimental data. Our investigation of the binding patterns of DQ3.2 $\beta$  peptides illustrates that several registers exist within a candidate binding peptide. Further analysis reveals that peptides with multiple registers occur predominantly for high-affinity binders (specificity=0.95). We successfully predicted 20/23 (87%) binding registers with excellent discrimination of low-, medium-, and high-affinity binders. The results also proved that our method was of high precision with a sensitivity value as high as 0.81 (81%) for low-, medium-, and high-affinity binders.

In the second experiment we carried out docking of 51 DRB1\*0402-specific desmoglein 3 (Dsg3) peptides with known  $IC_{50}$  values, 25 DRB1\*0402-specific Dsg3 peptides with experimental T cell proliferation values and 6 DQB1\*0503-specific Dsg3 peptides with experimental T cell proliferation values into the binding groove of *Pemphigus vulgaris* (PV) associated MHC class II alleles DRB1\*0402 and DQB1\*0503, respectively (Tong et al. 2006c). Docking of anchor peptide residues is performed using the docking procedure prior described followed by ab initio modeling of flanking residues. Our models present the best fit of each peptide into the binding cleft of each disease associated allele based on the following criteria: (i) pattern of hydrogen bonding to the MHC molecule; (ii) pattern of hydrophobic burial of peptide side-chains, and (iii) the absence of atomic clashes or repulsive contacts. Figure 2 illustrates the immunological hotspots that were predicted for both the alleles across the Dsg3 glycoprotein proteome. These immunological hotspots are the regions that contain immunogenic peptide epitopes that are highly suitable for vaccine design.

We were able to successfully predict all 25 and 5/6 peptides in test set II as high-binders for both DR and DQ alleles, respectively, with high accuracy ( $A_{\text{ROC}} = 0.93$ ) and specificity ( $SP = 0.80$ ). These results were consistent with our earlier qualitative structural studies (Tong et al. 2006d). Furthermore, these results confirm that both DRB1\*0402 and DQB1\*0503 are strongly associated with PV. Our analysis revealed the existence of multiple immunodominant epitopes that may be responsible for both disease initiation and propagation in PV and also suggests that DRB1\*0402 and DQB1\*0503 may share similar specificities by binding peptides of different binding registers, thus providing a molecular mechanism for the dual HLA association observed in PV. In this example, pMHC residues were considered to be in contact if at least one pair of their non-hydrogen (“heavy”) atoms was found to be within 4.00 Å radius (Fischer and Marquese 2000). Intra-peptide interactions and intra-MHC interactions were not considered as they have minor influence on backbone structure. Any atom in the peptide and any atom in the MHC were considered to be experiencing atomic clash if their separation is below 2.00 Å (Samudrala and Moult 1997) for non-hydrogen atoms and below 1.60 Å for atoms participating in hydrogen bonds (Samanta et al. 2002; Wallace et al. 1995).

We also performed a docking and binding prediction study of the repertoires for HIV-1 p24gag and gp160gag glycoproteins that are known to bind the MHC class I allele HLA-Cw\*0401 which plays a major role in the control of human immunodeficiency virus type 1 (HIV-1) infection. The analysis of predicted Cw\*0401-

