

High-resolution structure of HLA-A*1101 in complex with SARS nucleocapsid peptide

Thomas Blicher,^{a,b}
Jette Sandholm Kastrup,^a Søren
Buus^b and Michael Gajhede^{a*}

^aBiostructural Research, Department of Medicinal Chemistry, The Danish University of Pharmaceutical Sciences, Universitetsparken 2, DK-2100 Copenhagen, Denmark, and

^bDepartment of Medical Microbiology and Immunology, The Panum Institute, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen, Denmark

Correspondence e-mail: mig@dfuni.dk

The structure of the human MHC-I molecule HLA-A*1101 in complex with a nonameric peptide (KTFPPTPEPK) has been determined by X-ray crystallography to 1.45 Å resolution. The peptide is derived from the SARS-CoV nucleocapsid protein positions 362–370 (SNP_{362–370}). It is conserved in all known isolates of SARS-CoV and has been verified by *in vitro* peptide-binding studies to be a good to intermediate binder to HLA-A*0301 and HLA-A*1101, with IC₅₀ values of 70 and 186 nM, respectively [Sylvester-Hvid *et al.* (2004), *Tissue Antigens*, **63**, 395–400]. In terms of the residues lining the peptide-binding groove, the HLA-A*1101–SNP_{362–370} complex is very similar to other known structures of HLA-A*1101 and HLA-A*6801. The SNP_{362–370} peptide is held in place by 17 hydrogen bonds to the α -chain residues and by nine water molecules which are also tightly bound in the peptide-binding groove. Thr6 of the peptide (Thr6p) does not make efficient use of the middle (E) pocket. For vaccine development, there seems to be a potential for optimization targeted at this position. All residues except Thr2p and Lys9p are accessible for T-cell recognition.

Received 4 February 2005

Accepted 25 April 2005

PDB Reference:

HLA-A*1101–SNP_{362–370}
complex, 1x7q, r1x7qsf.

1. Introduction

Major histocompatibility complex class I (MHC-I) molecules play a pivotal role in the recognition of intracellular pathogens (Klein *et al.*, 1993). They bind peptides derived from cytosolic proteins and transport them to the cell surface. This enables cytotoxic T lymphocytes (CTLs) to monitor the metabolic status of every nucleated cell in the body. Infected (or transformed) cells displaying foreign (or aberrant) peptides bound to their surface MHC-I molecules can be identified and subsequently lysed by CTLs, thus eliminating infections (or potentially cancerous cells; Townsend & Bodmer, 1989).

MHC-I molecules consist of three chains: the polymorphic transmembrane α -chain (365 amino acids), the monomorphic β -chain (β_2m , 99 amino acids) and a peptide of 8–11 amino-acid residues. Assembly of the complexes takes place in the ER, where the process is controlled by a number of chaperones and the TAP peptide transporter (Cresswell, 2000). The extracellular part of MHC-I molecules has a very characteristic structure. The membrane-proximal β_2m/α_3 immunoglobulin domains serve as a rigid support for the membrane-distal α_1/α_2 -domains consisting of two α -helices on top of a β -sheet. Here, peptides are bound to MHC-I molecules between the two α -helices through interactions with polymorphic as well as conserved residues in the α -helices and the β -sheet floor. Several invariant interactions with the peptide

backbone, especially the N- and C-terminus, stabilize the peptide–MHC-I complex (Bjorkman *et al.*, 1987), but the ultimate determinant of the half-life of MHC-I complexes seems to be sequence-dependent interactions with the so-called specificity pockets in the groove. The peptide-binding groove is conventionally subdivided into six such specificity pockets named A to F (Garrett *et al.*, 1989). Each pocket can accommodate the side chain of one or a few closely related amino-acid residues, thus determining the peptide-specificity of the MHC-I molecule. The peptide residues preferred by a given MHC-I molecule constitute what is known as a sequence motif. Variation in the pool of bound peptides not defined by sequence motifs includes attachment of non-peptide groups to peptides (Otvos *et al.*, 1996; Speir *et al.*, 1999) as well as binding of extended or shortened peptides (Stryhn *et al.*, 2000). Currently, about 70% of all peptides binding to a given MHC-I molecule can be identified using extended motif definitions (Kast *et al.*, 1994).

Although each MHC-I allele has distinct peptide-binding properties, it has been suggested that most MHC-I alleles can be clustered into a few supertypes represented in all ethnic groups. So far, a comparison of sequence motifs has allowed the extraction of 9–12 (Lund *et al.*, 2004; Sette & Sidney, 1999) supertypes out of the 975 different MHC-I A and B alleles (January 2005; see <http://www.anthonynolan.org.uk/HIG/index.html>) and more than 99% of the world population carry at least one of these supertypes (Sette & Sidney, 1999). Hence, it seems that the number of alleles that have to be considered can be reduced. This may represent a significant simplification of the requirements for the design of peptide-based vaccines with broad population coverage.

HLA-A*1101 (A11) is one of the most frequently occurring MHC-I alleles, with an allele frequency of up to 27% in some Southeast Asian populations (Bodmer *et al.*, 1999). It is known to be important in the control of HIV (Culmann *et al.*, 1991; Johnson & Walker, 1991; Walker *et al.*, 1989), EBV (Gavioli *et al.*, 1993) and HBV infections (Achour *et al.*, 1986). A11 belongs to the A3 supertype, which is the second most common supertype; at least 44% of the world population carry A3 supertype MHC-I molecules (Sette & Sidney, 1999). A3 supertype molecules are characterized by a preference for small and/or hydrophobic residues in the second position of the peptide and arginine or lysine at the C-terminus (Sidney *et al.*, 1996). Specifically, A11 preferentially binds peptides with isoleucine, threonine or valine in position 2 and a C-terminal lysine or arginine residue (Falk *et al.*, 1994; Gavioli *et al.*, 1995; Kubo *et al.*, 1994; Zhang *et al.*, 1993). Structures of MHC-I molecules belonging to the A3 supertype improve our understanding of peptide binding to these molecules. Furthermore, such models will enable the construction of reliable A3 supertype homology models with any desired peptide, an approach which promises to become a useful way of predicting peptide–MHC-I affinities in the near future (Flower *et al.*, 2003; Logean *et al.*, 2001; Rognan *et al.*, 1999; Schueler-Furman *et al.*, 2000).

Although A3 is one of the most common supertypes, only a few structures belonging to the A3 family have been published

to date. Three structures of HLA-A*6801 (A68) exist. In one, the peptide identity is unknown (PDB code 2hla; resolution 2.6 Å; Garrett *et al.*, 1989), in another the peptide is partially missing (PDB code 1hsb; resolution 1.9 Å; Guo *et al.*, 1992) and in the third one the α_3 domain was proteolytically removed prior to crystallization (PDB code 1tmc; resolution 2.3 Å; Collins *et al.*, 1995). In addition, two complete structures of HLA-A*1101 have been published recently (Li & Bouvier, 2004; PDB codes 1q94 and 1qvo; resolution 2.4 and 2.22 Å, respectively). In this study, we describe the first high-resolution X-ray structure of HLA-A*1101 in complex with a peptide derived from SARS coronavirus (CoV) nucleocapsid protein (see below). This peptide has been suggested as a putative vaccine candidate (Sylvester-Hvid *et al.*, 2004).

