How cytotoxic T cells manage to discriminate nonself from self at the nonapeptide level

(nonapeptides/similarity/promiscuity)

SUSUMU OHNO

Beckman Research Institute of the City of Hope, 1450 East Duarte Road, Duarte, CA 91010-0269

Contributed by Susumu Ohno, February 6, 1992

ABSTRACT Class I major histocompatibility complex (MHC) antigens are confronted with an apparently insurmountable dilemma. Each should show a binding preference to a common enough variety of nonapeptides, so that one relevant nonapeptide can be found in at least every other viral protein to provoke a cytotoxic T-cell response. By so doing, however, the chance of that viral T epitope being self is greatly increased. Examination of human and viral nonapeptides preferred by HLA-B27 led to the following conclusions. (i) In normal cells, peptide fragments originating from 5000 or more diverse proteins vie for a finite number of class I MHC sites. Consequently, only those nonapeptides having the optimal binding affinity to a given class I MHC antigen can gain access to the plasma membrane. (ii) Tolerance is rendered only to those host nonapeptides with the optimal binding affinity. (iii) Because of the above noted tolerance, viral nonapeptides with the optimal binding affinity are invariably ignored. (iv) Viral T epitopes actually chosen are always second-echelon nonapeptides that are endowed with slightly less than the optimal binding affinity to a given class I MHC antigen. (v) Since such second-echelon nonapeptides would not gain access to the plasma membrane in normal cells, the issue of self or nonself is rendered irrelevant by this choice. (vi) Since viral T epitopes are of this type, cytotoxic T-cell responses against infected cells are expected to be effective only when a few viral proteins are made in large amounts at the expense of host proteins.

Inasmuch as peptide fragments presented by class II major histocompatibility complex (MHC) antigens are tetradecapeptide to heptadecapeptide in length and are derived from extracellular proteins (1), no problem with self versus nonself discrimination is anticipated for CD4+ helper T cells that recognize the above complex. Extracellular proteins encoded by the host genome are rather small in variety, and the longer the length of a peptide fragment, the greater is the chance of that peptide being unique either to the host or to a parasite.

The same cannot be said of CD8+ cytotoxic T cells that are designed to recognize class I MHC antigen-bound nonapeptides (occasionally octapeptides) originating from intracellular proteins (2, 3). It follows that self to cytotoxic T cells is derived from all of the 50,000 or more varieties of proteins encodable by the mammalian genome. Even extracellular proteins to the body as a whole are major intracellular proteins within specific cell types that manufacture them. Faced with this enormity of self, why did cytotoxic T cells choose to deal with short nonapeptides? Relevant to the above question is the raison d'être of class I MHC antigens, which is to provoke cytotoxic T-cell response by presenting peptide fragments derived from intracellular parasites. In order to do so, each class I MHC antigen should show the binding affinity to a common enough variety of nonapeptides, so that one relevant nonapeptide can at least be found in every other protein of the parasite origin. Indeed, certain oligopeptides are common enough to be found among totally unrelated proteins (4, 5).

When the amino acid composition at each of the nine positions was determined on host nonapeptides eluted from one human and three murine class I MHC antigens purified from uninfected cells, it became evident that each allelic form of class I MHC antigen preferentially binds to a specific set of nonapeptides (2). In the case of HLA-A2, for example, all host nonapeptides eluted from it had either leucine or methionine at the second position and almost exclusively valine at the ninth position. In sharp contrast, viral nonapeptides having these attributes were seldom chosen as T epitopes of HLA-A2-restricted cytotoxic T-cell responses (6). More often than not, T-epitope nonapeptides were those having isoleucine instead of leucine or methionine at the second position and residues other than valine at the ninth position, thus indicating their less than optimal binding affinity to HLA-A2 (6). One decisive advantage of choosing a second preference viral nonapeptide as a T epitope is found in the fact that by so doing the problem of self or nonself is rendered irrelevant. In normal cells in which peptide fragments derived from thousands of host proteins compete for the finite number of class I MHC antigen sites, such a second preference nonapeptide will be overwhelmed by a group of primarily preferred nonapeptides. Only in infected cells where a few viral proteins are made in large amounts, while the host protein synthesis is suppressed, can such a second preference nonapeptide gain a sufficient density to become the effective target of cytotoxic T cells (6).

Inasmuch as more information became available on the human HLA-B27 class I MHC antigen as to its interaction with human and viral nonapeptides (3, 7), this question of nonapeptides with less than the optimal binding affinity as preferred viral epitopes of cytotoxic T cells is reexamined with regard to HLA-B27.

