Immunogenetics (1998) 48: 89–97 © Springer-Verlag 1998

ORIGINAL PAPER

Kiley R. Prilliman · Mark Lindsey
Kenneth W. Jackson · Jeffrey Cole · Ron Bonner
William H. Hildebrand

Complexity among constituents of the HLA-B*1501 peptide motif

Received: 7 October 1997 / Revised: 10 December 1997

Abstract Analysis of peptides derived from HLA class I molecules indicates that thousands of unique peptides are bound by a single molecular type, and sequence examination of the pooled constituents yields a motif which collectively defines the peptides bound by a given class I molecule. Motifs resulting from pooled sequencing are then used to infer whether particular viral and tumor protein fragments might serve as class I-presented peptide therapeutics. Still undetermined from a pooled motif is the breadth or range of peptides in the population which are brought together to form the pooled motif, and it is therefore not yet known how representative of the population a pooled motif is. By employing hollow fiber bioreactors for large-scale production of HLA class I molecules, sufficient peptides are produced to investigate individual subsets of peptides comprising a motif. Edman sequencing and mass spectrometric analysis of peptides eluted from HLA-B*1501 reveal that many peptide sequences fail to align with either the N- or C-terminal anchors predicted for the B*1501 peptide motif through whole pool sequencing. These analyses further reveal auxiliary anchors not previously detected and peptides significantly larger and smaller than the predicted nonamer, ranging from 6 to 12 amino acids in length. These results demonstrate that constituents of the B*1501 peptide pool vary markedly in comparison with one another and therefore in comparison with previously established B*1501 motifs, and such complexity indicates that many of the peptide ligands presented to CTL cannot be predicted using class I consensus motifs as search criteria.

Key words HLA-Bantigens · Ligands · Antigen presentation

K. Prilliman · M. Lindsey · J. Cole · W.H. Hildebrand (✉)
Department of Microbiology and Immunology,
University of Oklahoma Health
Sciences Center, Oklahoma City, OK 73190, USA

K.W. Jackson
William K. Warren Medical Research Institute,
University of Oklahoma Health
Sciences Center, Oklahoma City, OK 73190, USA

R. Bonner
SCIEX, Concord, Ontario L4K4V8, Canada

Introduction

Major histocompatibility complex (MHC) class I molecules, designated human leukocyte antigens (HLA) in our species, are responsible for binding peptides produced within the cytosol of a cell and displaying them upon the cell surface to CD8+ T lymphocytes (Björkman and Parham 1990; Townsend and Bodmer 1989). In this manner, T cells can survey the inner contents of a cell, vigilantly inspecting for virus- or malignancy-associated peptides in the context of class I HLA molecules (Mazzocchi et al. 1996; Townsend et al. 1985; Walker et al. 1988). Given the number and potentially variable nature of parent proteins (i.e., viruses) from which peptide ligands are derived, class I MHC molecules are themselves highly polymorphic and vary widely in allelic composition among members of the human population (Parham et al. 1989, 1995). This class I diversity dictates that a virus which escapes presentation by one person's HLA class I molecules will not necessarily have evaded capture and presentation by different class I molecules expressed in a second individual (Hughes and Nei 1988; Nei and Hughes 1992; Parham 1992; Parham and Ohta 1996).

Class I polymorphisms are focused about the cleft where peptide ligands are bound such that divergent class I molecules have different peptide binding grooves (Björkman et al. 1987; Parham et al. 1995). To facilitate a further understanding of the functional significance of class I polymorphism, many studies have been carried out to determine the sequences of peptide pools eluted from given class I molecules (Barber et al. 1995, 1996; Falk et al. 1993; Jardetzky et al. 1991). Indeed, amino acid sequencing of such peptide pools acquired from individual HLA

class I molecules reveals that thousands of peptides are bound by a specific molecule (Huczko et al. 1993; Hunt et al. 1992). Pooled amino acid sequencing therefore provides a composite of the thousands of peptides bound within a class I complex and typically discloses a motif whereby peptides are anchored into the antigen binding groove at their amino (N)- and carboxy (C)- termini (Elliott et al. 1993; Rammensee et al. 1995). While this census of pooled class I peptides elucidates what the average peptide bound by a given class I molecule resembles and allows comparisons of peptides bound by divergent class I molecules, it remains unclear precisely how heterogeneous are the constituents of a pooled sequence (Rammensee et al. 1993). In fact, the characterization of HLA class I peptides with cytotoxic T lymphocytes (CTLs) demonstrates that T cells recognize bound peptides which do not conform to the motif that emerges from the pooled amino acid sequencing of class I-eluted peptides (Malarkannan et al. 1996). The realization that peptides inconsistent with a respective class I molecule's designated "motif" may represent important T-cell epitopes begs the question, "How representative of the population is a peptide motif?".

Assessing the complexity of peptides bound by a given class I molecule is restricted in that (1) thousands of peptides are bound by a single class I molecule and (2) small (nanogram or lesser) amounts of this complex peptide mixture are typically available for study (Appella et al. 1995; Falk et al. 1991; Hunt et al. 1992). Examination of the individual constituents collectively defining a peptide motif is therefore limited in that quantities of the extracted peptides are easily exhausted before particular peptides can be separated from the complex milieu for analysis. To facilitate assessing the overall diversity of peptides which contribute to a class I motif we therefore utilize hollow-fiber bioreactors to provide milligram quantities of peptides for subsequent separation by HPLC and analysis by mass spectrometry and Edman sequencing (Prilliman et al. 1997). Our results show that the peptides eluted from HLA-B*1501 vary markedly in comparison with one another and therefore in comparison with the previously established B*1501 peptide motif. The implications of these results as they pertain to applying traditional peptide motifs in the selection of suitable T-lymphocyte ligands are discussed.