1.1. SARS

The first known case of Severe Acute Respiratory Syndrome (SARS) occurred in the Guangdong Province in southeast China in November 2002 (Cherry, 2004). At the beginning of 2003, a large number of cases appeared and SARS quickly spread to more than 30 countries all over the world. By March 2003, the causative agent had been identified as a novel member of the coronavirus family (order *Nidovirales*, family *Coronaviridae*, genus *Coronavirus*). This class of viruses is responsible for up to 30% of all mild upper respiratory tract illnesses in humans (Siddell, 1995). To date, over 8400 cases of SARS CoV-infected patients have been identified worldwide. Approximately 10% of all infected individuals died from progressive respiratory failure (Centers for Disease Control and Prevention, 2003) and in elderly people the mortality was as high as 55%. Since the initial outbreak, the number of reported cases has dropped significantly; however, SARS-CoV is far from eradicated. Of special concern is the existence of SARS-CoV in several non-human species. Indeed, evidence suggests a non-human origin of SARS-CoV (reviewed in Poon *et al.*, 2004). The consequences of a worldwide SARS pandemic are grim and the World Health Organization still considers the development of a vaccine against SARS-CoV to be of great importance.

1.1.1. The N protein. The product of the SARS N gene, the nucleocapsid or N protein, is a protein of 46 kDa (422 amino-acid residues). It interacts with the viral RNA and presumably forms a helical arrangement in the centre of the virus particle as do N proteins from other coronaviruses (Siddell, 1995). Furthermore, the N protein has been shown to induce actin rearrangement and apoptosis in cells under stress (Surjit, Liu, Jameel *et al.*, 2004). A highly basic sequence unique to SARS-CoV in the C-terminal part of the protein (residues 362–381, KTFPPTPEPKDKKKKTDEAQ) resembles a nuclear localization signal (Marra *et al.*, 2003). Although its function remains unknown, the high pI of the peptide suggests binding to RNA. In addition, the C-terminal half of the protein is known to contain a dimerization domain (Surjit, Liu, Kumar *et al.*, 2004), the function of which remains to be elucidated.

2. Materials and methods

2.1. Cloning

A*1101 α -chain was cloned from the B cell line KHAGNI (European Collection of Cell Cultures) grown in RPMI with 10% FCS and 1% penicillin/streptomycin at 310 K, 5.0% CO₂. mRNA was extracted using a combination of two mRNA-purification kits: the RNA Isolation Kit from Stratagene and the RNeasy kit from Qiagen. The mRNA was reverse-transcribed with the GeneAmp RNA PCR kit from Perkin-Elmer and the DNA encoding the 275 N-terminal (extracellular) amino acids of the A*1101 α -chain was amplified using Vent_R DNA polymerase (New England Biolabs) and primers purchased from Hobolth DNA Syntese (Allerød, Denmark). The amplified DNA was purified (QIAquick PCR Purification kit) and cut with *Xho*I and *Nco*I restriction enzymes (New England Biolabs). Digested A*1101 DNA was ligated into a pET-28a vector using T4 DNA ligase (New England Biolabs) and transformed into chemically competent *Escherichia coli* BL21-DE3. The correctness of the DNA was verified by sequencing.

2.2. Protein preparation

Recombinant human β_2m and A*1101 α -chain were generated according to protocols described in previous work (Sidney *et al.*, 1996). Briefly, β_2m and A*1101 α -chain were produced in *E. coli* as inclusion bodies. Following resolubilization of the inclusion bodies in 8.0 M urea, the crude protein was purified by size-exclusion and ion-exchange chromatography. β_2m was further purified and refolded on an Ni-chelating column by virtue of an N-terminal His₆ tag. Prior to use, the His₆ tag was removed by cleavage with factor Xa followed by size-exclusion chromatography. This β_2m is indistinguishable from β_2m purified from natural sources (Pedersen *et al.*, 1995).

Folding of HLA-A*1101 complexes was performed by rapid 100-fold dilution of A11 solution (approximately 200 μ M/0.6 mg ml⁻¹) into a folding buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 3 mM EDTA, 1.0 mM β_2m and 2.0 mM SARS-CoV nucleocapsid peptide (residues 362–370). The folding solution was left at 291 K for 16–24 h and then concentrated to 1/20 of its original volume. After another 16–24 h at 277 K, misfolded protein was removed by centrifugation at 5000g for 30 min at 277 K. The supernatant was further concentrated to 1/20 of its original volume, left for another 16–24 h and finally centrifuged at 20 000g for 20 min at 277 K. The supernatant was applied onto a Superdex 75 column and eluted with a buffer containing 150 mM NaCl and 100 mM MES pH 6.5. Fractions containing the HLA-A*1101 complexes were concentrated after buffer exchange (0.1 M MES pH 6.5) to 6.5 mg ml⁻¹. The complex was purified to more than 95% purity (estimated by SDS-PAGE; data not shown).

The SARS peptide stock used in the preparation of the A11 complexes was purchased from Schafer-N (Copenhagen) as >99% HPLC-purified preparations. The peptide was synthesized using Fmoc chemistry.

Table 1

Data-collection and refinement statistics.

Values in parentheses refer to the highest resolution shell.

Data collection	
Space group	<i>P</i> 2 ₁
Resolution range (Å)	30–1.45 (1.48–1.45)
Completeness (%)	100 (100)
Total observations	265525 (16075)
Unique reflections	81925 (4079)
Average <i>I</i> / σ (<i>I</i>)	15.9 (3.1)
<i>R</i> _{sym} † (%)	0.104 (0.446)
Model refinement	
<i>R</i> _{cryst} ‡ (%)	13.3/12.3§
<i>R</i> _{free} ¶ (%)	18.1/16.8§
No. of atoms/average <i>B</i> factor (Å ²)	
Protein atoms	3224/17.3
Water molecules	574/32.1
Sulfate ions	15/25.8
Glycerol molecules	42/42.8
Ramachandran plot††	
Most favoured (%)	94.0
Disallowed (%)	0.0

† *R*_{sym} = $\sum_{h,i} |I_{h,i} - \langle I_h \rangle| / \sum_{h,i} I_{h,i}$. ‡ *R*_{cryst} = $\sum_{h,i} |F_o - F_c| / \sum F_o$. § Data with $F_o/\sigma(F_o) > 4$. ¶ *R*_{free} as *R*_{cryst} but calculated using 5% of the data excluded from refinement. †† Calculated with *PROCHECK* (Laskowski *et al.*, 1993).

2.3. Crystallization and data collection

Crystals of HLA-A*1101 in complex with the SNP_{362–370} peptide were prepared using the hanging-drop technique. Crystallization experiments were set up at 293 K with individual drops consisting of 2 μ l protein solution (6 mg ml⁻¹ in 0.1 M MES pH 6.5) and 2 μ l reservoir solution. The crystals grew from a crystallization buffer containing 30% PEG 5000 MME, 0.2 M ammonium sulfate (Crystal Screen 2, Hampton Research), appearing as clusters of monoclinic plates after four weeks. These initial crystals were used to streak-seed new drops and produced single crystals of final dimensions 0.2 × 0.1 × 0.05 mm in three weeks. Single crystals were collected, soaked in reservoir solution containing 5% glycerol as a cryoprotectant and flash-frozen in liquid nitrogen. Two data sets (36.0–2.54 and 15.0–1.45 Å resolution) were collected on the same crystal at 100 K at a wavelength of 0.9815 Å and were processed with the *HKL* suite (Otwinowski & Minor, 1997), giving the data statistics outlined in Table 1. Data were collected at beamline 14-2 at BESSY (Berlin, Germany) using a MAR345 image-plate detector. The space group is *P*2₁, with unit-cell parameters *a* = 58.5, *b* = 80.6, *c* = 56.6 Å, β = 116.8° and one complex per asymmetric unit.

