# Amino Acids Specifying MHC Class Preference in TCR V$\alpha$2 Regions[1]

**Margarida Correia-Neves, Caroline Waltzinger, Jean-Marie Wurtz, Christophe Benoist,[2] and Diane Mathis[2]**

**Some TCR variable regions are preferentially expressed in CD4$^+$ or CD8$^+$ T cells, reflecting a predilection for interacting with MHC class II or class I molecules. The molecular basis for MHC class bias has been studied previously, in particular for V$\alpha$3 family members, pointing to a dominant role for two amino acid positions in complementary-determining regions (CDRs) 1 and 2. We have evaluated the generality of these findings by examining the MHC class bias of V$\alpha$2 family members, an attractive system because it shows more variability within the CDR1 and -2, exhibits variation in the framework regions, and includes a member for which the crystal structure has been determined. We find that preferential recognition of MHC class I or II molecules does not always depend on residues at the same positions of CDR1 and -2; rules for one family may be reversed in another. Instead, there are multiple influences exerted by various CDR1/2 positions as well as the CDR3s of both the TCR $\alpha$- and TCR $\beta$-chains.** ***The Journal of Immunology,*** **1999, 163: 5471–5477.**

T lymphocytes recognize Ags in the context of MHC molecules (1). T cells that bear a CD8 coreceptor express an Ag-specific receptor able to bind MHC class I molecules, while CD4$^+$ T cells display TCRs that engage MHC class II molecules. The molecular interactions that underlie this dichotomy are still poorly understood.

The expression of several mouse TCR variable gene segments is differential in the CD4$^+$ and CD8$^+$ subsets. For example, most T cells expressing V$\alpha$11.1 are found in the CD4$^+$ compartment; conversely, V$\alpha$3.2$^+$ cells are more frequent in the CD8$^+$ population (reviewed in Ref. 2). The differences suggest that germline-encoded V regions have an intrinsic bias for interaction with MHC class I or II molecules. This preference can vary for the different alleles of a given V region, but is mostly independent of MHC haplotype (3). These properties are quite consistent with recently described crystal structures of TCR/peptide/MHC class I complexes, which have suggested that all TCRs dock on MHC molecules in a common manner, regardless of the particular MHC molecule or the receptor composition or specificity (4–7).

Two hypotheses might explain the MHC class bias of TCR V regions: 1) the conformation of the V region or of particular V residues on its surface favors the interaction with either MHC class I or II complexes; and 2) a differential interaction with CD4 or CD8 indirectly promotes engagement with MHC class I or II molecules. Evidence for the former idea has been presented recently; Sim et al. showed that the opposite MHC class preferences of two members of the V$\alpha$3 family result from only two amino acid differences in complementarity-determining regions 1 (CDR1)[3] and 2 (8); extrapolation from the crystal structures mentioned above indicates that these residues most likely contact peptide/MHC, not the coreceptors. This study presented the first evidence for a decisive role for specific residues located within CDR1 or CDR2 in distinguishing MHC class.

These results raised the question of whether particular V$\alpha$ positions play a general role in MHC class discrimination. How important are these residues for recognition of MHC class by other V$\alpha$ families? What is the influence of residues outside of the V$\alpha$ CDR1/2 regions? In other words, are there conserved modes of interaction, and thus conserved contacts, that underlie recognition of MHC class I vs II molecules?

To address these questions, we sought a V$\alpha$ family with more variability within CDR1/2 than in the V$\alpha$3 family and also with variation in the framework regions (FRs). The V$\alpha$2 family consists of seven expressed members, according to an analysis of the B10.A mouse (9). The amino acid sequences in the V$\alpha$2-coding region are well conserved, but exhibit more variability than in the V$\alpha$3 family, with both conservative and nonconservative amino acid replacements in CDR1/2 as well as the FRs. Furthermore, the three-dimensional structure of a TCR using V$\alpha$2.3 has been reported, providing a sound structural backdrop for data interpretation (10). Studies with the B20.1 mAb, directed against V$\alpha$2 family members, showed that the family as a whole is preferentially displayed on CD4$^+$ T cells (11), but preliminary indications from transgenic (tg) mice indicated that the bias may not be shared by all family members (M. Correia-Neves, unpublished observations). Thus, the V$\alpha$2 family offers an attractive system for evaluating the impact of particular V region residues on MHC class preference.

In this study we limited our analysis to the TCR $\alpha$-chain by using a tg mouse line expressing an already rearranged TCR$\beta$ gene. We examined the distribution of V$\alpha$2 family members in the CD4$^+$ and CD8$^+$ T cell compartments and observed a differential

Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, Strasbourg, France

Received for publication June 21, 1999. Accepted for publication August 26, 1999.