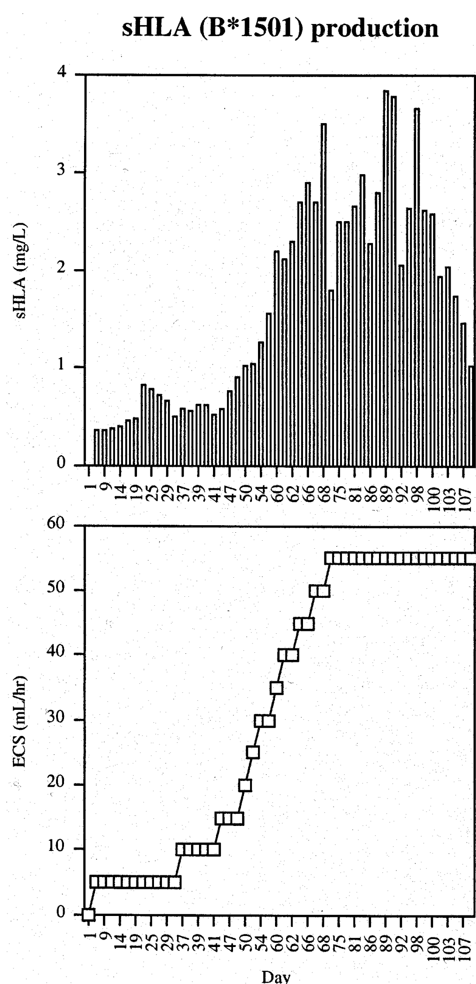

Fig. 1 Production yields of sB*1501. During culture of the soluble B*1501-producing transfectant in Cell-Pharm 3000 hollow-fiber bioreactor cartridges, basal media (RPMI-1640) provides glucose by circulating through the fibers, while the extracapillary space (*ECS*) housing the cells in each cartridge is continuously turned over with 10% FCS; this allows for the constant harvest of sB*1501-containing culture supernatant. Harvest concentration (*upper graph*) and collection rate (*lower graph*) data from a typical 12- to 15-week long bioreactor run demonstrates that from the 60–70 L of harvest accumulated during this span a total of 150–200 mg of sHLA complexes, as quantitated by the ELISA described, is obtained

Materials and methods

*Large-scale production and purification of soluble B*1501 (sB*1501) molecules*

The establishment and culture of a cell line which secretes a truncated form of the HLA class I molecule B*1501 have been discussed elsewhere (Prilliman et al. 1997). Briefly, the HLA-negative mutant B-lymphoblastoid cell line (B-LCL) 721.221 (Kavathas et al. 1980) was transfected with a pBJ1-neo construct containing the B*1501 coding region, which was truncated by the introduction via reverse transcription-polymerase chain reaction (RT-PCR) of a TGA stop at codon 300, thereby eliminating the transmembrane and cytoplasmic domains (exons 5–8), using primers HLA5UT (Domena et al. 1993) and sHLA3TM (Prilliman et al. 1997). Clones surviving G418 (Sigma, St. Louis, Mo.) selection were tested for sB*1501 secretion with a sandwich enzyme-linked immunosorbent assay (ELISA) employing W6/32 (Barnstable et al. 1978) as the capture antibody and α-β_2m-horseradish peroxidase conjugate (DAKO Corp., Carpinteria, Calif.) as the detecting antibody. Cells from positive wells were subcloned by limiting dilution to obtain transfectants of optimum expression. A selected sB*1501-producing transfectant was subsequently expanded to 4×10^9 viable cells in roller bottles of RPMI-1640+10% fetal calf serum (FCS) prior to inoculating them into the bioreactor cartridges of a Cell-Pharm 3000 unit (UNISYN Technologies Inc., San Diego, Calif.). Harvests collected during the run were centrifuged and passed over a CNBr-activated Sepharose 4B (Pharmacia Biotech, Piscataway, N.J.) column coupled with the light chain-specific monoclonal antibody (mAb) BBM.1 (Brodsky et al. 1979) for purification of soluble B*1501 trimers, which were then eluted with 0.2 N acetic acid; sB*1501 complexes from a portion of the harvests were purified using the heavy chain-specific MAb W6/32 for the purpose of comparison.

(W6/32-purified B*1501 complexes)

position:	1	2	3	4	5	6	7	8	9	10	11	12
dominant	–	**Q** (8.0)	**K** (8.0) **F** (5.0) **R** (4.0)	–	–	–	–	–	**Y** (4.0)	–	–	–
strong	–	**M** (3.0) **L** (2.0) **V** (2.0)	**Y** (3.0) **P** (2.0) **N** (2.0) **H** (2.0)	**P** (2.5) **D** (2.5) **G** (2.0) **E** (2.0)	**G** (2.0)	–	–	–	**F** (3.0)	–	–	–

(BBM.1-purified B*1501 complexes)

position:	1	2	3	4	5	6	7	8	9	10	11	12
dominant	–	**Q** (7.5)	–	–	–	–	–	–	**Y** (4.0)	–	–	–
strong	–	P (3.5) **L** (2.5) **V** (2.0)	**F** (2.5) **K** (2.5) **R** (2.0) **P** (2.0) **N** (2.0)	**P** (2.2) **D** (2.0) **G** (2.0)	**I** (2.0)	–	–	–	**F** (3.0)	–	–	–

Fig. 2 Pooled sequencing of peptides eluted from sB*1501. Data derived from the peptide eluates of complexes purified using either W6/32 (*upper panel*) or BBM.1 (*lower panel*) indicates amino acid residues which have been previously attributed to the corresponding positions in B*1501 motifs (*solid characters*; Barber et al. 1997; Falk et al. 1995), as well as residues which are not included in prior motif descriptions (*outline characters*). Relative picomolar fold increases are indicated in *parentheses* to the *right* of each residue; *dashes* indicate that no distinct residue(s) could be detected as dominant or strong at the given positions; for clarity, weakly detected residues are not shown here. Amino acid abbreviations comply with standard nomenclature (IUPAC-IUB Joint Commission on Biochemical Nomenclature 1984)

*Isolation and Edman sequencing of sB*1501-eluted peptides*

The sB*1501-associated peptides were liberated from immunoaffinity-purified complexes by acid denaturation and separation from α-chains and β2m by passage through a stirred cell (Amicon, Beverly, Mass.) equipped with a 3000 M_r exclusion membrane and collection of the eluate as previously performed (Prilliman et al. 1997). To examine relative sequence complexity, Edman degradation was carried out for at least 12 cycles on a model 492 A pulsed liquid phase protein sequencer (Perkin-Elmer Applied Biosystems Division, Norwalk, Conn.) with cysteine underivatized. Residue assignments were made from raw data as described (Barber et al. 1995; Stevanović and Jung 1993) through examining the picomolar fold increases of amino acids observed for each cycle as follows: dominant, >3.5-fold; strong, 2.0–3.5-fold; and weak, 1.5–2.0-fold. Data were obtained for peptides both from whole extracts (respectively obtained from BBM.1 and W6/32-purified complexes) and from fractions obtained in separating 150 μg of extract (from BBM.1-purified complexes only) by reversed-phase HPLC (RP-HPLC) on a C18 column (1.0×150 mm; Microm Bioresources Inc., Auburn, Calif.) using an initial gradient of 2–10% acetonitrile in 0.02 min followed by a linear gradient of 10–60% acetonitrile in 60 min at a 40 μl/min flow rate; absorbance was monitored at 214 nm, and fractions were collected every minute.