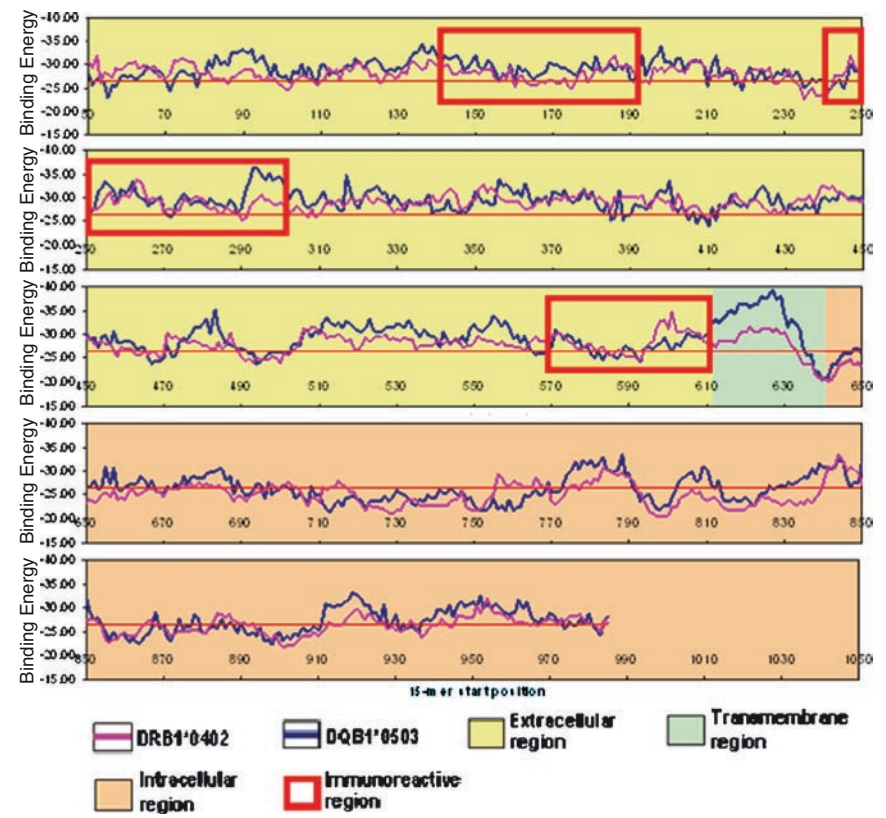


Fig. 2 Proteome-wide screening of Dsg3 peptides

binding peptides showed that anchor residues may not be restrictive and the Cw\*0401 binding pockets may possibly accommodate a wide variety of peptides with common physico-chemical properties (Tong et al. 2007c). The potential Cw\*0401-specific T cell epitopes are well distributed throughout both glycoproteins, with 14 and 9 immunological hotspots for HIV-1 p24gag and gp160gag glycoproteins, respectively. External validation results indicate that our Cw\*0401 predictive model provides excellent discrimination between binding and the non-binding ligands with high accuracy ( $A_{ROC}=0.93$ ) and a sensitivity of 76% ( $SE=0.76$ ) and a specificity as high as 95% ( $SP=0.95$ ). Our results strongly indicate that Cw\*0401 can bind antigenic peptides in amounts comparable to both HLA-A and -B molecules, and support the existence of a potentially large number of Cw\*0401-specific T cell epitopes.

Table 1 gives a summary of the overall comparison of the results from the three experiments described above. An important thing to note is the diminishing training datasets used in the three experiments which suggests that our prediction model can

**Table 1** Comparison of the training set, test set, and results of the three prediction experiments

DQ3.2β (Tong et al. 2006b)	Training set	56 binding and 30 non-binding conformations from experimentally determined binding and non-binding peptides
	Test set	I – 68 peptides with known IC <sub>50</sub> values II – 12 peptides with known T cell proliferation values
	Results	I – A <sub>ROC</sub> =0.88, SE=0. 81, and SP=0.95. 20/23 (87%) binding registers were predicted correctly. II – Top five predictions (SE=0.95) have known T cell proliferation values
DR and DQ (Tong et al. 2006c)	Training set	8 DRB1*0402-specific Dsg3 peptides 8 DQB1*0503-specific Dsg3 peptides
	Test set	I – 51 DR-specific Dsg3 peptides with known IC <sub>50</sub> values II – 25 DR and 6 DQ-specific Dsg3 peptides with experimental T cell proliferation values
	Results	I – A <sub>ROC</sub> =0.93, SE=0.70 and SP=0.95. II – All 25 DR-specific peptides in test set II were predicted as high binders (SE=0.65, SP=0.80). 5/6 DQ-specific peptides (all true positives) were determined with a cutoff of –26.64 kJ/mol
HLA-Cw*0401 (Tong et al. 2007c)	Training set	6 peptide sequences with known IC <sub>50</sub> values
	Test set	58 peptides known to bind Cw*0401
	Results	Test set accuracy: A <sub>ROC</sub> =0.93, SE=0.76, and SP=0.95. 14 and 9 potential Cw*0401-specific T cell immunogenic regions or hotspots were predicted for HIV-1 p24gag and gp160gag glycoproteins, respectively

give excellent results with as low as six peptides in the training dataset. It, however, is yet to be determined as to what proportion of these predicted peptides may be expressed at the cell surface and are capable of eliciting functional T cell responses. One aspect of this would be to look at the TR/pMHC complex on the whole.