Materials

As representatives of the host, seven human proteins totally unrelated to each other were chosen. In descending order of their sizes, they were 858-residue-long elongation factor 2 (8), 782-residue-long DNA repair helicase (9), 724-residue-long 90-kDa heat shock protein (10), 595-residue-long estrogen receptor (11), 585-residue-long serum albumin (12), 461-residue-long M1 muscarinic acetylcholine receptor (13), and 417-residue-long phosphoglycerate kinase (14), with the grand total of 4422 residues.

Of the above seven, elongation factor 2, phosphoglycerate kinase, and probably also the estrogen receptor are ubiquitous proteins: they are expressed by every somatic cell type of the body. DNA repair helicase and heat shock protein, on

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; HIV, human immunodeficiency virus.
the other hand, are inducible ones; either ultraviolet radiation or various stresses trigger their induction. Were the tolerance not rendered to relevant nonapeptides contained in these inducible proteins, damaged as well as stressed cells would become unintended targets of cytotoxic T cells. Serum albumin is a major extracellular protein; its peptide fragments, therefore, should be surveyed by class I MHC antigens. Yet, for its manufacturers, which are hepatocytes, it must also be a major intracellular protein. The M1 muscarinic acetylcholine receptor represents rather numerous transmembrane proteins. Their peptide fragments are surveyed by both helper T cells and cytotoxic T cells.

As representatives of invading parasites, three proteins from each of two viruses that target humans as their primary host were chosen. Three from human immunodeficiency virus type 1 (HIV-1) (15) were 1006-residue-long transcriptase, 863-residue-long envelope protein, and 512-residue-long gag p24 protein. Three proteins from influenza A virus were 550-residue-long hemagglutinins (16), 498-residue-long nucleoprotein (17), and 252-residue-long matrix protein (18). The grand total of these six viral proteins was 3681 residues. Upon infection, peptide fragments of these viral proteins have been shown to provoke cytotoxic T-cell responses restricted to diverse HLA alleles.

Eleven Host Nonapeptides Eluted from HLA-B27 of IG-2 Lymphoblastoid Cells

Eleven host nonapeptides identified in eluates from purified HLA-B27 of uninfected human cells are shown in Fig. 1. They are aligned in the order of their yields (3). Although two other nonapeptides derived from HLA-B27 itself were also included in the original report (3), the amino acid sequences of these two showed little resemblance to the above 11. Accordingly, these two were discarded as degraded contaminants. At first glance, marked differences in yields of these 11 nonapeptides appear to reflect differences in their binding affinity toward HLA-B27 more than anything else. Five of the 11 were apparently derived from a single protein, a human counterpart yet to be sequenced of the rat 60S ribosomal protein L28 (3). Yet, their yields were very different, ranging from 14 pmol to 1.5 pmol. However, there is an alternative explanation. Those of higher yields, such as nonapeptides 1 and 2, might have been derived not from one, but from several proteins. It was previously shown that, within a hypothetical giant protein comprised of several totally unrelated ones, a pair of identical pentamers is found, on the average, once every 390 residues; a pair of identical hexamers is found once every 1200 residues (5, 6). The incidence of a pair of nonamers differing from each other by two conservative substitutions was about the same as that of identical hexamers. Thus, the possibility of even unrelated proteins yielding the identical nonamer is real. Furthermore, many proteins are comprised of several isozymic forms. For example, human muscarinic acetylcholine receptors are present in five forms (M1–M5) of very similar amino acid sequences (13). Accordingly, if the optimal affinity nonapeptide is present in one, the other four would also contribute the same nonapeptide.

Whereas the presence of arginine at the second position was an absolute prerequisite for the optimal binding affinity, residues at the first position influenced the choice of residues at the ninth position and vice versa. The presence of a small hydrophobic residue such as glycine or alanine at the first position dictated the presence of a basic residue at the ninth position as in nonapeptides 2 and 3. Conversely, the presence of a basic residue at the first position was compatible with the occupancy of the ninth position by a noncharged residue such as leucine, tyrosine, or alanine, as shown in nonapeptides 1, 4, and 8 of Fig. 1.

Aside from the 11 nonapeptides listed in Fig. 1, there must have been other HLA-B27-bound host nonapeptides that escaped individual identification.