*Analysis of sB*1501-associated peptides by mass spectrometry*

Liquid chromatography/mass spectrometry (LC/MS) was employed to examine relative size complexity. The mass/charge (m/z) range of 300–1600 was broken into three overlapping and equivalent sets of Δ500 amu for examination as follows: (1) 300–800, (2) 700–1200, and (3) 1100–1600 m/z. For each of the three consecutive runs, approximately 50 μg of B*1501 peptide extract was desalted and eluted from the C18 column mentioned above directly into the IonSpray source (operated at 4 keV in positive ion mode) of a PE SCIEX API III triple quadruple mass spectrometer. Gradient parameters consisted of 2–10% acetonitrile in 0.02 min followed by 10–60% acetonitrile in 25 min at a 40 μl/min flow rate. Spectra were obtained by scanning the respective Δ500 m/z every 5 s and summing individual spectra. Inclusion of Δ100 amu overlaps between the ranges provided internal points of reference for evaluating consistency of the gradients and offsetting chromatograms as necessary.

Results

Large-scale production of soluble class I complexes and purification of peptides

As Fig. 1 indicates, 150–200 mg of sB*1501 is obtainable through large-scale bioreactor culture. Estimating that 2.2% of each class I complex is peptide ligand by weight, 3–5 mg of sB*1501-associated peptides are therefore available for isolation and analysis. Whole pool sequencing data were subsequently obtained by subjecting a sample of the low molecular weight protein material isolated from BBM.1- and W6/32-purified sB*1501 through acid elution to Edman degradation (Fig. 2). Amino acid assignments obtained with W6/32 purified material (upper panel) were consistent with previous characterization of B*1501 peptides, and the majority of amino acid assignments obtained

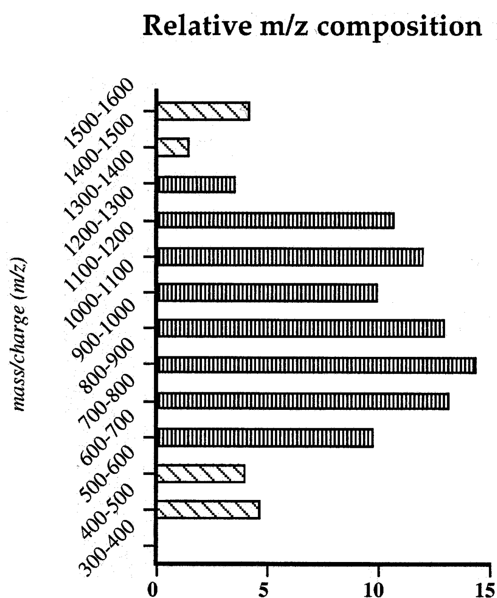

Fig. 3 Mass complexity of sB*1501-bound peptides. For simplification of the LC/MS data obtained for the peptides, which past analysis describes as highly complex (Prilliman et al. 1997), the data obtained from each of the three LC/MS runs as described in the text were broken into the 100 amu m/z range increments shown, thresholded at a signal-to-noise level of 2, and centroided using a merge distance of 1 amu and a minimum width of 0.5 amu. The number of centroided data points within each incremental m/z region was then considered with respect to the total number of centroided data points contained within the △1300 amu to obtain the relative percentages shown. Mass ranges corresponding to typical (nonameric) or atypical (smaller or larger) peptides as compared with the pooled motif are respectively denoted by *vertical* or *diagonal shading*

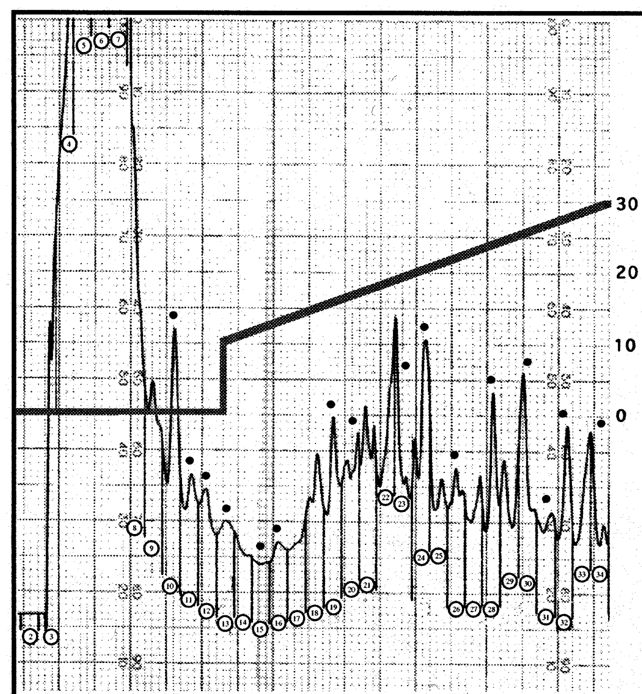

Fig. 4 RP-HPLC separation of sB*1501 peptide pool constituents. The UV trace obtained from running aproximately 150 µg of B*1501-derived peptides on a 10–60% acetonitrile gradient (*bold line*) over 60 min (chart speed = 0.5 cm/min) reveals that the region shown corresponding to the B*1501-associated material eluted mostly from 0–30% acetonitrile (*right*). The fractions collected are *numbered*; though all were examined, only the representatives subsequently selected for analysis and data presentation in Figs. 5 and 6 are marked by *dots*. Note that the chosen fractions were distributed evenly across the entire region of interest and included a wide variety which were of high as well as low UV absorbance and/or resolution

with BBM.1 conformed with past data obtained for B*1501 in defining Q and Y as dominant residues at positions 2 (P2) and P9 respectively (Barber et al. 1997; Falk et al. 1995). However, with BBM.1 a previously unreported strong anchor (P) was consistently identified for P2 (Fig. 2, lower panel) which failed to manifest in data obtained from peptides eluted from W6/32-purified class I complexes (Fig. 2, upper panel).