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

[1] This work was supported by institute funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Hôpital Universitaire de Strasbourg, Bristol-Myers Squibb, and a grant (to C.B. and D.M.) from the Association pour la Recherche sur le Cancer. M.C.-N. was funded by the Portuguese Gulbenkian Ph.D. Program in Biology and Medicine and the Fundação para a Ciência e Tecnologia in Portugal.

[2] Address correspondence and reprint requests to Dr. Christophe Benoist or Dr. Diane Mathis, Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch Cedex, France. E-mail address: cbdm@igbmc.u-strasbg.fr

[3] Abbreviations used in this paper: CDR, complementary-determining region; B6, C57BL/6; FR, framework region; LN, lymph node; tg, transgenic.



0022-1767/99/$02.00

expression of family members, which we interpreted in the context of the crystal structure of Vα2.

## Materials and Methods

### *DNA constructs and transgenesis*

The rearranged Vβ5.2Dβ2Jβ2.6 segment encoding the variable region of the TCR β-chain expressed by the $CD8^+$ cytotoxic T cell clone B3 (12), specific for the chicken OVA SIINFEKL peptide in the context of H-$2K^b$, was engineered from genomic DNA and cloned into a TCR β-chain expression cassette (13). The β-chain gene fragment was then excised from the plasmid and injected into fertilized B6×SJL F2 eggs. Screening of tg founders was performed on Southern blots with a 1.6-kb *Eco*RI fragment from the germline Jβ2 region. One founder was obtained that expressed a $V\beta5^+$ TCR in >99% of the T lymphocytes. Its offspring were crossed with a line carrying a null mutation in the TCRα locus ($C\alpha^{o/o}$) (14) to obtain $V\beta5^+$ tg mice heterozygous for the Cα mutation. These mice were then crossed with C57BL/6 (B6) animals, and $V\beta5^+$ $C\alpha^{o/+}$ progeny were used.

### *Cell staining and sorting*

Expression of Vβ5.2 was monitored by flow cytometric analysis of lymphocytes using the anti-Vβ5 mAb MR9-4. Thymocytes were sorted after staining with FITC-labeled anti-CD8α; PE-labeled anti-CD4 (Caltag Laboratories, South San Francisco, CA), and B20.1, specific for the Vα2 family (11), were revealed by Cy5-conjugated anti-rat IgG (Jackson ImmunoResearch Laboratories, West Grove, PA). Peripheral lymphocytes were sorted after four-color staining with the same Abs plus biotinylated anti-CD62L (Caltag Laboratories), followed by streptavidin-Texas Red (Jackson ImmunoResearch Laboratories). Cell sorting was performed on a Coulter Elite cytometer equipped with an automatic cell deposition unit (Coulter, Hialeah, FL). This deposition system was programmed to sort single cells directly into wells of microtiter PCR plates containing 10 μl of RT mix.

### *Single-cell RT-PCR*

The RT-PCR method was an adaptation of the protocol of Chang et al. (15). RT was performed in 10 μl of RT buffer (25 mM Tris-HCl, 37.5 mM KCl, and 1.5 mM $MgCl_2$) containing 2% Triton X-100, 1 μg of BSA, 500 μM dNTP, 50 ng of $oligo(dT)_{12–18}$, 8 U of RNasin, and 30 U of Moloney murine leukemia virus reverse transcriptase (Life Technologies, Grand Island, NY). As soon as the sort was complete, the plates were incubated for 90 min at 37°C in a humidified incubator. For the first PCR round, 10 μl of the resulting cDNA was amplified by adding 40 μl of Taq buffer (50 mM KCl, 10 mM Tris-HCl pH 9, 3 and 2, 5 mM $MgCl_2$) containing 2.5 U Taq polymerase, 500 μM dNTP, 400 ng of sense primer specific for the Vα2 family (5′-CAGCAGCAGGTGAGACAAAGT-3′), and antisense primer specific for the TCRα constant region (5′-GTTTTGTCAGTGAT GAACGT-3′; 3 min at 93°C; 35 cycles: 30 s at 93°C, 45 s at 50°C, and 30 s at 72°C; 10 min at 72°C). For the second PCR round, 2 μl of the first amplification product was amplified in 50 μl of Taq buffer containing 1 U of Taq polymerase, 500 μM dNTP, 200 ng of sense (5′-AAGGC CCGGGTCTCTGACAGTCTGGGAAGGA-3′) and antisense (5′-AATCTGCAGCGGCACATTGATTTGGGA-3′) nested primers (3 min at 93°C; 22 cycles: 30 s at 93°C, 30 s at 50°C, and 30 s at 72°C; 10 min at 72°C). The PCR products were purified by polyethylene glycol (20% polyethylene glycol 6000 and 2.5 M NaCl) precipitation followed by two washes in 75% ethanol. The PCR products were sequenced using one of the primers for the second round of PCR (5′-AATCTGCAGCGGCACATT GATTTGGGA-3′). In a few instances the Vα2 family members were identified not by sequencing but by hybridization with member-specific oligonucleotides (details available upon request).