2.4. Structure determination and refinement

The structure of the A11–SNP_{362–370} complex was solved by molecular replacement (MR) in *AMoRe* (Navaza, 1994) using the structure of HLA-A*6801 without peptide and water molecules (PDB code 1hsb; Guo *et al.*, 1992) as a search model. The program *ARP/wARP* was used to remove bias from the MR model and subsequently managed to unambiguously exchange the polymorphic side chains and to build the peptide. All subsequent refinements were carried out

with *SHELXL* (Sheldrick & Schneider, 1997), with 5% of the data reserved for the calculation of R_{free} .

The model after *ARP/wARP* rebuilding included all 275 residues of the α -chain, all 99 residues of $\beta_2\text{m}$, 171 water molecules and the peptide. The R value was 22.4% and R_{free} was 25.8%. Following the addition of 204 water molecules, minor side-chain adjustments using program *O* (Jones *et al.*, 1991) and a round of positional and B -factor refinement, R and R_{free} fell to 20.4 and 24.2%, respectively. Introduction of another 210 water molecules (now 585 in total) along with seven glycerol molecules gave R and R_{free} values of 19.5 and 23.6%, respectively. The resolution of the data allowed all

atoms to be refined anisotropically. After anisotropic B -factor refinement and deletion of 28 water molecules, R and R_{free} were 16.1 and 21.3%, respectively. Subsequent addition of riding H atoms resulted in R and R_{free} values of 14.1 and 19.0%, respectively. Cleaning up the water molecules and a few more rounds of positional and B -factor refinement gave a final R factor of 13.3% and an R_{free} of 18.1% (see Table 1). No B -factor cutoff was used during the last rounds of water refinement. Rather, all water molecules with clear $2F_o - F_c$ density at 1σ and potential hydrogen-bonding partners were retained. The following 20 residues were modelled with double conformations: Ser4 α , Ser11 α , Arg35 α , Arg82 α , Asn86 α , Ser88 α , Met98 α , Thr134 α , Gln141 α , Arg145 α , Gln156 α , Arg157 α , Thr216 α , Lys243 α , Ser251 α , Gln255 α , Thr4 β , Ser20 β , Val27 β and Pro4p. There are no direct contacts between the peptide and neighbouring molecules in the crystal. The closest contacts arise from packing of $\beta_2\text{m}$ from a neighbouring molecule against the outside of the middle part of the α_1 helix.

3. Results and discussion

The SARS-CoV N protein peptide (KTFPPTEPK; SNP_{362–370}) was crystallized in complex with HLA-A*1101. The structure was determined at 1.45 Å resolution to an R factor of 13.3% ($R_{\text{free}} = 18.1\%$). The final model contains all 275 residues of the extracellular domain of A11 (α -chain), all 99 residues of $\beta_2\text{m}$ (β -chain), the SNP_{362–370} peptide (p chain), seven glycerol molecules, three partially occupied sulfate ions (65, 57, 62%) and 574 water molecules (see Fig. 1 for an overview). The Ramachandran plot calculated with *PROCHECK* (Laskowski *et al.*, 1993) shows 94.0% of all residues in the most favoured regions, 5.7% in additional allowed regions and only Asp29 of the α -chain in a generously allowed region. The unusual but well defined conformation of Asp29 has been reported for other MHC-I molecules (Hillig *et al.*, 2001).

3.1. The A3 supertype molecules

A comparison of the A11–SNP_{362–370} structure with the A11–Nef and A11–RT structures (Li & Bouvier, 2004) reveals very similar overall conformations, with only minor variations in the angles between the domains. These minor variations reflect the inherent flexibility of MHC-I molecules as well as differences in crystal packing. Despite differences in the identity of the bound peptide, most residues lining the peptide-binding groove are in almost identical positions (Fig. 2). The only differences are seen at the edge of the binding groove, where side chains from the peptide allow different sets of interactions with otherwise solvent-exposed residues of the α -chain. However, the contribution of these mainly polar interactions to the peptide binding is probably negligible, as they are easily replaced with solvent interactions. Minor local displacements of the α_1 and α_2 helices caused by the peptide can also be discerned and are of the order of 0.5–1.0 Å. These displacements centre on the middle part of the α_1

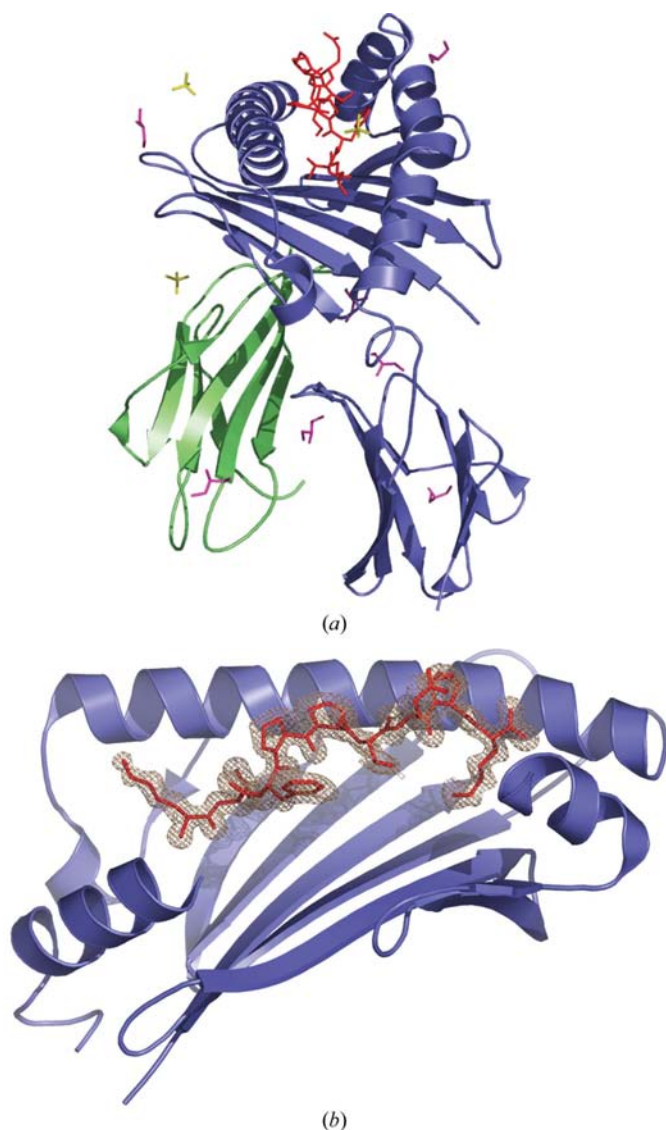


Figure 1
Schematic representation of HLA-A*1101 in complex with the SNP_{362–370} peptide. The α -chain is shown in blue, $\beta_2\text{m}$ in green, the peptide in red, sulfate ions in yellow and glycerol molecules in magenta. (a) Overview, peptide N-terminus closest to the viewer. (b) Close-up view of the $\alpha_1\alpha_2$ domain holding the SNP_{362–370} peptide. Part of the α_2 helix has been removed to allow a better view of the peptide electron density. The $2F_o - F_c$ electron density surrounding the peptide is contoured at 1σ . The peptide N-terminus is on the left. All figures were prepared with *PyMOL* (DeLano, 2002).

