MHC-restricted antigen presentation and recognition: constraints on gene, recombinant and peptide vaccines in humans

E. Cunha-Neto

Laboratório de Imunologia de Transplantes, Instituto do Coração, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brasil

Abstract

The target of any immunization is to activate and expand lymphocyte clones with the desired recognition specificity and the necessary effector functions. In gene, recombinant and peptide vaccines, the immunogen is a single protein or a small assembly of epitopes from antigenic proteins. Since most immune responses against protein and peptide antigens are T-cell dependent, the molecular target of such vaccines is to generate at least 50-100 complexes between MHC molecule and the antigenic peptide per antigen-presenting cell, sensitizing a T cell population of appropriate clonal size and effector characteristics. Thus, the immunobiology of antigen recognition by T cells must be taken into account when designing new generation peptide- or gene-based vaccines. Since T cell recognition is MHC-restricted, and given the wide polymorphism of the different MHC molecules, distinct epitopes may be recognized by different individuals in the population. Therefore, the issue of whether immunization will be effective in inducing a protective immune response, covering the entire target population, becomes an important question. Many pathogens have evolved molecular mechanisms to escape recognition by the immune system by variation of antigenic protein sequences. In this short review, we will discuss the several concepts related to selection of amino acid sequences to be included in DNA and peptide vaccines.

Key words
- Vaccines
- MHC
- Antigen recognition
- Antigen processing
- T cells
- Molecular evolution

Molecular mechanisms involved in MHC-restricted antigen presentation and recognition by T cells

Vaccination with DNA- or peptide-based vaccines relies on T cell recognition, either in the afferent (T cell-dependent antibody-mediated responses) or in both the afferent and effector limbs of the immune response (T cell-mediated responses - e.g., delayed hypersensitivity). In order to induce protective immunity, epitopes contained in these new generation subunit vaccines must i) match epitopes naturally presented to the immune system during infection; ii) be recognized by the entire target population, and iii) induce an immune response of an adequate isotype. The central event in the cellular immune response to invading microorganisms is the specific recognition of foreign peptides bound to major histocompatibility complex (MHC) molecules by the αβ
T cell receptor (TCR). In most immunological systems, including post-infectious protective immune responses, only some peptides from a large number of potential candidates are actually the target of a vigorous immune response. The epitopes to which a natural immune response is generated upon immunization with the whole protein are called dominant epitopes. Immunization with overlapping peptides induces suboptimal responses to additional peptides, the so-called “cryptic” epitopes. A large number of factors influence dominance and crypticity of peptide epitopes, basically availability for MHC binding, MHC binding itself, and the recognition of the MHC:peptide complex by T cells via their antigen receptors (1). While peptide vaccines may bind directly to the MHC, recombinant vaccines must undergo proteolytic processing through the MHC class II pathway in endosomal vesicles, and expression products of DNA vaccines enter both the cytoplasmic/endoplasmic reticulum MHC class I pathway and can also be uptaken by professional antigen-presenting cells (2).

The steps involved in antigen presentation and recognition by T cells that may influence the selection of antigenic protective epitopes from whole proteins will be reviewed.

Availability of peptides for MHC binding in antigen presentation pathways

Several factors influence the availability of a given peptide sequence for processing and presentation, at both the quantitative and qualitative levels. An abundant protein has a higher chance of yielding peptides available for presentation. In intracellular eukaryotic parasites like *Trypanosoma cruzi*, proteins that are secreted may be a major source of peptides for the MHC class I presentation pathway (3). Cleavage site-specific proteases and transporter proteins involved in the processing of protein antigens into peptides also seem to play a role in the selection of antigenic peptides. Proteasomes, which generate peptides from cytoplasmic proteins for the class I pathway, seem to have preferred cleavage sites flanking dominant CD8+ T cell epitopes in protein sequences (4). The human TAP (transporter associated with antigen presentation) molecule selects peptides according to a binding motif on their three amino-terminal and two carboxy-terminal residues (5). Cathepsin and carboxypeptidase, which are present in the phagolysosomal compartment and are involved in proteolytic processing of endocytosed proteins in the MHC class II pathway, display preferential cleavage of dibasic (RR, KK, KR or RK) sites (6). Peptides eluted from human HLA class II molecules frequently display proline residues close to their N-terminus, consistent with N-aminopeptidase activity (7).

HLA polymorphism and binding of T cell epitopes

T cells recognize antigen via the variable regions of the clonotypic T cell antigen receptor molecule (TCR). Zinkernagel's (8) pioneering experiments and succeeding refinements performed by many authors disclosed that T cells recognize antigen as proteolytic peptide fragments together with the individual's own MHC molecule. CD8+ T cells recognize peptide antigen as 8-10-residue-long peptides embedded into the antigen-binding groove of MHC class I molecules (the products of gene loci H2-K, -S, and -L in mice; HLA-A, -B and -C in humans). CD4+ T cells recognize peptide fragments of 13 residues or longer embedded into the open-ended antigen-binding groove of MHC class II molecules (the products of gene loci H2-I-A and I-E in mice; HLA-DR, -DQ and -DP in humans).