*General assessment of mass complexity within the sB*1501-bound peptide pool*

By splitting the m/z range to be analyzed into three distinct regions (300–800, 700–1200, and 1100–1600 m/z) over separate LC/MS runs, sensitivity was increased as compared with what would be detected when an entire m/z range of Δ1300 is collectively examined. Scanning Δ500 amu at the resolution afforded by a 0.2 amu step size with a dwell time of 2 msec allowed each Δ500 amu subrange to be scanned once every 5 s, whereas applying the same parameters to the entire Δ1300 amu range would have permitted a scan only once every 13 s. Examination of the sB*1501-associated peptide pool via LC/MS subsequently indicated that approximately 80% of the observed ion signals (Fig. 3) corresponded with singly $(M+H)^+$ and doubly $(M+2H)^{+2}$ charged species across a mass range of 600–1300 amu, which agrees with the sizes of specific 9- to 10-mers formerly characterized among B*1501-associated peptides (Barber et al. 1997; Falk et al. 1995). The remaining 20% (ranges of 400–600 and 1300–1600 amu) demonstrated m/z values representing peptides both smaller and larger than those contained within the predominant size range: several of the signals detected, including for example $(M+2H)^{+2}$ ions at m/z 1116 and $(M+H)^+$ ions at m/z 1573, 481, and 430 respectively, specifically indicate the presence of peptide species which deviate in size from the nonameric motif assigned to B*1501 by pool sequencing.

*General assessment of sequence complexity within the sB*1501-bound peptide pool*

Since consensus class I motifs consist of thousands of various peptides, we subfractionated this complex peptide milieu and then amino acid sequenced peptide subsets to examine diversity around the pooled motif. Edman sequencing of peptide-containing fractions collected from the RP-

Fraction 10

position:	2	3	4	5	6	7	8	9	10	11	12
dominant	P Q	-	-	-	-	-	-	-	-	-	-
strong	W	K R H	G F	-	-	-	-	-	-	-	-
weak	S I	E	D N M I	S W P	H S	V T	I	Y	K	V	E

Fraction 15

position:	2	3	4	5	6	7	8	9	10	11	12
dominant	***A*** P	-	***G***	***Y***	-	-	-	-	-	-	-
strong	Q	***V***	-	M	***V***	I	-	H	-	-	-
weak	-	Y R S	F	Q	R A	D Y	Q F	-	Y	S	R K

Fraction 28

position:	2	3	4	5	6	7	8	9	10
dominant	-	K	-	W	-	-	-	-	-
strong	S Q P	H	N	R M	S Y	F V	-	G	F
weak	L	Y L V	A	-	L K	-	Y S K	Y	-

Fraction 31

position:	2	3	4	5	6	7	8	9	10	11	12
dominant	K	-	-	-	-	-	-	-	-	-	-
strong	Q V P	H N R	G	L S H	R	I	D N	K	V	E	-
weak	-	P F K D	E	P E	V A F	P M	Q E A H W	Y F	-	F P	S K

Fig. 5 Sequence complexity of sB*1501-bound peptides. Each of the fractions in Fig. 4 was subjected to Edman sequencing; the results obtained from fractions 10, 15, 28, and 31 are shown here to illustrate by example some of the diversity and trends observed among constituents of the peptide mixture. The positions called in fraction 15 which form part of a series defining the hexamer are identified in ***bold italics***. *Dashes* are as described in Fig. 2

HPLC gradient shown in Fig. 4 supports the existence of peptides up to 12 residues long and reveals significant positional diversity among the peptides isolated from sB*1501, as data obtained from 16 representative fractions illustrate (Figs. 5, 6). Assessment of dominant, strong, or weak amino acid residues present at the various positions indicates that the dominant anchors previously defined by whole pool sequencing do not predominate in subfractions of the peptide pool (Fig. 5). In subfractions 15 and 31 the dominant P2 glutamine was replaced by a dominant alanine and a dominant lysine respectively, while the pooled motif Y fell below residues such as histidine and lysine at P9 in the same subfractions. Additional diversity among the subfractions included K, R, S, T, Y, and W manifest as P2 occupants (Figs. 5, 6). While variations on the consensus motif were frequent at P2, this diversity was not concentrated at the N-termini of the bound peptides as demonstrated by the occurrences of K, R, V, A, S, and H at P9 (Fig. 6). Multiple other substitutions inconsistent with preferred residues of the motif are also seen, as evident in comparing the pool-derived data in Fig. 2 with information extracted from the subfractions in Figs. 5 and 6.

Of the 16 LC subfractions studied, 12 demonstrated weak sequence yields out to 12 cycles of degradation; in 9 of these fractions, P12 was occupied by the charged or polar residues E, R, K, S, H, or Q (Figs. 5, 6). Additionally, the presence of numerous decamers within the sB*1501 peptide population was suggested in that 11 of the subfractions sequenced exhibited a typical P9 residue Y, F, V, or H at P10. For example, in Fig. 5 the strong F and V presences at P10 in fractions 28 and 31 suggests that P10 serves as an anchor in these decamers. The shifting of a P9 anchor preference to P10 is consistent with the P10 occupancies of individual 10-mers formerly characterized from B*1501 peptide pools (Barber et al. 1997; Falk et al. 1995). Although residues representing P9 anchors are seen at P10 and P12, amino acids unique to these longer ligands (P12 E and Q; P10 D and F) were also detected; the shifting of a previously reported P9 anchor is not the only means by which longer peptides are bound within the groove.