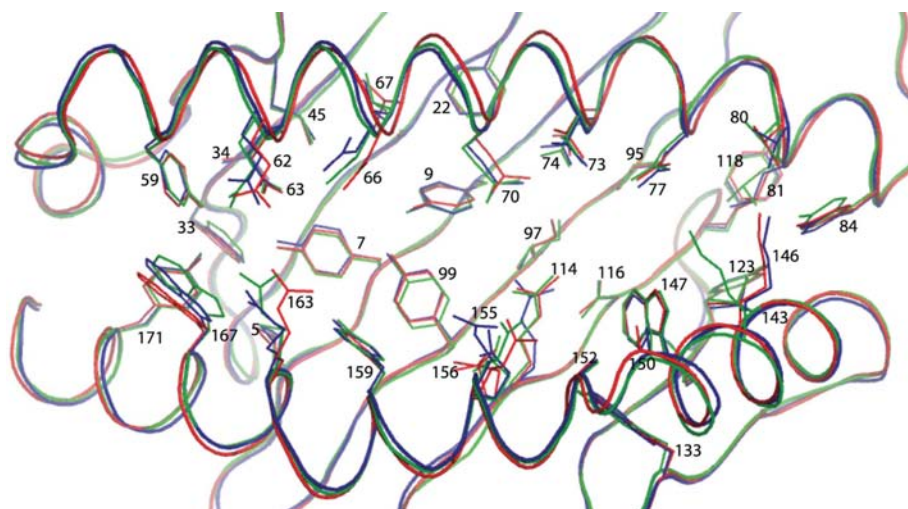


Figure 2
Comparison of the peptide-binding grooves of three A11 structures. The A11-SNP₃₆₂₋₃₇₀ structure is shown in red, A11-Nef in blue and A11-RT in green (Li & Bouvier, 2004). Residues facing the peptide-binding groove are in almost identical positions or adopt similar rotamer conformations.

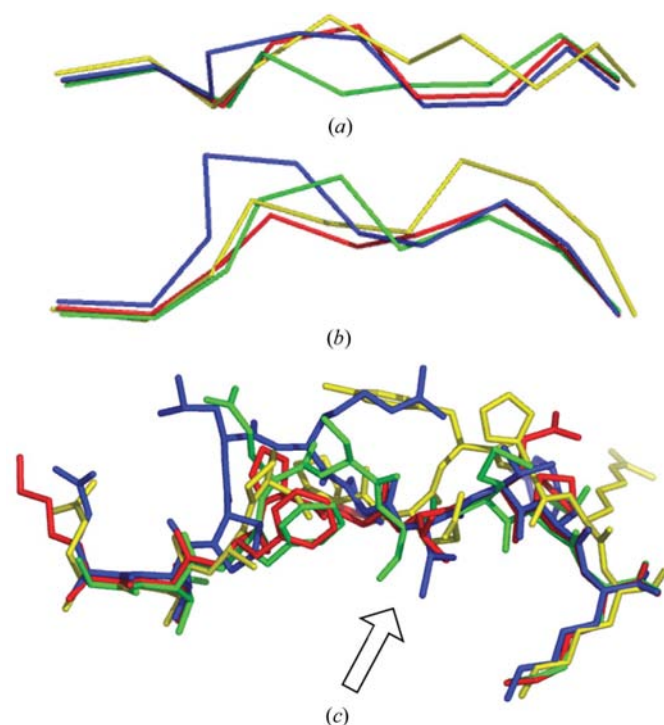


Figure 3
Peptide conformations. Superposition of four peptides bound to A3 supertype MHC-I molecules, N-termini on the left. The nonameric SNP₃₆₂₋₃₇₀ peptide is shown in red, the decameric Nef (A11; Li & Bouvier, 2004) in blue, the nonameric RT (A11; Li & Bouvier, 2004) in green and an endogenously derived decameric peptide (A68; Collins *et al.*, 1995) in yellow. (a) C α trace, top view. (b) C α trace, viewed from the α_2 -helix side of the binding groove. (c) All atoms, same view as (b). Residues binding in the E pocket are marked by an arrow.

helix (B/C pockets) and the N-terminal half of the α_2 helix (D/E pockets).

A comparison of the A11-SNP₃₆₂₋₃₇₀ structure with the three existing A68 structures shows the same trends as the

comparison with existing A11 structures. Those residues in the binding groove that differ between the two alleles (A11 \leftrightarrow A68: Q62R, E63N, I97M, A152V, Q156W, R163T) occupy approximately the same space. The only significant difference is found at the bottom of the central part of the binding groove, where A68 residues Met97 α and Trp156 α , which are sitting opposite each other, are more bulky than the corresponding A11 residues Ile97 α and Gln156 α . The A11 \leftrightarrow A68 change in position 156 to a less polar environment (Q \leftrightarrow W) in addition to the increase in size makes the outermost part of the side chain of the shared Arg114 α flip over, which in turn forces the side chain of Asp116 α to adopt another conformation. These differences affect the shape and charge

geometry at the bottom of the peptide-binding groove (F pocket) and might explain why A68 has a slight preference for arginine over lysine at the peptide C-terminal end, while the opposite is true for A11 (see the SYFPEITHI database at <http://www.syfpeithi.de/>; Rammensee *et al.*, 1999). The long side chain of an arginine residue requires a deeper and wider pocket in order to maintain the conserved network of hydrogen bonds surrounding the peptide C-terminus and this is achieved by substituting Ile97 α in A11 with methionine in A68. Although other residue differences do exist between A11 and A68 in the $\alpha_1\alpha_2$ region, they are all found on the outside far away from the peptide and are not likely to be important for peptide binding.

In general, the positions of the peptide N- and C-termini are independent of the peptide length. Variations in length are accommodated in the central part of the peptide-binding groove where the peptide adjusts itself to allow optimal interactions between the primary anchor residues (P₂ and P_C for the A3 supertype) and the corresponding peptide-binding pockets (B and F, respectively). Short peptides (octamer) optimize their interactions with the B and F pockets by adopting an extended conformation (Reiser *et al.*, 2000, 2002) compared with nonamer or decamer peptides, which adopt various forms of bulging or zigzagging conformations in the central part of the groove (Hillig *et al.*, 2001). The straight conformation tends to place the octameric peptides deeper in the groove. Thus, owing to steric interference caused by Met97 α and Trp156 α at the floor of the peptide-binding groove, A68 would be expected to have a stronger preference for nonamers or decamers than A11. The set of A11 and A68 example peptides in the SYFPEITHI database confirms this tendency (Rammensee *et al.*, 1999). The other members of the A3 supertype would be expected to have the same preferences as A11 as they all (A3, A31 and A33) have a leucine residue in position 156 and either isoleucine (A3 and A11) or methionine (A31 and A33) in position 97.

3.2. Peptide-binding interactions

A superposition of the A11 structure with that of other MHC-I complexes demonstrates that the SNP_{362–370} peptide adopts a standard nonameric conformation with the canonical positioning of the N- and C-terminal residues (see Fig. 3). The interactions between the SNP_{362–370} peptide and the A11 binding groove are summarized in Table 2 and shown in Fig. 4. The SNP_{362–370} peptide is held in place by 17 hydrogen bonds to the α -chain residues and by additional hydrogen bonds to nine water molecules which are also tightly bound in the peptide-binding groove. A total of 13 water molecules occupy key positions in the groove, being either directly involved in peptide binding or contributing to the specific architecture of the various binding pockets. Seven of the water molecules (W19, W46, W87, W152, W159, W178 and W283) form a hydrogen-bonded network starting at the C pocket and ending at the amino group of the C-terminal lysine residue located in the F pocket. A detailed description of the peptide-binding interactions is given below.