The structure of MHC molecules complexed with different peptides was solved by X-ray crystallography studies. Bound peptides are buried in the antigen-binding groove.
formed between the helices of the MHC molecule, leaving only a few of their side chains available for direct TCR contact (9-12). Peptides bind to MHC class II in an extended conformation where the side chain of each amino acid residue is rotated 120° relative to the previous one; nine residues lie within the antigen-binding groove (10) while flanking residues may also interact with the TCR (13). A network of hydrogen bonds between the peptide main chain and residues on the groove assures a tight interaction between peptide and MHC; specific binding of certain peptides to an MHC molecule comes from the interaction of peptide side chains with the irregular surface of the floor and sides of the groove, the pockets and ridges formed by the protrusion of MHC residues. The major pockets in the floor of the groove of HLA-DR molecules are occupied by the side chains of residues 1, 4, 6, and 9 of the bound peptide (10,11).

Each human individual carries up to twelve different alleles from the six HLA class I and class II loci. HLA molecules are highly polymorphic, with more than 600 allelic forms for HLA class I and class II (14). This polymorphism is concentrated in the region encoding the peptide-binding groove, yielding very diverse amino acid sequences in this region among different HLA alleles. Thus, most pockets in the HLA groove are shaped by clusters of polymorphic HLA residues, with peculiar chemical and size characteristics in different HLA molecules. For example, a positively charged side chain (e.g. lysine) in one HLA molecule may interact preferentially with negatively charged residues (e.g. aspartic or glutamic acid) in a bound peptide; conversely, a hydrophobic residue in the same position of another HLA molecule may only accept peptides with hydrophobic residues in the corresponding position (15). It follows that each allelic HLA molecule only binds peptides with amino acid sequences that are capable of interacting with its antigen-binding groove. These preferences can be evident after the sequence alignment of peptides known to interact with a given HLA molecule, disclosing certain peptide positions where amino acids with similar side chains occur at increased frequency, the peptide-binding preferences or "motifs" (16). The study of large, MHC-selected peptide pools, and the sequencing of endogenous MHC-bound peptides (17) or the analysis of large peptide pools from phage display libraries selected by binding to HLA-DR (18) allowed a generalization of rules for peptide binding to individual HLA molecules (15,19). Single-chain substitution of designer (optimally binding) peptides led to the establishment of motifs based on quantitative matrices, where the effect of each amino acid residue in each peptide position was taken into account, generating a numeric score, rather than the all-or-none result for simple motifs (20). Further, study of peptides eluted from HLA-DR and DQ suggested that residues flanking the bound nonamers may influence overall peptide binding preferences in an allele-specific manner (21), confirming what was observed in the mouse system (6). Since each of the different allelic HLA molecules has different peptide-binding preferences, a distinct set of epitopes from a given protein antigen will be presented to T cells in each individual bearing a different HLA molecule.

On the whole, this implies that different individuals may be able to recognize different antigenic peptides from a given protein. This interindividual variation of antigen recognition can be a problem for selecting immunogens for DNA, recombinant and peptide vaccines, since they must contain epitopes recognized by individuals with a wide range of different HLA molecules. This can be circumvented by including additional epitopes that can be presented by different HLA alleles covering most of the population, which can increase the complexity of the vaccine. On the other hand, it has been observed that several distinct HLA class I...
molecules share peptide-binding preferences ("supermotifs") and can bind certain common peptides with high affinity. Thus, the identification of single peptides that can bind multiple HLA types, the so-called "promiscuous" epitopes, could lead to effective coverage of the human population by a peptide-based vaccine (19).

**T cell receptor recognition of MHC/peptide complex: structure and function**

Recently, the structure of two trimolecular complexes (TCR bound to MHC/peptide) was solved: a complex containing a human \( \alpha \beta \) TCR bound to an HLA-A2 molecule loaded with a nonapeptide derived from the HTLV-1 virus (22) and a mouse TCR complexed to an H-2Kb/peptide ligand (23). In both structures, TCR V\(\alpha\)V\(\beta\) regions interact with the composite surface made of residues belonging to the antigenic peptide and to both MHC H\(\alpha\)1 and H\(\alpha\)2 helices. The TCR is oriented diagonally across the peptide-binding site of HLA-A2, with the TCR V\(\alpha\) contacting the left end of the peptide-binding groove, toward the amino-terminal part of the MHC H\(\alpha\)1 helix and of the peptide, while the TCR V\(\beta\) is positioned toward the carboxy terminus of the peptide. The third complementarity-determining regions (CDR3) of the TCR V\(\alpha\) and V\(\beta\) chains, which display extremely diverse sequences due to junctional diversity that occurred during VDJ rearrangement, contact the center of the bound peptide and seem to have extensive conformational flexibility. In TCR:MHC class II:peptide complexes, a similar orientation seems to apply; however, N- and C-terminal residues flanking the peptide's nine MHC-bound residues may interact with the TCR (13). Flanking residues may also hinder TCR interaction with MHC/peptide complex (6,13).