Variation on the pooled motif is not the only interesting result of subfractionating the pool of eluted peptides. For example, fractions 12 and 15 reveal distinct sequence information, indicating that one peptide population predominates in these fractions. Subfraction 12 resulted in the distinct sequence GQRKGAGSVF, a decamer derived from human ribosomal protein L8$_{7-16}$ and previously reported to bind B*1501 (Barber et al. 1997). Subfraction 15 distinctly contained the 6-mer IAVGYV derived from HLA class I α-chain$_{23-28}$; P1 determinations (not shown) for these two fractions were estimated by comparing total amino acid yield percentages for the first cycle of degradation.

Discussion

The typical binding motif obtained from pooled sequencing has been applied as a tool in the prediction of CD8^{+} T lymphocyte epitopes for given class I molecules (Kawakami et al. 1994; Ressing et al. 1995), in essence providing an "anchor template" to guide the development of immunomodulatory therapeutics. While such allele-specific motifs describe ligands which result from the physiological processing and transport preferences inherent to the cell, the resultant pooled consensus sequence offers merely a sketch of the immensely heterogeneous peptide collection. To better understand how representative a peptide motif is of a peptide population, large-scale B*1501-derived peptide extracts were subfractionated and analyzed for size and sequence complexity (Figs. 3, 5). Results show that peptides bound by B*1501 include (1) species which are both

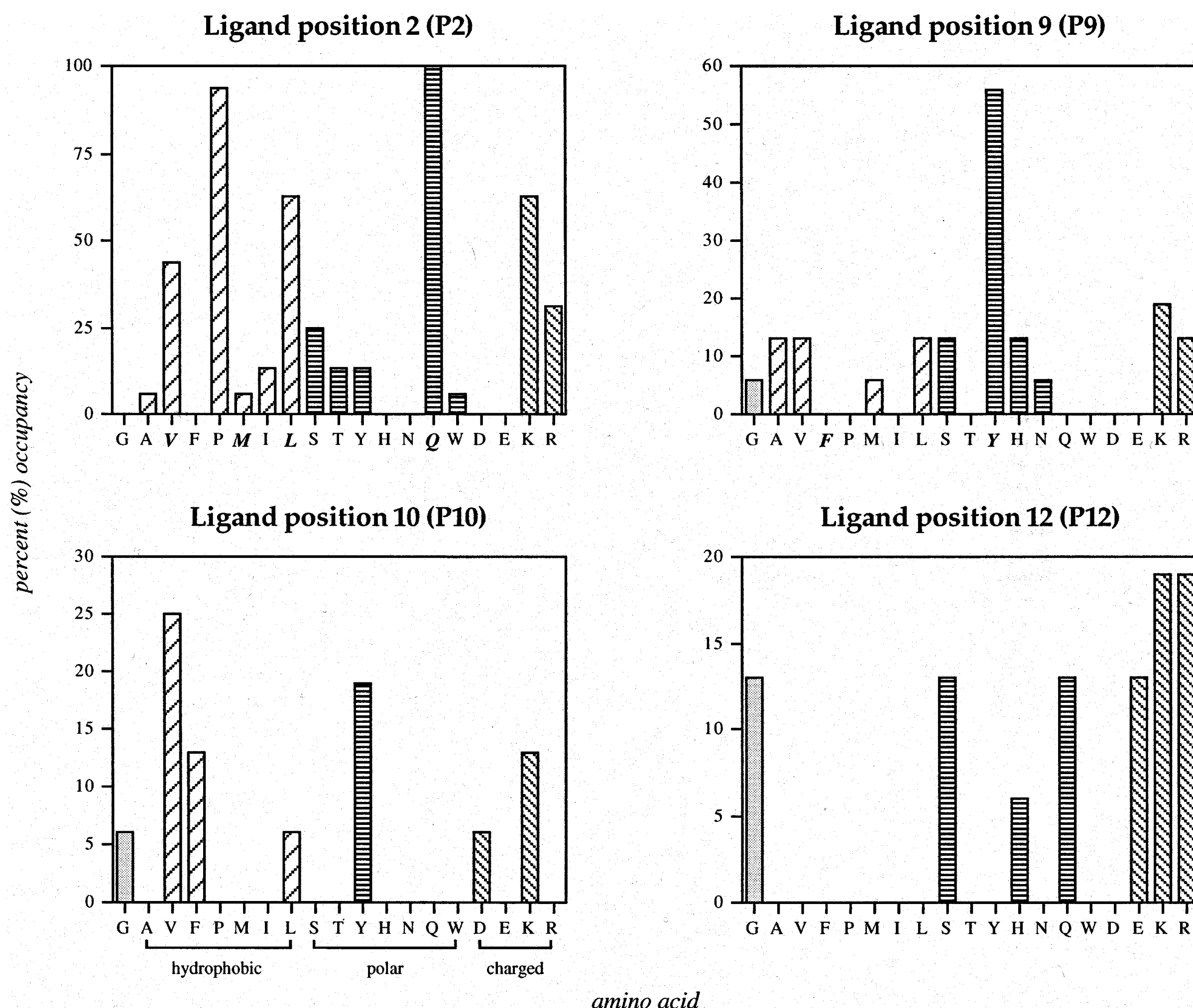

Fig. 6 Percentage of subfractions with a particular amino acid at P2, P9, P10, and P12. Sequence complexity among sB*1501-derived peptides was summarized by averaging the frequency of occurrence for amino acid residues at positions 2 through up to 12 along the ligands among the fractions marked in Fig. 4. Amino acids are grouped along the *x*-axis according to their physicochemical natures (charged, polar, or hydrophobic), as specifically designated in the chart for P10 and indicated within all four charts by *differential shading*. Residues observed at P2 and P9 in former B*1501 motif descriptions (Barber et al. 1997; Falk et al. 1995) are indicated in ***bold italics***. Since it was not derivatized, cysteine was undetectable and is therefore excluded

longer and shorter than the nonameric size typically indicated by pooled sequencing alone, and (2) many species which exhibit primary sequences different from those predicted via whole pool sequencing.