3.2.1. The A pocket. The backbone N atom of Lys1p forms the characteristic hydrogen bonds to Tyr7 α and Tyr171 α and the N-terminal carbonyl O atom makes a hydrogen bond to Tyr159 α . The aliphatic part of the Lys1p side chain is engaged in van der Waals (vdW) interactions with the side chains of Tyr59 α and Trp167 α , and the distal amino group (N⁵) forms hydrogen bonds to Gln62 α as well as to a water molecule (W421) and a sulfate ion (4003). A conserved water molecule (W17) stabilizes the A pocket by hydrogen-bond formation to the carboxyl group of Glu63 α and to the hydroxyl groups of Tyr7 α and Tyr59 α . The latter in turn forms a hydrogen bond to the hydroxyl group of Tyr171 α .

3.2.2. The B pocket. Thr2p is fixed at the entrance to the B pocket *via* hydrogen bonds from the hydroxyl group to the side chains of Glu63 α and Asn66 α . Another hydrogen bond from Glu63 α to the Thr2p amide NH group, in addition to vdW interactions with the side chains of Tyr7 α , Tyr9 α , Asn66 α and Tyr99 α , also at the pocket entrance, further restricts the flexibility of Thr2p. The architecture of the entrance to the B pocket seems uniquely suited to accommodate amino-acid residues branched at C ^{β} . Indeed, A11 prefers peptides with isoleucine, valine and threonine in the second position (Gavioli *et al.*, 1995), but other aliphatic residues such as methionine, leucine, alanine and serine are also accepted (Falk *et al.*, 1994; Sidney *et al.*, 1996). Met45 α and Val67 α define the bottom of the B pocket and since Thr2p does not fill the pocket entirely, these residues would have to rearrange slightly to accommodate larger side chains. Opposite the B pocket, Thr2p rests against Tyr159 α and forms a hydrogen bond to the Arg163 α side chain *via* the carbonyl group. Finally, a water molecule (W6) forms hydrogen bonds to the side chains of Glu63, Asn66 α and Arg163 α as well as to the Thr2p carbonyl group. Together, these interactions completely bury Thr2p in the peptide-binding groove.

3.2.3. The C pocket. Pro4p and Pro5p are bound in a similar solvent-exposed manner: Pro4p in the region between the B and C pockets formed by the α_1 helix and Pro5p in the C

pocket. Both residues stack against the face of a side-chain amide group (Asn66 α and Gln70 α , respectively). Hydrogen bonds are seen between their carbonyl groups and bound water molecules (Pro4p to W296, Pro5p to W134 and W136 and water molecules W46 and W283 to both). Pro5p makes additional vdW interactions with Ala69 α , Thr73 α and the backbone of Gln70 α . The open nature of the C pocket allows a broad range of residues to be bound.

3.2.4. The D pocket. The D pocket is wide and imposes few restrictions on the nature of the peptide side chain in position 3. Together with the C pocket, it forms the central open part of the peptide-binding groove. While two tyrosine residues (Tyr99 α and Tyr159 α) allow interactions with aromatic and hydrophobic residues including methionine, the presence of Arg114 α and Gln156 α at the peptide-distal end of the pocket explains how acidic residues and their derivatives (Asp, Glu, Asn and Gln) are also accepted (Falk *et al.*, 1994).

The amide NH group of Phe3p forms a hydrogen bond with the hydroxyl group of Tyr99 α and the carbonyl group forms hydrogen bonds to two water molecules (W46 and W283), which are themselves caught below the peptide in the binding groove. The Phe3p side chain makes vdW interactions with Tyr159 α and Gln156 α , the latter assuming two clearly visible conformations. Since Phe3p does not make use of the hydrogen-bonding abilities of Gln156 α , a water molecule (W178) is inserted between Arg114 α , Gln156 α and W283, pushing the phenyl ring of Phe3p up into a more superficial position close to Pro4p. This means that although the Phe3p backbone is hydrogen-bonded to the peptide-binding groove, the side chain rests on a cushion of water molecules (W178 and W283) and polar residues. The resulting relatively high mobility (*B* factors of 27 Å² for C ^{γ} increasing to 40 Å² for C ^{ϵ}) of the Phe3p side chain indicates that the interactions between Phe3p and the D pocket are rather weak.

3.2.5. The E pocket. The side chain of Thr6p points downwards into the groove of the E pocket, where its hydroxyl group forms a hydrogen bond to a water molecule (W391). The side-chain methyl group is in contact with Arg114 α and Trp147 α . W391 forms additional hydrogen bonds with the Ala152 α amide carbonyl group and one of the two side-chain rotamers of Gln156 α . Residual electron density in the vicinity of the Thr6p hydroxyl group and W391 suggests the presence of a disordered water molecule (possibly W391), in agreement with the relatively high *B* factors of the Thr6p side chain (*B* factors between 24 and 38 Å²). This disorder, which is not explicitly modelled, might also explain the relatively short hydrogen-bonding distance between W391 and Thr6p as it represents the distance between average positions. The backbone NH of Thr6p forms a hydrogen bond with W159 in the binding groove, which is itself hydrogen bonded to W283 and W152 on opposite sides, also in the groove. The Thr6p carbonyl group only binds surface water molecules.

Thr6p seems not to be directly involved in any specific interactions essential for binding of the SNP_{362–370} peptide. Rather, it merely helps to fill the peptide-binding groove, where a moderately sized polar residue such as asparagine or glutamine would probably be able to make more favourable

Table 2
Peptide-binding interactions.

Peptide		Hydrogen-bond partner [†]		Distance (Å)	vdW contact residues [‡]
Residue	Atom	Residue	Atom		
Lys1p	N	Tyr7 α	OH	3.0	Met5 α , Tyr7 α , Tyr59 α , Glu63 α , Trp167 α
	N	Tyr171 α	OH	2.7	
	O	Tyr159 α	OH	2.6	
	N ⁵	Gln62 α	O ⁵¹	3.1	
	N ⁵	(W421)‡	OW	3.1	
Thr2p	N ⁵	(SO ₄ ²⁻ 4003)‡	O2	2.8	Tyr7 α , Tyr9 α , Met45 α , Glu63 α , Asn66 α , Val67 α , Tyr99 α , Tyr159 α
	N	Glu63 α	O ⁵¹	2.9	
	O	Arg163 α	NH2	3.1	
	O	W6	OW	2.7	
	O ¹	Glu63 α	O ⁵¹	2.8	
Phe3p	O ¹	Asn66 α	N ⁸²	3.0	Gln156 α , Tyr159 α , W178, W283, (W296)‡, (W440)‡, Thr6p
	N	Tyr99 α	OH	3.0	
	O	W46	OW	2.7	
Pro4p	O	W283	OW	2.9	Asn66 α
	O	(W134)‡	OW	2.8	
	O	(W136)‡	OW	2.6	
Pro5p	O	(W296)‡	OW	2.8	Asn66 α , Ala69 α , Glu70 α , Thr73 α , (W296)‡
	O	(W296)‡	OW	2.8	
Thr6p	N	W159	OW	2.9	Arg114 α , Trp147 α , Ala152 α , W87, W152, Phe3p
	O	(W85)‡	OW	2.8	
	O ¹	W391	OW	2.3	
Glu7p	O	Trp147 α	N ⁵¹	3.1	Trp147 α , Ala150 α , Phe3p
	O	W87	OW	2.7	
	O ⁵¹	(W300)‡	OW	2.5	
	O ⁵²	(W375)‡	OW	2.6	
Pro8p	O	Trp147 α	N ⁵¹	3.1	Thr73 α , Trp147 α , W501, Glu7p
Lys9p	N	Asp77 α	O ⁵¹	2.9	
Lys9p	OT1	Tyr84 α	OH	2.7	Asp77 α , Leu81 α , Ile95 α , Tyr123 α , Thr143 α , Trp147 α
	OT1	Thr143 α	O ¹	2.7	
	OT2	Lys146 α	N ⁵	2.8	
	OT2	W16	OW	2.7	
	N ⁵	Asp116 α	O ⁸²	2.7	
	N ⁵	W19	OW	2.7	
	N ⁵	W121	OW	3.5	