TR/pMHC Interaction

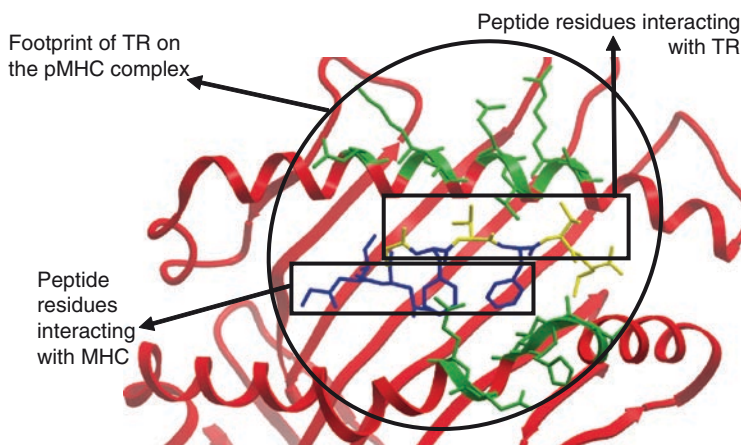
TR/pMHC interaction is the most essential binding step in the entire adaptive immune response cascade. It is this interaction that is responsible for eliciting a T cell response. After an endogenous or exogenous peptide is bound to the MHC class I or class II, respectively, the pMHC complex is transported to the membrane of the antigen presenting cell (APC) and is presented at the cell surface for surveillance by the TR which then binds to the pMHC and forms the TR/pMHC complex. This binding is also partly determined by the cluster of differentiation (CD) molecules present on the membrane of the T cells as these bind specifically to MHC class I and class II proteins. Therefore, class I pMHC molecules stimulate CD8+ cytotoxic T cells which directly kill the infected cells whereas class II pMHC molecules stimulate CD4+ helper T cells which in turn activate B cells, leading to antibody production. Three-dimensional structures of the pMHC complex and the TR are essential and play a vital role in the activation of the Adaptive immune system. With the chance of 1 in 2,000 antigenic peptides being able to stimulate T

cells (Yewdell and Bennink 1999), finding immunogenic peptide epitopes poses a great challenge. In view of this enormity, it is experimentally impossible to scan all the putative peptides arising from the proteomes of all pathogens known today. An in-depth analysis of TR/pMHC complexes using structural immunoinformatics combines the power of computational analysis with detailed structural data to accelerate immune system research and provides clues for the development of vaccines for immunotherapeutic applications.

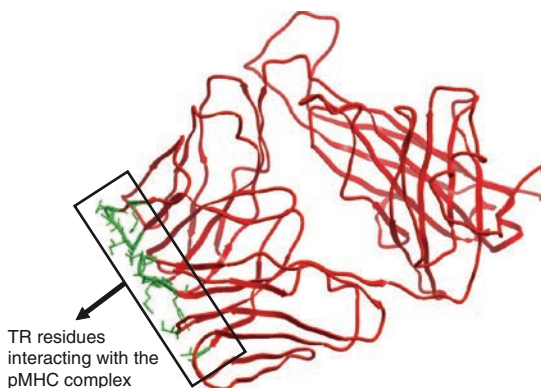
Figure 3 depicts the TR footprint on a pMHC complex. The residues (both MHC and peptide residues) that comprise the footprint are the key to pMHC recognition by the TR molecule via the variable regions of the  $\alpha$  and  $\beta$  chains. The residues on the TR molecule that interact with the residues of the pMHC complex (Fig. 4) are also equally essential to this process as they recognize the corresponding residues on the pMHC complex and form H-bonds with them to anchor the TR molecule to the pMHC complex.

However, it is the peptide epitope that is important for vaccine development as a few of its residues take part in the peptide–MHC binding and the others take part in the TR/pMHC recognition and binding. Therefore, it can be inferred that it is the epitope that acts as a key to unlock the immune cascade and thereby plays the most important role in this major defense mechanism in higher vertebrates.