In terms of mass the LC/MS data presented here indicate that the nonameric peptide size previously assigned to B*1501 is generally representative of the peptides bound by B*1501. Approximately 80% of the pooled peptides bound by B*1501 fall into the mass range of a nonamer or decamer, data which are in agreement with what others have noted and with the observation that the closed-ended nature of the class I binding groove restricts the length of the peptides bound (Madden 1995). For the 20% or so of those peptides with masses suggesting lengths of greater than ten amino acids we can only speculate as to how they are positioned within the binding groove. Past evidence has tended to support the bulging of longer peptide ligands outwards centrally from the peptide binding groove (Fremont et al. 1992; Guo et al. 1992; Madden et al. 1993). For B*1501 peptides which generated sequence through ten Edman degradation cycles we often found P2/P10 anchors which coincided with the P2/P9 anchors of nonamers from our subsets (Fig. 6). In contrast, peptide subfractions which generated sequence to 12 amino acids seldom possessed previously detected anchors at both the N- and C-proximal regions of the peptide. Rather, peptide subpools sequencing to 12 cycles typically demonstrate a P2 anchor position which coincides with P2 positions observed among 9-to 10-mers, while charged or polar amino acids are commonly found in the 12th position of longer peptides. Thus, for peptides of ten amino acids in length, our data suggest that peptides are anchored at P2 and P10 with various bulges accommodating slight (single residue) length fluctuations. For peptides of 12 residues in length we typically identify a common anchor at P2 with C-proximal residues exhibiting charged or polar characteristics (Figs. 5, 6). We hypothesize that such peptides of non-traditional length remain anchored at their N-terminus, while the lack of a strong anchor at the opposing terminus of the peptide facilitates

a peptide folding pattern such that the charged or polar portions of longer epitopes extend out of the groove in a solvent-accessible manner.

The whole pool sequencing data herein are in agreement with the dominant P2 and P9 anchors (Q and Y/F, respectively) and with the auxiliary anchor of I at P5 (Falk et al. 1995). This result demonstrates that the peptides we subfractionate coincide with previously characterized B*1501 peptide pools. However, several amino acids in our whole pool sequencing data arising from BBM.1 purified material indicate that a greater number of peptides than originally realized contribute to the consensus B*1501 motif. First, while Q and residues including L, M, and V have been reported at P2, the aliphatic side chain proline is strongly detected here as well. More subtle differences with previous B*1501 motif assignments noted here throughout the interior of the motif include: (1) neither Y nor A were detected at P3; (2) no E was detected at P4; (3) no preferred residues could be defined for P6 or P7, and (4) Q and S were weakly detected at P8 rather than the Y, V, or T called by Falk and co-workers (1995). Primary structural differences become even more pronounced once the pool of B*1501-derived peptides is separated by RP-HPLC prior to Edman sequence analysis, at which point P2 is seen in varying degrees to also accommodate the aliphatic residues A and Y as well as both polar (S, W, and T) and charged (K and R) side chains. As mentioned earlier, P9 exhibits a similar physicochemical flexibility in permitting occupancy by a number of aliphatic, charged, and nonpolar amino acids which fail to be represented by the confines of the motif derived from whole pool sequencing. Furthermore, glycine also appears to be a residue observed frequently at position 4, as demonstrated in three of the four fractions illustrated in Fig. 5 and in 10 of the 16 fractions marked in Fig. 4, where it could either serve as an influential auxiliary anchor or simply be favored due to its unobtrusive size. Thus, subfractionation of the B*1501 peptide pool shows that much diversity surrounds the consensus motif.

The information obtained from examining the B*1501 peptide pool by fractionation leads one to believe that while the typically-defined anchor residues satisfactorily describe the characteristics of the most abundant peptides bound by this class I product, they do not provide an adequate description of the numerous constituents comprising the pool. In fact, as formerly implied (Engelhard 1994), the concept of a motif as described by whole pool and/or strong UV peak sequencing data alone may not even be sufficient for describing minimal rules on ligand binding. Indeed, our data suggest that in many cases binding is governed more by entire primary structures on an individual basis rather than by P2/P9 anchor rules: complex interactions along the entire length of a peptide ligand may be of greater significance to binding than previously credited.

A new dimension to this study is the purification of sHLA complexes with the β_2m-specific mAb BBM.1. We chose BBM.1 to avoid biases imposed upon the class I heavy chain by bound peptides (Bluestone et al. 1992; Solheim et al. 1993). Of interest here is that proline has not been reported before as a strong or even weak P2 anchor in the B*1501 peptide motif. It has been suggested that B-locus subtypes which present peptides with proline at P2 demonstrate a shallower B pocket within their binding grooves than does B*1501, which exhibits serine at α-chain position 67 rather than a more constricting residue such as phenylalanine (Barber et al. 1997). However, we observed proline strongly not only through whole pooled sequencing but dispersed widely among individual fractions derived from the BBM.1-purified pool (for examples, see Fig. 6). Furthermore, proline was also identified at P2 in a B*1501-associated nonamer we previously characterized by MS/MS collision-induced dissociation (Prilliman et al. 1997). This suggests the possibility that proline binds amicably within this deeper pocket but perhaps induces an altered heavy chain conformation that negatively biases purification of complexes by mAbs such as W6/32. The observation of a P2 proline in a pocket with suboptimal physical complementarity is corroborated by a similar occurrence among peptides bound by the murine class I molecule L^d (Balendiran et al. 1997; Corr et al. 1992). Purification methodology could therefore serve as a factor in allele-specific motif predictions, and the whole pool sequencing with peptides extracted from BBM.1 and W6/32-purified B*1501 complexes demonstrates that a strong proline anchor at P2 is antibody dependent.

In summary, both the selection of therapeutic class I peptide ligands and the understanding of how class I diversification functionally acts requires detailed information pertaining to which peptides are presented by individual class I molecules. The data presented here clearly demonstrate that peptide motifs represent a composite of the population while they fail to elucidate variability in the population from which the composite is drawn. Our finding that many peptides naturally bound by B*1501 fail to conform with the pooled motif is relevant given that class I-presented peptide-based therapeutics have so far met with limited success and that the hunt for therapeutic peptides often uses binding motifs as blue prints for modeling immunogenic peptides (Celis et al. 1994; Chicz and Urban 1994; Loftus et al. 1995). Coupled with data showing that (1) peptides which do not fit a collectively-defined allele-specific motif or which are presented on the cell surface in remarkably low quantities are nevertheless efficiently recognized by CTLs (Malarkannan et al. 1996; Wang et al. 1997) and (2) that nonimmunodominant peptides may effectively elicit protective CTL responses (Oukka et al. 1996), the peptide complexity results presented herein indicate that data beyond the standard motif will augment the selection of therapeutic class I peptide ligands.