[†] The cutoff distance of a hydrogen bond is 3.5 Å; vdW contact residues are within 4.0 Å of any SNP_{362–370} peptide atom. [‡] Residues in parentheses are judged not to be important for peptide binding.

interactions with the side chains of Arg114 α , Trp147 α and Gln156 α and the backbone of Ala152 α . In the two other published structures of A11, either a serine (AIFQSSMTK) or methionine residue (QVPLRPMTYK) fills part of the E pocket from the P_{C-3} position (see also Fig. 3c). In analogy with the C pocket, the E pocket also accepts a rather broad range of amino acids owing to its open nature and residues of inadequate size are supplemented with water molecules.

The most solvent-exposed residue of the entire peptide is Glu7p. The backbone NH group approaches the Thr6p hydroxyl group to form a weak hydrogen bond and the backbone carbonyl group forms hydrogen bonds with the indole NH of Trp147 α and a water molecule (W87). The side chain makes vdW interactions with the side chains of Trp147 α and Ala150 α as well as with Pro8p. A water molecule binds between the carbonyl group of Pro8p and the carboxylate group of Glu7p.

Pro8p is bound in a manner similar to that of Pro4p and Pro5p, with its carbonyl group forming a hydrogen bond to a surface water molecule (W300). The rest of the residue makes no direct interactions with the peptide-binding groove. The absence of restraints on the side chain allows Pro8p to switch

between two conformations. The carbonyl group also forms a very weak hydrogen bond with the Trp147 α indole NH group. In fact, the indole NH group is well outside the amide plane of the Glu7p-Pro8p amide bond and seems to interact with the amide-bond π electrons. A similar interaction is seen between the Trp147 α indole NH group and the carbonyl group of the Thr6p-Glu7p amide group. However, this bifurcated hydrogen bond is far from ideal hydrogen-bond geometry and its contribution to peptide binding is probably correspondingly low.

3.2.6. The F pocket. The prime determinant of peptide binding is the interaction between the C-terminal residue and the F pocket, which in the case of A11 is tailored to fit lysine (and arginine) residues. This is evident from the large number of interactions to the backbone and the side chain of Lys9p, which fits in between the negatively charged residues Asp74 α , Asp77 α and Asp116 α .

The carboxylate group of Asp77 α forms a direct hydrogen bond to the Lys9p amide NH group and an indirect bond to the carboxylate and N⁵ groups through water molecules W16 and W19, respectively. W16, which also makes a hydrogen bond to the Thr80 α hydroxyl group, is not merely a surface-bound water molecule but an integral part of

the F pocket and its *B* factor is as low as that of the Lys9p carboxylate (~ 14 Å²). The carboxylate of Lys9p also forms a number of direct hydrogen bonds to residues Tyr84 α , Thr143 α and Lys146 α of the peptide-binding groove.

The N⁵ group is engaged in hydrogen bonding to Asp116 α and two water molecules, W19 and W121, the first forming part of a hydrogen-bonded string of water molecules which stretches all the way back to Phe3p in the C pocket. W19 and W121 are further bound to the side chain of Asp74 α and the carbonyl group of Ala117 α , respectively. Finally, the aliphatic part of Lys9p makes vdW interactions with Asp77 α , Leu81 α , Ile95 α , Tyr123 α and Trp147 α . Interestingly, the C⁵–N⁵ part is the most flexible part of Lys9p (C⁵, 22.7 Å²; N⁵, 22.3 Å²; the remainder, 12–15 Å²), despite the large number of interactions holding the side chain. This is also evident from a slight fan-like smearing of the corresponding electron density.

4. Implications for vaccine design

In principle, all residues except Thr2p and Lys9p are directly accessible for T-cell recognition (Fig. 5). However, the side chains of Phe3p and Thr6p are so deeply placed in the binding