A detailed analysis of the types of inter-residue interactions observed in this complex is shown in Fig. 5. The peptide residues, Gly4(C), Thr8(C), and Val6(C), interact with TR residues, Gln52(E), Asp32(E), Gln52(E), and Ser99(E), respectively, to form H-bonds and thereby stabilize the TR/pMHC complex besides anchoring TR to the pMHC complex. Gly97(D) and Ile53(E) of the TR are also



**Fig. 3** A schematic structure of the top view of the MHC groove and the bound peptide in a class I TR/pMHC complex (PDB ID – 1OGA) (Stewart-Jones et al. 2003) in  $\alpha$  trace ribbon representation, with MHC in red and peptide in blue. Residues interacting with TR are highlighted in green (MHC) and yellow (peptide), with heavy side-chain atoms shown in stick representation. The black oval contains these residues or the footprint of the TR on the pMHC complex



**Fig. 4** Schematic C $\alpha$  trace ribbon representation of the T cell receptor (PDB ID – 1OGA) in red, with the residues in its variable regions interacting with the pMHC complex highlighted in green

involved in hydrophobic interactions with Gly4(C) and Phe5(C), Thr8(C), respectively, thereby contributing to the overall stability of the TR/pMHC complex. Such interactions are also seen between the MHC and the peptide residues suggesting the importance of peptide epitopes for immunotherapy and vaccine development. Water bridges (not shown) also play a significant role in this TR/pMHC complex formation and stability.