The Number of High-Affinity HLA-B27-Restricted Nonapeptides in Seven Unrelated Human Proteins

As already noted, residues at the first, second, and ninth positions of nonapeptides were of prime importance in determining the binding affinity of these nonapeptides shown in Fig. 1 to HLA-B27. Indeed, the x-ray crystallographic analysis of HLA-B27 has shown that the first, second, third, seventh, and ninth residues of bound nonapeptides were buried deep in the groove provided by two parallel α-helices of this class I MHC antigen (7). Under the assumption that third and seventh positions in addition to the first, second, and ninth positions are more critical than the remaining four in determining the binding affinity of nonapeptides to HLA-B27, the amino acid compositions of these five positions were constructed from 11 nonapeptides of Fig. 1 in a manner comparable to those constructed from nonapeptides bound to other class I MHC antigens (Fig. 2) (2). The diverse yields of the 11 nonapeptides shown in Fig. 1 were taken into account. Shown immediately below the amino acid compositions, in the first and second rows, are two nonapeptides found among seven unrelated human proteins. Residues at the above five positions of these two nonapeptides were all represented in the amino acid compositions shown above them, thus suggesting that the binding affinities of these two to HLA-B27 were comparable to those of the 11 listed in Fig. 1. In fact, the first, from the estrogen receptor, resembled nonapeptide 1 of Fig. 1, as to residues at the first, second, and ninth positions, whereas the second, from DNA repair helicase, resembled nonapeptide 8 of Fig. 1.

Seven unrelated human proteins also yielded two additional nonapeptides that were likely to show nearly as high a binding affinity to HLA-B27 as any thus far discussed. These two are identified in the third and fourth rows of Fig. 2. Although each had a mismatch at one of the five critical positions, each mismatch was a negligible one. For example, a mismatch in one nonapeptide representing positions 449–457 of the 90-kDa heat shock protein was leucine at the
volumes of nonapeptides, the sequence determination of any of them would have been impossible. If it is assumed that nonapeptide 11, which had the lowest yield of any of the nonapeptides shown in Fig. 1, still occupied 1.0% of the HLA-B27 sites, nonapeptide 1, as well as nonapeptide 2, would have to have occupied 10.7% of the sites each. It follows that together the 11 nonapeptides of Fig. 1 accounted for 52.7% of the available HLA-B27 sites on IG-2 lymphoblastoid cells. At most, the remaining 47.5% of the sites can be occupied by 475 additional high-affinity nonapeptides, each with an occupancy rate of 0.1% of the HLA-B27 sites. Any other nonapeptide with an occupancy rate below 0.1% is of no consequence; it cannot serve as a potentially effective target of the cytotoxic T-cell response (19).

This apparent discrepancy can again be resolved by the assumption that identical and nearly identical nonapeptides having the optimal binding affinity to HLA-B27 are often contributed by more than one protein, either related or unrelated to each other.

Indeed, within Fig. 1 nonapeptides 6 and 7 differed from each other by only two conservative substitutions; thus, this pair enjoyed the identity score of 8.4/9—each position occupied by an identical residue received 1.0 point, whereas a position occupied by conservatively substituted residues received 0.7 point (5, 6). Nonapeptide 6 also shared 5.7/9 identity with nonapeptide 9. As to Fig. 2, the nonapeptide derived from the 90-kDa heat shock protein (positions 449–457) enjoyed the identity score of 6.4/9 with nonapeptide 4 of Fig. 1 (Fig. 2, bottom two rows). It also had an identity score of 5.8/9 with nonapeptide 6 as well as with nonapeptide 7 of Fig. 1. The nonapeptide derived from the estrogen receptor (positions 362–370) enjoys 5.0/9 identity with the nonapeptide of phosphoglycerate kinase (positions 16–24) (Fig. 2). Inasmuch as only 15 nonapeptides shown in Figs. 1 and 2 showed considerable homology with each other, it is rather expected that 5000 diverse proteins would supply only 500 or so varieties of optimal binding affinity nonapeptides rather than 1332 distinct entities.

**On Thymic Educator Cells of Self to Cytotoxic T Cells**

What if the thymic educator cell of self to cytotoxic T cells is unlike any other cell type and simultaneously synthesizes all the proteins encodable by the genome? Inasmuch as 3000 or so proteins for household chores have to be produced in fixed amounts, only minuscule amounts of the remaining 4700 strong would be produced to no avail. The net result would be that tolerance can be achieved with regard to only those nonapeptides derived from household chore proteins. What if thymic educator cells become multiclonal, each clone resembling one or the other of the somatic cell types that comprise the body? As often as not, an autoreactive cytotoxic T cell would be lost in an island of thymic educator cells representing a wrong cell type; thus, by failing to encounter its intended target, it escapes suppression and elimination. Since neither of the two alternatives noted above is a viable option, we are left with the most mundane of all the possibilities: the thymic educator cell produces roughly 5000 diverse proteins as all other somatic cell types do. Yet, the tolerance rendered to 500 or so varieties of optimal affinity nonapeptides presented by HLA-B27 sites of this cell type is expected to anticipate all the additional optimal affinity nonapeptides to be derived from the remaining 45,000 proteins encodable by the host genome. What makes this type of anticipation possible is the required commonness of nonapeptides enjoying the optimal binding affinity to HLA-B27 or to any other class I MHC antigen.