Along a peptide, the side chains of certain residues interact with the TCR (TCR contact residues) while residues in other positions interact with the MHC molecule (MHC contact residues or anchor residues). Substitutions at MHC contact residues may yield peptide analogs binding MHC with different avidities from the parent peptide. Avidity of the peptide for MHC and dosage seems to influence the cytokine profile resulting after immunization. It has been reported that among peptide analogs binding to MHC with different affinities over a 10,000-fold range, the higher affinity peptide always elicits a T1-type cytokine profile, while the lower affinity analog yields a T2-type profile at low doses (24) but a T1-type profile at higher doses (25). Peptide analogs with conservative substitutions at TCR contact residues may be full antagonists or partial agonists, inducing anergy or partial or differential T cell activation (26,27).

**Prediction of T cell epitopes**

Ideally, modern subunit vaccines should contain highly immunogenic T cell epitopes intentionally chosen from the available pathogen proteins. However, a comprehensive screening of T cell epitopes by random cloning or direct synthesis of overlapping peptides from large numbers of gene sequences is often impossible. With the growing knowledge on the nature of TCR recognition of the MHC-peptide complex, several investigators tried to develop methods to predict T cell epitopes. The first algorithms, based on the simple alignment of antigenic peptides (and often disregarding diversity of MHC alleles), were no better than random when tested on larger databases (28). MHC allele-specific peptide-binding motifs were used in the prediction of T cell epitopes (29-31), minimizing the number of peptides to be synthesized and assayed. The success of MHC-based epitope prediction seems to lie in the filtering out of potential T cell epitopes, since most natural peptides seem to lack the capacity to interact with MHC mol-
MHC-restricted antigen recognition and molecularly defined vaccines

 molecules (20). MHC binding algorithms based on quantitative matrices yielded scores which correlated with binding affinity (20) like the TEPITOPE described for up to 25 distinct HLA-DR molecules (15), or the Internet server BIMAS (32) for several mouse and human MHC class I alleles, which permit the selection of high affinity binding peptides, the ones with the greater chance of eliciting T cell responses (33).

From the vaccine immunologist’s point of view, however, identification of MHC allele-specific T cell epitopes may not be enough, since one is searching for vaccine epitopes that can effectively cover the human population. This implies the identification of promiscuous epitopes that can bind to several MHC alleles whose combined frequency in the population approaches 100%. The use of HLA class I supermotifs (19), the alignment of peptides binding to several distinct HLA-DR molecules with TEPITOPE (15) and the identification of sequences containing clustered motifs for several distinct MHC molecules (34; Cunha-Neto E and Gruber A, unpublished observations) can identify such potentially promiscuous epitopes.

**Escape from presentation and recognition by “molecular evolution”**

For millions of years pathogens have evolved molecular mechanisms to escape effective presentation and recognition to the immune system by variation of antigenic protein sequence, after interplaying with the host’s immune system (35). Antigen processing and peptide transport mechanisms are ancient in evolutionary terms. The major features of the MHC peptide-binding groove, like the hydrophobic nature of most pockets, leading to preferential binding of peptides with hydrophobic residues, are phylogenetically conserved from cartilaginous fish and sharks to man (~250 million years) (36). Sequence variation in or flanking an epitope may lead to an inadequate response (37) by several mechanisms such as loss of processing signals (e.g., loss of sites for sequence-specific proteases or a different folding pattern), epitope competition and epitope sequestration (6,38), abrogation of MHC binding (39), antagonism or partial agonism of T cell receptor signalling (40), establishment of ineffective mock dominant epitopes (e.g., those in variant regions of CS proteins from Plasmodia, stronger than those in conserved regions), and molecular mimicry with self or abundant environmental epitopes (41,42). In the case of molecular mimicry between host and pathogen epitopes, it is important to identify potentially pathogenic crossreactive epitopes to be excluded from a protective immunogen, as in the case of Streptococcus pyogenes M protein and rheumatic fever (43). Tandemly repeated domains of parasitic protozoan proteins, which are immunodominant B-cell epitopes, display nonrandom amino acid usage with few hydrophobic residues leading to deficient binding to MHC and T cell recognition (Abel LCJ, Gruber A and Cunha-Neto E, unpublished observations; 44). Thus, the immune response of infected individual to intact pathogen protein antigens may reflect the evolutionary success of the parasite. DNA or recombinant vaccines encoding whole proteins or their genes will carry most molecular sequence variation escape mechanisms evolved by the pathogen. Immunization with substituted synthetic peptides or DNA/proteins can present neoepitopes or alter the hierarchy of dominant/cryptic T cell epitopes, bypassing recognition escape mechanisms (45).

**Concluding remarks**

The astonishing rate at which knowledge of T cell antigen recognition has been accumulating in the last decade is setting the stage for the comprehensive search for ideal epitopes and further engineering of immu-
References