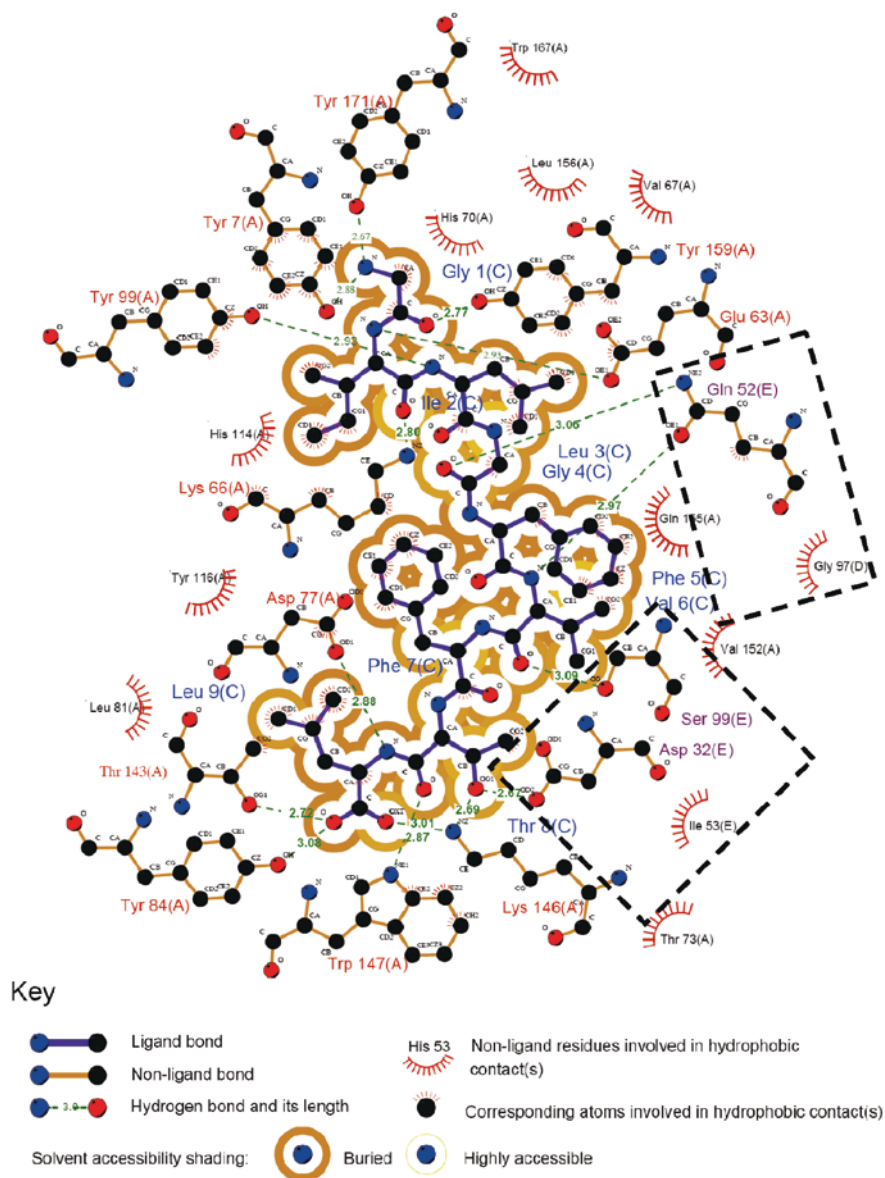
## Analysis of the 1OGA Complex

We now extend our interaction parameters to the TR level in order to analyze the 1OGA TR/pMHC structure, using the formulae described earlier. The interface area for the TR/pMHC complex is 733.55 Å<sup>2</sup> which is lower compared to that of the pMHC complex (856.00 Å<sup>2</sup>) for the same structure. This suggests that the TR/pMHC binding is localized. As seen in Fig. 5, the peptide is sandwiched between the TR and the pMHC and is fairly well buried in terms of the accessible surface area.

Figure 5 indicates the presence of four H-bonds between the peptide and the TR residues [Gly4(C) with Gln52(E), Val6(C) with Gln52(E), and Ser99(E), Thr8(C) with Asp32(E)]. The average H-bond length between the peptide and the TR residues is 2.95 Å. H-bonding (not shown) between the TR residue, Arg98(E) and the MHC residues, Ala150(A) and Gln155(A) with an average bond length of 2.92 Å, also plays a vital role in anchoring the TR onto the pMHC complex. Notably, 12 hydrogen bonds are seen to stabilize and anchor the peptide firmly onto the MHC molecule, of length 2.67–3.08 Å. Hydrophobic interactions are also seen to occur specially between the MHC and the TR molecules, involving Val152(A), Gln155(A), and Lys66(A) residues of MHC (interacting TR residues not shown in Fig. 5), while Thr73(A) interacts strongly with its neighboring TR residue, Ile53(E).

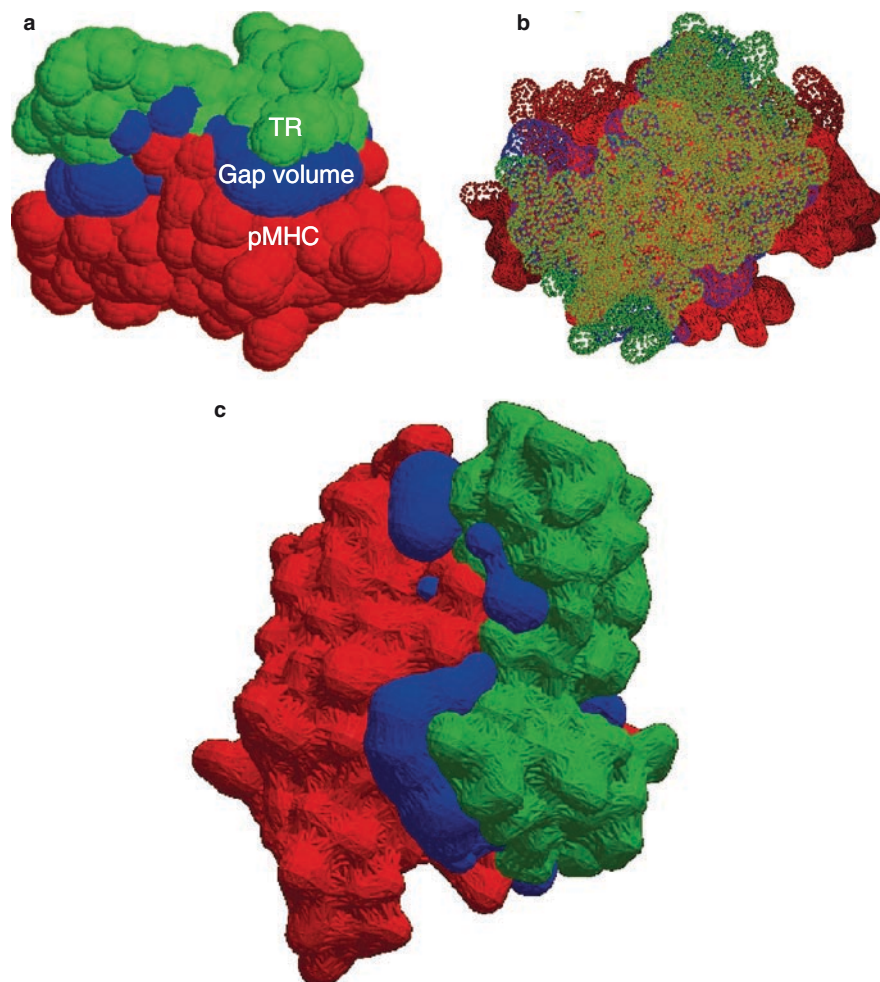
The gap volume between pMHC and TR was calculated using a molecular model of only the interacting regions from both pMHC and TR, generated from the