Acknowledgments We thank Dave Yamane (PE SCIEX) for instrumentation assistance. This study was funded by NIH contract NO-1-AI-45243. K.R.P. was supported by NIH predoctoral training grant 5-T32-AI07364–05.

References

Appella, E., Padlan, E.A., and Hunt, D.F. Analysis of the structure of naturally processed peptides bound by class I and class II major histocompatibility complex molecules. *In* P. Jollès and H. Jörnvall (eds.): *Interface between Chemistry and Biochemistry*, pp. 105–119, Birkhäuser, Basel, 1995

Balendiran, G.K., Solheim, J.C., Young, A.C.M., Hansen, T.H., Nathenson, S.G., and Sacchettini, J.C. The three-dimensional structure of an H-2L^d-peptide complex explains the unique interaction of L^d with beta-2 microglobulin and peptide. *Proc Natl Acad Sci USA 94:* 6880–6885, 1997

Barber, L.D., Gillece-Castro, B., Percival, L., Li, X., Clayberger, C., and Parham, P. Overlap in the repertoires of peptides bound *in vivo* by a group of related class I HLA-B allotypes. *Curr Biol 5:* 179–190, 1995

Barber, L.D., Percival, L., Valiante, N.M., Chen, L., Lee, C., Gumperz, J.E., Phillips, J.H., Lanier, L.L., Bigge, J.C., Parekh, R.B., and Parham, P. The inter-locus recombinant HLA-B*4601 has high selectivity in peptide binding and functions characteristic of HLA-C. *J Exp Med 184:* 735–740, 1996

Barber, L.D., Percival, L., Arnett, K.L., Gumperz, J. E., Chen, L., and Parham, P. Polymorphism in the $\alpha 1$ helix of the HLA-B heavy chain can have an overriding influence on peptide-binding specificity. *J Immunol 158:* 1660–1669, 1997

Barnstable, C.J., Bodmer, W.F., Brown, G., Galfre, G., Milstein, C., Williams, A.F., and Ziegler, A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens. New tools for genetic analysis. *Cell 14:* 9–20, 1978

Björkman, P.J. and Parham, P. Structure, function and diversity of class I major histocompatibility molecules. *Annu Rev Biochem 59:* 253–288, 1990

Björkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., and Wiley, D.C. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature 329:* 512–518, 1987

Bluestone, J.A., Jameson, S., Miller, S., and Dick, R. Peptide-induced conformational changes in class I heavy chains alter major hisotocompatibility complex recognition. *J Exp Med 176:* 1757–61, 1992

Brodsky, F.M., Bodmer, W.F., and Parham, P. Characterization of a monoclonal anti-β2-microglobulin antibody and its use in the genetic and biochemical analysis of major histocompatibility antigens. *Eur J Immunol 9:* 536–545, 1979

Celis, E., Tsai, V., Crimi, C., DeMars, R., Wentworth, P.A., Chestnut, R.W., Grey, H., Sette, A., and Serra, H. Induction of anti-tumor cytotoxic T lymphocytes in normal humans using primary cultures and synthetic peptide epitopes. *Proc Natl Acad Sci USA 91:* 2105–2109, 1994

Chicz, R.M. and Urban, R.G. Analysis of MHC-presented peptides: applications in autoimmunity and vaccine development. *Immunol Today 15:* 155–159, 1994

Corr, M., Boyd, L.F., Frankel, S.R., Kozlowski, S., Padlan, E.A., and Margulies, D.H. Endogenous peptides of a soluble major histocompatibility complex class I molecule, H-2L^{d}s: sequence motif, quantitative binding, and molecular modeling of the complex. *J Exp Med 176:* 1681–1692, 1992

Domena, J.D., Little, A.-M., Madrigal, J.A., Hildebrand, W.H., Johnston-Dow, L., du Toit, E., Bias, W.B., and Parham, P. Structural heterogeneity in HLA-B70, a high frequency antigen of black populations. *Tissue Antigens 42:* 509–517, 1993

Elliott, T., Smith, M., Driscoll, P., and McMichael, A. Peptide selection by class I molecules of the major histocompatibility complex. *Curr Biol 3:* 854–866, 1993

Engelhard, V.H. Structure of peptides associated with class I and class II MHC molecules. *Annu Rev Immunol 12:* 181–207, 1994

Falk, K., Rötzschke, O., Stevanović, S., Jung, G., and Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self peptides eluted from MHC molecules. *Nature 351:* 290–293, 1991

Falk, K., Rötzschke, O., Grahovac, B., Schendel, D., Stevanović, S., Gnau, V., Jung, G., Strominger, J.L., and Rammensee, H.-G. Allele-specific motifs revealed by sequencing of self peptides eluted from MHC molecules. *Nature 90:* 12005–12009, 1993

Falk, F., Rötzschke, O., Takiguchi, M., Gnau, V., Stevanović, S., Jung, G., and Rammensee, H.-G. Peptide motifs of HLA-B58, B60, B61, and B62 molecules. *Immunogenetics 41:* 165–168, 1995

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A., and Wilson, I.A. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science 257:* 919–927, 1992

Guo, H.-C., Jardetzky, T.S., Garrett, T.P.J., Lane, W.S., Strominger, J.L., and Wiley, D.C. Different length peptides bind to HLA-Aw68 similarly at their ends but bulge out in the middle. *Nature 360:* 364–366, 1992

Huczko, E.L., Bodnar, W.M., Benjamin, D., Sakasuchi, K., Zhu, N.Z., Shabanowitz, J., Henderson, R.A., Appella, E., Hunt, D.F., and Engelhard, V. H. Characteristics of endogenous peptides eluted from the class I MHC molecule HLA-B7 determined by mass spectrometry and computer modeling. *J Immunol 151:* 2572–2587, 1993

Hughes, A.L. and Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature 335:* 167–170, 1988

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A.L., Appella, E., and Engelhard, V.H. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science 255:* 1261–1263, 1992