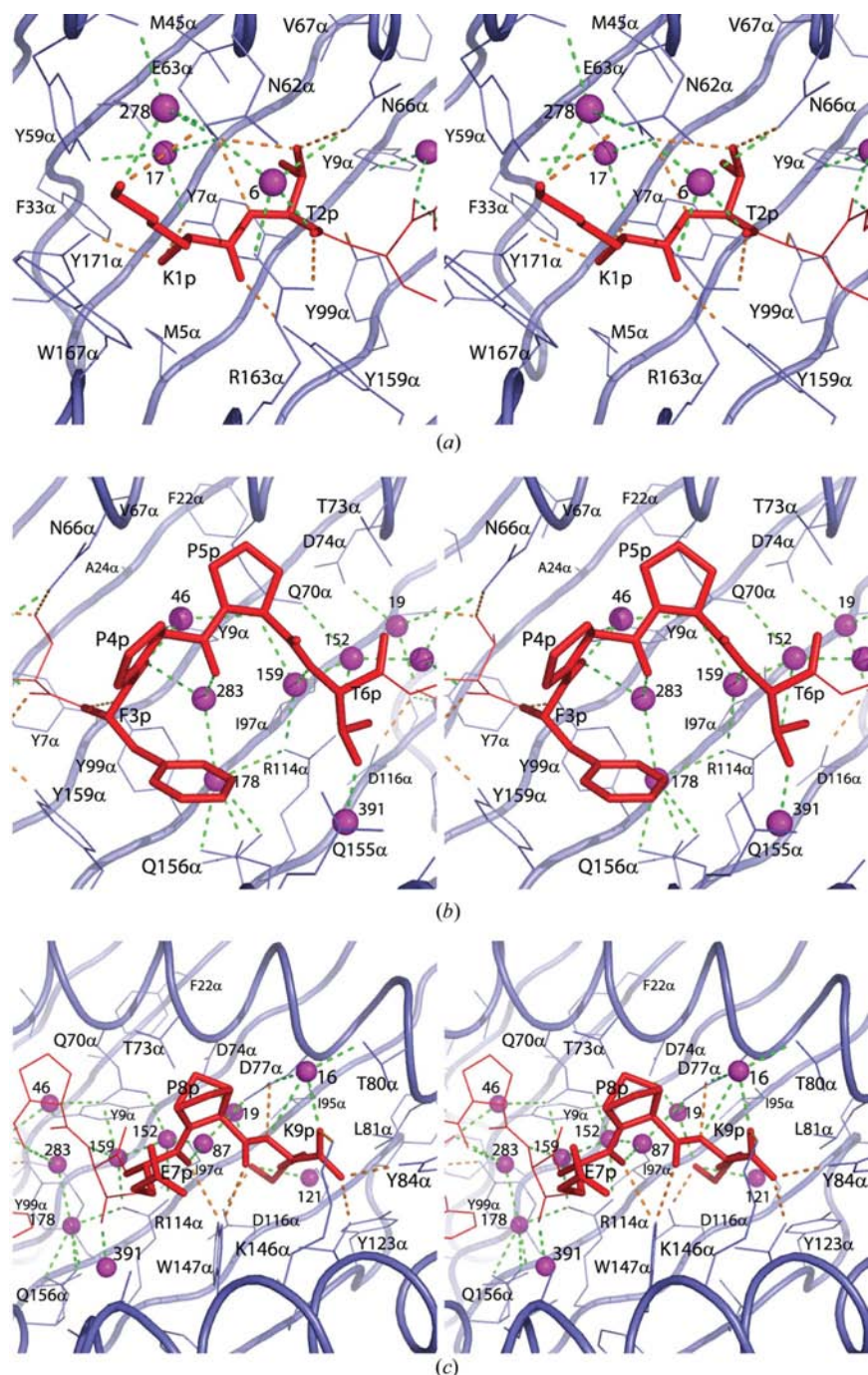


Figure 4 Peptide-binding pockets shown in stereo. The SNP_{362–370} peptide is shown in red, the α -chain in blue and water molecules involved in peptide binding in purple. Hydrogen bonds between the peptide and the α -chain are shown as dashed orange lines, while hydrogen bonds involving water molecules are green. The central part of the peptide is highlighted in bold stick representation in each panel. (a) The A and B pockets (left half and upper right, respectively) holding Lys1p and Thr2p. (b) The central part of the peptide-binding groove flanked by the C and E pockets (upper/middle left and lower right, respectively) holding Phe3p, Pro4p, Pro5p and Thr6p. Notice the network of water molecules, which stretches below the peptide in the groove. (c) The F pocket (right half) holding Lys9p. Notice the low number of interactions between the α -chain and Pro4p and Pro5p (in b) and Pro8p which is shown in double conformation (in c).

groove that their interaction with the T-cell receptor (TCR) might be limited. The residues in those positions could likely be optimized for A11 binding without severely affecting TCR

binding. The large number of proline residues is probably responsible for the intermediate affinity of the SNP_{362–370} peptide for two reasons. Firstly, they do not contribute any significant hydrogen bonds in the peptide-binding groove. Rather, water molecules are present in the central part of the binding groove to satisfy the polar residues. Secondly, they restrict the flexibility of the peptide, thereby preventing optimization of interactions with the binding groove. This is evident from the increased mobility of the Lys9p side chain, which is usually the most tightly bound part of the peptide (Li & Bouvier, 2004). In this case, however, the proline residues cannot be exchanged, since they are all solvent-exposed and therefore expected to interact directly with the TCR.

In recently published work, the best MHC-I-binding peptides from the SARS-CoV were identified using an artificial neural network approach (Sylvester-Hvid *et al.*, 2004). In addition to MHC-I binding, peptides were selected such that they had a high probability for being generated by proteasomal processing and subsequently translocated into the ER by TAP. The predicted peptides were tested for MHC-I binding in a biochemical binding assay. From this set of verified binders, we selected a peptide with the sequence KTFPPTEPK (the SNP_{362–370} peptide) corresponding to a conserved putative nuclear localization signal (see above). This peptide has an IC₅₀ ($\sim K_d$) value of 70 nM for binding to HLA-A*1101 and 186 nM for binding to HLA-A*0301, placing it in the range of intermediate MHC-I binders. As the SNP_{362–370} peptide is part of a SARS-specific sequence from the N protein which seems to be involved in several functions, it is likely that it is and will remain well conserved among different SARS-CoV strains. Therefore, the SNP_{362–370} peptide is a potential candidate for inclusion in a peptide-based vaccine against SARS.

Based on a comparison of several A3 supertype MHC-I molecules, the SNP_{362–370} peptide is expected to bind to other members of the A3 supertype, as the differences between these appear not to interfere with the SNP_{362–370} binding interactions. Previously published structures of A11 (Li & Bouvier, 2004) and A68 (Collins *et al.*, 1995) have suggested the use of additional anchor residues in the P_{C-3} position, resulting in a double-bulge conformation of the

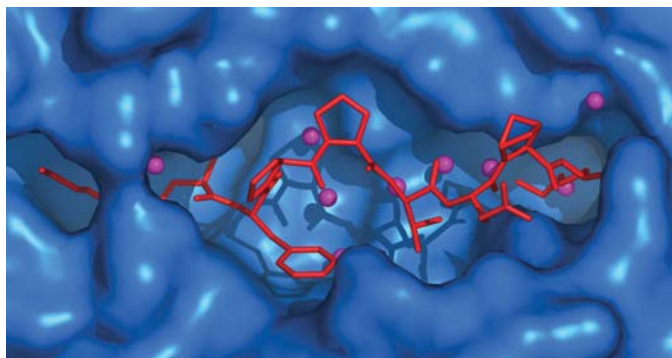


Figure 5

The HLA-A*1101-SNP₃₆₂₋₃₇₀ peptide complex as seen from the T-cell perspective (peptide N-terminus on the left). The α -chain is shown in a blue surface representation and the peptide as red sticks. The surface is drawn with a 1.4 Å probe over the α -chain without water molecules. Water molecules involved in peptide binding are shown in purple. Notice the prominent proline residues in the peptide. All residues except Thr2p and Lys9p are accessible for direct interaction with the TCR.

central part of the peptide (see Fig. 3). In the A11-SNP₃₆₂₋₃₇₀ complex, Thr6p does not make efficient use of this middle (E) pocket and hence there seems to be a potential for optimization targeted at the P₆ position. Ideally, Thr6p should be replaced by a residue which could displace loosely bound water molecules in the groove and make direct contacts with residues lining the D pocket.

The Human MHC Project (Buus, 1999) is an effort to design peptide-based vaccines against any pathogen by means of a complete mapping of MHC-based immune reactivity. As direct screening of every allele is an insurmountable task, there is a need for alternative methods to extend our current knowledge. A promising strategy involves the use of known MHC-I structures as templates for modelling of any desired peptide-MHC-I complex followed by estimation of the interaction energy (*i.e.* affinity) for the complex (Flower *et al.*, 2003). An encouraging fact in this context is the observation of the structural invariance of the peptide-binding groove within a supertype. Although the bound peptides are very different, the A11-SNP₃₆₂₋₃₇₀ complex is very similar to the other two A11 structures in terms of positions of the residues lining the peptide-binding groove. Even A68 is so similar that it should be possible to model it (and presumably A03, A31 and A33 as well) reliably based on the A11 structures. Extrapolating this to the other MHC-I superotypes means that a few well determined structures belonging to each supertype should suffice as a reliable basis for the modelling of any desired MHC-I-peptide complex. The A11-SNP₃₆₂₋₃₇₀ complex structure presented here adds to this expanding list of MHC-I template structures, which hold the information necessary for a direct structural approach to peptide-based vaccine design.

The authors would like to thank Ole Kristensen for assisting with data collection. This work was supported by the Danish Medical Research Council (grant 9601615 and 22-01-0272), the Fifth Framework program of the European Commission

(grant QLRT-1999-00173), the NIH (grant AI49213-02) and the Danish National Research Foundation project DANSYNC.