**Fig. 5** A LIGPLOT schematic diagram of the peptide ligand from the three dimensional structure of class I TR/pMHC complex (PDB ID – 1OGA) showing the solvent accessibility and the interactions between its residues and the corresponding MHC and TR residues. The TR residues, Gln52(E), Asp32(E), Ser99(E), Ile53(E), and Gly97(D) that together anchor the TR on the pMHC are shown in dotted rectangles. The letters in the brackets correspond to the respective chain IDs to which the residues belong. (A) – MHC  $\alpha$  chain, (C) – peptide, (D) and (E) – TR  $\alpha$  and  $\beta$  chains, respectively

template structure, 1OGA using MODELLER (Sali and Blundell 1993). The gap volume was then computed to be  $3304.43 \text{ \AA}^3$  using SURFNET (Laskowski 1991), depicted by the blue region in Fig. 6. This value of the gap volume between the TR and pMHC is very large compared to that between the peptide and MHC suggesting that the binding between peptide and MHC is much stronger than that between



**Fig. 6** (a) A space filling representation of the interacting residues of TR (green) and pMHC (red) from the 1OGA crystal structure, showing the gap volume (blue) between the two complexes. A fairly large gap volume supports the theory that the binding is not very strong, unlike the pMHC binding. The surface representation file used for visualization was generated using SURFNET (Laskowski 1991). (b) Top view of the TR/pMHC complex shown in an atomic mesh representation. The TR binds to the pMHC at an angle of  $69^\circ$  (Stewart-Jones et al. 2003). (c) Side view of the TR/pMHC interacting regions, in ball and stick surface representation

pMHC and TR. These results underline the importance of strong peptide binders in the first step of the entire adaptive immune response cascade. A large gap volume and a small interface area indicate that the gap index of the TR/pMHC complex is high (4.50 Å), compared to that of the peptide and the MHC (0.60 Å). This suggests that the electrostatic and the geometric complementarities of the TR and pMHC are not as significant as between the peptide and MHC.

## Conclusion

Our analysis and extensive studies on peptide–MHC interactions have revealed structural features that can be analyzed in terms of the parameters governing the pMHC complex formation. We have now extended this formalism to defining the interaction between TR and pMHC, relevant for immune system activation. Based on our pMHC analyses, we have developed methods to successfully predict T cell epitopes in accordance with their MHC binding specificities. The next challenge is to extend this methodology to the unexplored TR level as this would greatly improve the efficacy of our prediction model, in separating a large number of predicted MHC-binding peptides from true T cell epitopes. The complexities involved in methodology development and the computational costs incurred in docking peptides and proteins have hindered the progress of structure-based prediction techniques. In the era of high throughput and distributed computing over global grids, the necessary computational requirements for large-scale structure-based screening of potential T cell epitopes are now available. We can therefore expect new structure-based approaches to predicting promiscuous peptide epitopes for MHC supertypes and TR activation, for the design of sub-type-specific vaccines with wide population coverage. Large-scale structure-based screening helps overcome the barriers of insufficient training data and the lack of peptide binding motifs, especially for MHC class II alleles by cutting down the lead time involved in experimental vaccine development methods, resulting in the production of effective and highly specific peptide vaccines.