IUPAC-IUB Joint Commission on Biochemical Nomenclature: Nomenclature and symbolism for amino acids and peptides: recommendations 1983. *Biochem J 219:* 345–373, 1984

Jardetzky, T.S., Lane, W.S., Robinson, R.A., Madden, D.R., and Wiley, D.C. Identification of self peptides bound to purified HLA-B27. *Nature 353:* 326–329, 1991

Kavathas, P., Bach, F H., and DeMars, R. Gamma ray-induced loss of expression of HLA and glyoxalase I alleles in lymphoblastoid cells. *Proc Natl Acad Sci USA 77:* 4251–4255, 1980

Kawakami, Y., Eliyahu, S., Sakaguchi, K., Robbins, P.F., Rivoltini, L., Yannelli, J.R., Appella, E., and Rosenberg, S A. Identification of the immunodominant peptides of the MART-1 human melanoma antigen recognized by the majority of HLA-A2-restricted tumor infiltrating lymphocytes. *J Exp Med 180:* 347–352, 1994

Loftus, D.J., Kubo, R.T., Sakaguchi, K., Celis, E., Sette, A., and Appella, E. Analysis of MHC-specific peptide motifs: applications in immunotherapy. *In* M. Z. Atassi and G. S. Bixler (eds.): *Advances in Experimental Medicine and Biology*, pp. 201–210, Plenum Press, New York, 1995

Madden, D. The three-dimensional structure of peptide-MHC complexes. *Annu Rev Immunol 13:* 587–622, 1995

Madden, D.R., Garboczi, D.N., and Wiley, D.C. The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. *Cell 75:* 693–708, 1993

Malarkannan, S., Gonzalez, F., Nguyen, V., Adair, G., and Shastri, N.: Alloreactive CD8+ T cells can recognize unusual, rare, and unique processed peptide/MHC complexes. *J Immunol 157:* 4464–4473, 1996

Mazzocchi, A., Storkus, W.J., Traversari, C., Tarsini, P., Maeurer, M.J., Rivoltini, L., Vegetti, C., Belli, F., Anichini, A., Parmiani, G., and Castelli, C. Multiple myeloma-associated epitopes recognized by HLA-A3-restricted CTLs and shared by melanomas but not melanocytes. *J Immunol 157:* 3030–3038, 1996

Nei, M. and Hughes, A.L. C1.7: balanced polymorphism and evolution by the birth-and-death process in the MHC loci. *In* K. Tsuji, M. Aizawa, and T. Sasazuki (eds.): *Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, pp. 27–38, Oxford University Press, New York, 1992

Oukka, M., Manuguerra, J.C., Livaditis, N., Tourdot, S., Riche, N., Vergnon, I., Cordopatis, P., and Kosmatopoulos, K. Protection against lethal viral infection by vaccination with nonimmunodominant peptides. *J Immunol 157:* 3039–3045, 1996

Parham, P. C1.12: Evolution of class I HLA polymorphisms: selection and drift. *In* K. Tsuji, M. Aizawa, and T. Sasazuki (eds.): *Proceedings of the Eleventh International Histocompatibility Workshop and Conference*, pp. 72–82, Oxford University Press, New York, 1992

Parham, P., Lawlor, D.A., Salter, R.D., Lomen, C.E., and Ennis, P.D. HLA-A,B,C: patterns of polymorphism in peptide binding proteins. *In* B. Dupont (ed.): *Immunobiology of HLA*, pp. 10–33, Springer, New York Berlin Heidelberg, 1989

Parham, P., Adams, E.J., and Arnett, K.L. The origins of HLA-A, B, and C polymorphism. *Immunol Rev 143:* 141–180, 1995

Parham, P. and Ohta, T. Population biology of antigen presentation by MHC class I molecules. *Science 272:* 67–74, 1996

Prilliman, K., Lindsey, M., Zuo, Y., Jackson, K.W., Zhang, Y., and Hildebrand, W. Large-scale production of class I bound peptides: assigning a signature to HLA-B*1501. *Immunogenetics 45:* 379–385, 1997

Rammensee, H.-G., Falk, K., and Rötzschke, O. Peptides naturally presented by MHC class I molecules. *Annu Rev Immunol 11:* 213–244, 1993

Rammensee, H.-G., Friede, T., and Stevanović, S. MHC ligands and peptide motifs: first listing. *Immunogenetics 41:* 178–228, 1995

Ressing, M.E., Sette, A., Brandt, R.M.P., Ruppert, J., Wentworth, P.A., Hartman, M., Oseroff, C., Grey, H.M., Melief, C.J.M., and Kast, W.M. Human CTL epitopes encoded by human papillomavirus type 16 E6 and E7 identified through in vivo and in vitro immunogenicity studies of HLA-A*0201-binding peptides. *J Immunol 154:* 5934–5943, 1995

Solheim, J.C., Carreno, B.M., Smith, J.D., Gorka, J., Myers, N.B., Wen, Z., Martinko, J.M., Lee, D.R., and Hansen, T. Binding of peptides lacking consensus anchor residue alters H-2L^d serologic recognition. *J Immunol 151:* 5387–5397, 1993

Stevanović, S. and Jung, G. Multiple sequence analysis: pool sequencing of synthetic and natural peptide libraries. *Anal Biochem 212:* 212–220, 1993

Townsend, A. and Bodmer, H. Antigen recognition by class I restricted T lymphocytes. *Annu Rev Immunol 7:* 601–624, 1989

Townsend, A.R.M., Gotch, F.M., and Davey, J. Cytotoxic T cells recognize fragments of the influenza nucleoprotein. *Cell 42:* 457–467, 1985

Walker, B.D., Flexner, C., Paradis, T.J., Fuller, T.C., Hirsch, M.S., Schooley, R.T., and Moss, B. HIV-1 reverse transcriptase is a target for cytotoxicT-lymphocytes in infected individuals. *Science 240:* 64–66, 1988

Wang, W., Gulden, P.H., Pierce, R.A., Shabanowitz, J.A., Man, S.T., Hunt, D.F., and Engelhard, V.H. A naturally processed peptide presented by HLA-A*0201 is expressed at low abundance and recognized by an alloreactive CD8$^+$ cytotoxic T cell with apparent high affinity. *J Immunol 158:* 5797–5804, 1997