References

- Achour, A., Begue, B., Gomard, E., Paul, P., Sayagh, B., Van Pel, A. & Levy, J. P. (1986). *Eur. J. Immunol.* **16**, 597–604.
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). *Nature (London)*, **329**, 506–512.
- Bodmer, J., Cambon-Thomsen, A., Hors, J., Piazza, A. & Sanchez-Mazas, A. (1999). In *Proceedings of the Twelfth International Histocompatibility Workshop and Conference*, edited by D. Charron. Paris: EDK.
- Buus, S. (1999). *Curr. Opin. Immunol.* **11**, 209–213.
- Centers for Disease Control and Prevention (2003). *MMWR Morb. Mortal. Wkly Rep.* **52**, 1202–1206.
- Cherry, J. D. (2004). *Paediatr. Respir. Rev.* **5**, 262–269.
- Collins, E. J., Garboczi, D. N., Karpusas, M. N. & Wiley, D. C. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 1218–1221.
- Cresswell, P. (2000). *Traffic*, **1**, 301–305.
- Culmann, B., Gomard, E., Kieny, M. P., Guy, B., Dreyfus, F., Saimot, A. G., Sereni, D., Sicard, D. & Levy, J. P. (1991). *J. Immunol.* **146**, 1560–1565.
- Delano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.
- Falk, K., Rotzschke, O., Takiguchi, M., Grahovac, B., Gnau, V., Stevanovic, S., Jung, G. & Rammensee, H. G. (1994). *Immunogenetics*, **40**, 238–241.
- Flower, D. R., McSparron, H., Blythe, M. J., Zygouri, C., Taylor, D., Guan, P., Wan, S., Coveney, P. V., Walshe, V., Borrow, P. & Doytchinova, I. A. (2003). *Novartis Found. Symp.* **254**, 102–120.
- Garrett, T. P., Saper, M. A., Bjorkman, P. J., Strominger, J. L. & Wiley, D. C. (1989). *Nature (London)*, **342**, 692–696.
- Gavioli, R., Kurilla, M. G., Campos-Lima, P. O., Wallace, L. E., Dolcetti, R., Murray, R. J., Rickinson, A. B. & Masucci, M. G. (1993). *J. Virol.* **67**, 1572–1578.
- Gavioli, R., Zhang, Q. J., Marastoni, M., Guerrini, R., Reali, E., Tomatis, R., Masucci, M. G. & Traniello, S. (1995). *Biochem. Biophys. Res. Commun.* **206**, 8–14.
- Guo, H. C., Jardetzky, T. S., Garrett, T. P., Lane, W. S., Strominger, J. L. & Wiley, D. C. (1992). *Nature (London)*, **360**, 364–366.
- Hillig, R. C., Coulie, P. G., Stroobant, V., Saenger, W., Ziegler, A. & Hulsmeier, M. (2001). *J. Mol. Biol.* **310**, 1167–1176.
- Johnson, R. P. & Walker, B. D. (1991). *Biotechnol. Ther.* **2**, 137–146.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kast, W. M., Brandt, R. M., Sidney, J., Drijfhout, J. W., Kubo, R. T., Grey, H. M., Melief, C. J. & Sette, A. (1994). *J. Immunol.* **152**, 3904–3912.
- Klein, J., Satta, Y., O'hUigin, C. & Takahata, N. (1993). *Annu. Rev. Immunol.* **11**, 269–295.
- Kubo, R. T., Sette, A., Grey, H. M., Appella, E., Sakaguchi, K., Zhu, N. Z., Arnott, D., Sherman, N., Shabanowitz, J., Michel, H., Bodnar, W. M., Davis, T. A. & Hunt, D. F. (1994). *J. Immunol.* **152**, 3913–3924.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Li, L. & Bouvier, M. (2004). *J. Immunol.* **172**, 6175–6184.
- Logean, A., Sette, A. & Rognan, D. (2001). *Bioorg. Med. Chem. Lett.* **11**, 675–679.
- Lund, O., Nielsen, M., Kesmir, C., Petersen, A. G., Lundegaard, C., Worning, P., Sylvester-Hvid, C., Lamberth, K., Roder, G., Justesen, S., Buus, S. & Brunak, S. (2004). *Immunogenetics*, **55**, 797–810.
- Marra, M. A. *et al.* (2003). *Science*, **300**, 1399–1404.

- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Otvos, L. Jr, Cappelletto, B., Varga, I., Wade, J. D., Xiang, Z. Q., Kaiser, K., Stephens, L. J. & Ertl, H. C. (1996). *Biochim. Biophys. Acta*, **1313**, 11–19.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pedersen, L. O., Stryhn, A., Holter, T. L., Etzerodt, M., Gerwien, J., Nissen, M. H., Thogersen, H. C. & Buus, S. (1995). *Eur. J. Immunol.* **25**, 1609–1616.
- Poon, L. L., Guan, Y., Nicholls, J. M., Yuen, K. Y. & Peiris, J. S. (2004). *Lancet Infect. Dis.* **4**, 663–671.
- Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanovic, S. (1999). *Immunogenetics*, **50**, 213–219.
- Reiser, J. B., Darnault, C., Guimezanes, A., Gregoire, C., Mosser, T., Schmitt-Verhulst, A. M., Fontecilla-Camps, J. C., Malissen, B., Housset, D. & Mazza, G. (2000). *Nature Immunol.* **1**, 291–297.
- Reiser, J. B., Gregoire, C., Darnault, C., Mosser, T., Guimezanes, A., Schmitt-Verhulst, A. M., Fontecilla-Camps, J. C., Mazza, G., Malissen, B. & Housset, D. (2002). *Immunity*, **16**, 345–354.
- Rognan, D., Lauemøller, S. L., Holm, A., Buus, S. & Tschinke, V. (1999). *J. Med. Chem.* **42**, 4650–4658.
- Schueler-Furman, O., Altuvia, Y., Sette, A. & Margalit, H. (2000). *Protein Sci.* **9**, 1838–1846.
- Sette, A. & Sidney, J. (1999). *Immunogenetics*, **50**, 201–212.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Siddell, S. G. (1995) *The Coronaviridae*. New York: Plenum.
- Sidney, J., Grey, H. M., Southwood, S., Celis, E., Wentworth, P. A., del Guercio, M. F., Kubo, R. T., Chesnut, R. W. & Sette, A. (1996). *Hum. Immunol.* **45**, 79–93.
- Speir, J. A., Abdel-Motal, U. M., Jondal, M. & Wilson, I. A. (1999). *Immunity*, **10**, 51–61.
- Stryhn, A., Pedersen, L. O., Holm, A. & Buus, S. (2000). *Eur. J. Immunol.* **30**, 3089–3099.
- Surjit, M., Liu, B., Jameel, S., Chow, V. T. & Lal, S. K. (2004). *Biochem. J.* **383**, 13–18.
- Surjit, M., Liu, B., Kumar, P., Chow, V. T. & Lal, S. K. (2004). *Biochem. Biophys. Res. Commun.* **317**, 1030–1036.
- Sylvester-Hvid, C., Nielsen, M., Lamberth, K., Roder, G., Justesen, S., Lundegaard, C., Worning, P., Thomadsen, H., Lund, O., Brunak, S. & Buus, S. (2004). *Tissue Antigens*, **63**, 395–400.
- Townsend, A. & Bodmer, H. (1989). *Annu. Rev. Immunol.* **7**, 601–624.
- Walker, B. D., Flexner, C., Birch-Limberger, K., Fisher, L., Paradis, T. J., Aldovini, A., Young, R., Moss, B. & Schooley, R. T. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 9514–9518.
- Zhang, Q. J., Gavioli, R., Klein, G. & Masucci, M. G. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2217–2221.