## References

- Abagyan R, Maxim T (1999) Ab Initio Folding of Peptides by the Optimal-Bias Monte Carlo Minimization Procedure. *J Comput Phys* 151:402–421
- Barnard JM, Downs GM (1992) Clustering of chemical structures on the basis of two-dimensional similarity measures. *J Chem Inf Comput Sci* 32:644–649
- Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242
- Chothia C, Janin J (1975) Principles of protein–protein recognition. *Nature* 258:705–708
- Doytchinova IA, Flower DR (2005) *In silico* identification of supertypes for class II MHCs. *J Immunol* 174:7085–7095
- Doytchinova IA, Guan P, Flower DR (2004) Identifying human MHC supertypes using bioinformatic methods. *J Immunol* 172:4314–4323



- Fernández-Recio J, Totrov M, Abagyan R (2002) Soft protein–protein docking in internal coordinates. *Protein Sci* 11:280–291
- Fischer KF, Marqusee S (2000) A rapid test for identification of autonomous folding units in proteins. *J Mol Biol* 302:701–712
- Govindarajan KR, Kangueane P, Tan TW et al (2003) MPID: MHC-Peptide Interaction Database for sequence-structure-function information on peptides binding to MHC molecules. *Bioinformatics* 19:309–310
- Hubbard SJ, Thornton JM (1993) “NACCESS” computer Program, Department of Biochemistry and Molecular Biology, University College, London
- Jones S, Thornton JM (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci* 93:13–20
- Kangueane P, Sakharkar MK, Kolatkar PR et al (2001) Towards the MHC-peptide combinatorics. *Hum Immunol* 62:539–556
- Kangueane P, Sakharkar MK, Rajaseger G et al (2005) A framework to sub-type HLA supertypes. *Front Biosci* 10:879–886
- Laskowski RA (1991) “SURFNET” computer program, Department of Biochemistry and Molecular Biology, University College, London
- Lund O, Nielsen M, Kesmir C et al (2004) Definition of supertypes for HLA molecules using clustering of specificity matrices. *Immunogenetics* 55:797–810
- Morrison RT, Boyd RN (1992) Organic chemistry, vol 6. Prentice Hall, USA
- Ranganathan S, Tong JC, Tan TW (2008) Structural immunoinformatics. In: Schonbach C, Ranganathan S, Brusci V (eds) Springer, Immunomics Reviews Series, chap. 3, pp 51–61
- Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* 234:774–815
- Samanta U, Bahadur RP, Chakrabarti P (2002) Quantifying the accessible surface area of protein residues in their local environment. *Prot Eng* 15:659–667
- Samudrala R, Moult J (1997) Handling context-sensitivity in protein structures using graph theory: Bona fide prediction. *Proteins* 1(Suppl):43–49
- Sette A, Livingston B, McKinney D et al (2001) The development of multi-epitope vaccines: Epitope identification, vaccine design and clinical evaluation. *Biologicals* 29:271–276
- Sette A, Newman M, Livingston B et al (2002) Optimizing vaccine design for cellular processing, MHC binding and TCR recognition. *Tissue Antigens* 59:443–451
- Stewart-Jones GB, McMichael AJ, Bell JI et al (2003) A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* 4:657–663
- Tong JC, Bramson J, Kanduc D et al (2006a) Modeling the bound conformation of pemphigus vulgaris associated peptides to MHC class II DR and DQ alleles. *Immunome Res* 2:1
- Tong JC, Kong L, Tan TW et al (2006b) MPID-T: Database for sequence-structure-function information on TCR-peptide-MHC interactions. *Appl Bioinform* 5:111–114
- Tong JC, Tan TW, Ranganathan S (2004) Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein Sci* 13:2523–2532
- Tong JC, Tan TW, Sinha AA et al (2006c) Prediction of desmoglein-3 peptides reveals multiple shared T-cell epitopes in HLA DR4- and DR6-associated pemphigus vulgaris. *BMC Bioinform* 18:7
- Tong JC, Tan TW, Ranganathan S (2007a) Methods and protocols for prediction of immunogenic epitopes. *Brief Bioinform* 8:96–108
- Tong JC, Tan TW, Ranganathan S (2007b) *In silico* grouping of peptide/HLA class I complexes using structural interaction characteristics. *Bioinformatics* 23:177–183
- Tong JC, Zhang GL, August JT et al (2006d) Prediction of HLA-DQ3.2 $\beta$  Ligands: Evidence of multiple registers in class II binding peptides. *Bioinformatics* 22:1232–1238
- Tong JC, Zhang GL, August JT et al (2007c) *In silico* characterization of immunogenic epitopes presented by HLA-Cw\*0401. *Immunome Res* 3:7
- Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: A program to generate schematic diagrams of protein–ligand interactions. *Prot Eng* 8:127–134
- Yewdell JW, Bennink JR (1999) Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* 17:51–88