---

**Fig. 2.** The amino acid composition at each of five critical positions is shown at the top. These are deduced from the 11 nonapeptides of Fig. 1, taking their yields into consideration, and are presented in the manner emulating that in ref. 2. Arginine at the second position, shown in large capital letters, is the invariant anchor residue. As to other positions, prominent residues are shown in large capital letters, while sporadic residues are shown in small capital letters. Shown in the first and second rows immediately below the amino acid compositions are two nonapeptides found among seven human proteins that contained residues corresponding to those in the list at all of the five critical positions. The second position anchor residue, arginine, received three asterisks, while prominent residues at four other positions received 2 asterisks and sporadic residues received 1 asterisk each. The source of each nonapeptide is indicated to the right, and its position in the protein is shown by numbers above the first and ninth residues. For extrapolation per 5000 proteins, see the text. Shown in the third and fourth rows are an additional nonapeptides (from the 90-kDa heat shock protein and phosphoglycerate kinase) that apparently are also endowed with the optimal binding affinity to HLA-B27. A negligibly mismatched residue at the seventh position of the third row and the ninth position of the fourth row are marked by an $. Homology between nonapeptides is compared in the bottom two rows of this figure. All the residues of the one nonapeptide are shown in large capital letters and are underlined. In the other, mismatched residues are shown in small capital letters and not underlined, identical residues are shown in all large capital letters, and conservatively substituted residues are shown with the first two letters of the abbreviation in large capital letters and the third in a small capital letter. The identity score is shown at the left.
Invariable Selection of Second Preference Nonapeptides as Viral T Epitopes

Shown as the first group in Fig. 3 are 5 nonapeptides derived from six viral proteins totaling 3689 residues. Since their residues at the five critical positions were comparable to those of the 15 host nonapeptides identified in Figs. 1 and 2, they are expected to be endowed with the optimal binding affinity toward HLA-B27. Yet, none of them has been identified as a target of an HLA-B27-restricted cytotoxic T-cell response. Evidently, tolerance has been rendered to them by anticipation of their presence in viral proteins. As shown at the bottom of Fig. 3, the nonapeptide derived from the HIV envelope protein (positions 510–518), indeed, had 6.4/9 identity with one host nonapeptide contained in the 90-kDa heat shock protein. At the same time, it also enjoyed 5.7/9 identity with another host nonapeptide derived from phosphoglycerate kinase. These two host nonapeptides were identified in Fig. 2. Similarly, the nonapeptide of influenza A virus nucleoprotein (positions 174–182) enjoyed 5.0/9 identity with host nonapeptide 8 of Fig. 1.

All three viral nonapeptides that have been shown to provoke HLA-B27-restricted cytotoxic T-cell responses as T epitopes (3, 20–22) were found among the six nonapeptides belonging to the second group. Since each showed a non-conservative mismatch at one of the five critical positions, their binding affinity to HLA-B27 must be considered as less than optimal, albeit second only to those in the first group. The advantage inherent in the choice of second preference nonapeptides as T epitopes has already been noted (6).

Perhaps the most informative was the situation found in the HLA-B27-restricted response to influenza A virus nucleo-

![Fig. 3](image_url)

**Fig. 3.** Two groups of nonapeptides found in six viral proteins are shown. As to the five in the first group, their residues at each of the five critical positions are contained in the amino acid composition shown above. Therefore, they are expected to enjoy the optimal binding affinity toward HLA-B27. Each of the six in the second group, on the other hand, had a mismatched residue indicated by a vertical arrow at one of the five critical positions. Thus, they are expected to have slightly less than the optimal binding affinity toward HLA-B27. Proven T epitopes among them are so indicated. One nonapeptide marked by an asterisk at the left is derived from a HIV protein not presently analyzed (3). At the bottom, one viral nonapeptide and two host nonapeptides are compared for homology. INF, influenza.

![Fig. 4](image_url)

**Fig. 4.** The original HLA-B27-restricted T epitope residing in HIV gag p24 protein is shown at the top. Substituted residues in its four mutants are shown below. Three mutants were still recognized in the same manner as the original T epitope by relevant cytotoxic T cells, while the fourth mutant escaped detection.
mutant T epitope, which underwent a drastic amino acid substitution, from glycine to glutamic acid at the seventh position, however, was no longer recognized by these cytotoxic T cells (20). Since glutamic acid is also a favored residue at the seventh position, this drastic substitution is not expected to have reduced the binding affinity of the fourth mutant to HLA-B27. It would thus appear that even some of the buried residues at the second, third, seventh, and ninth positions are surveyed by the cytotoxic T cell receptors.

At any rate, a hoped-for degree of promiscuity by cytotoxic T-cell receptor appears to have been confirmed.