### *Model building and surface generation*

Models for Vβ5-Vα2 TCR were generated by homology from the coordinates of the KB5-C20 TCR (10) using the Modeller package (version 4.0) (16); the original TCR β-chain was substituted by the Vβ5 chain expressed in the transgenics (12). A model for each Vα2 family member was then realized. The various Vα2 sequences were aligned (9) (no gaps or insertions), and modeling by satisfaction of spatial restraints was used to generate the three-dimensional structure of the Vα2 members using the automatic features of Modeller 4.0. Briefly, the spatial restraints are first derived from the known three-dimensional template structure (e.g., Cα-Cα distances, and main chain and side chain dihedral angles) and from the statistical analysis of the relationship among various features of protein structures derived from a database of well-resolved structures. These are expressed as conditional probabilities' distribution, which are used to optimize the initial peptide chain of the target sequence. The final model is obtained by optimizing the molecular probability density function by summing of the probability density functions of the individual features (bond length, valence angle, van der Waals contacts, etc.). No additional energy minimization was performed. Views of predicted Vα2.2 and Vα2.6 TCR surfaces facing the peptide/MHC complex were generated with GRASP software (17) (http://TRANTOR.bioc.columbia.edu/grasp).

## Results

### *Vα2 segments are differentially expressed by $CD4^+$ and $CD8^+$ T cells*

To focus on the influence of TCR α-chain variable residues on MHC class preference, we maintained the TCR β-chain constant by employing mice carrying a rearranged TCRβ transgene. Fixing one TCR chain to highlight the influence of residues in the other is a strategy that has been exploited successfully in several contexts (12, 18, 19). For example, a $V\beta5^+$ TCR tg mouse line was used to study the TCR α-chain regions important for recognition of the $OVA_{257–264}$ peptide presented by $K^b$ molecules (12). The TCR β-chain encoded by this transgene pairs well, even preferentially, with $V\alpha2^+$ TCR α-chains (our unpublished observations). Therefore, we chose to use a tg line expressing this $V\beta5^+$ chain for our experiments, actually a novel line established for another project, expressing the same Vβ region (including the junctional and J regions), but carried within a different expression cassette (13) (M. Correia-Neves, unpublished observation). The phenotype of our $V\beta5^+$ TCR tg line on the B6 background is very similar to that described for the published line, with expression of the transgene-encoded β-chain in >99% of T lymphocytes. For the experiments detailed below, the mice also bore a null mutation at the TCRα locus (14) in the heterozygous state. Thereby, only one TCRα locus of the B6 haplotype can rearrange, avoiding misassignments due to cells expressing two rearranged TCRα-chains, a common occurrence (20).

Vα2 family member usage was analyzed by single-cell RT-PCR. Single-cell analysis was chosen for quantitative estimates, rather than the more commonly used approach of batch amplification and cloning, because it alone can give reliable frequencies. We also found an extremely high incidence of chimeric product formation during the amplification when starting from mixed populations. The Vα segments from single cells were amplified with primers equally active with all Vα2 family members, and the identity of the amplified products was determined by direct sequencing or by hybridization with allele-specific oligonucleotides. The sequences showed very diverse junctional regions and are thus representative of broad repertoires, free of any bias due to ongoing immune responses (with one exception, see below).

Our first observation was that the TCRα haplotype from B6 mice, present in the $V\beta5^+$ TCR tg mice, encodes the same seven Vα2 family members as described for the B10.A mouse strain (9). This finding is consistent with restriction fragment length polymorphism data (reviewed in Ref. 21). Thus, it was easy to assign each of the B6 sequences to the previously described family members on the basis of nucleotides at defined positions in the Vα region. We consequently use the nomenclature of Gahery-Segard et al. (9), which is internally consistent, rather than that of Arden et al. (22), which does not distinguish family members from allelic variants.

We analyzed Vα2 gene segment usage by mature $CD4^+CD8^-V\alpha2^{high}$ and $CD4^-CD8^+V\alpha2^{high}$ thymocytes from the same $V\beta5.2^+C\alpha^{+/0}$ tg mice. Results from 123 $CD4^-CD8^+$ and 118 $CD4^+CD8^-$ mature single-positive thymocytes are represented in Fig. 1*A*. (The full sequence data are presented as supplementary material on http://biblio-igbmc.u-strasbg.fr/cbdm.)

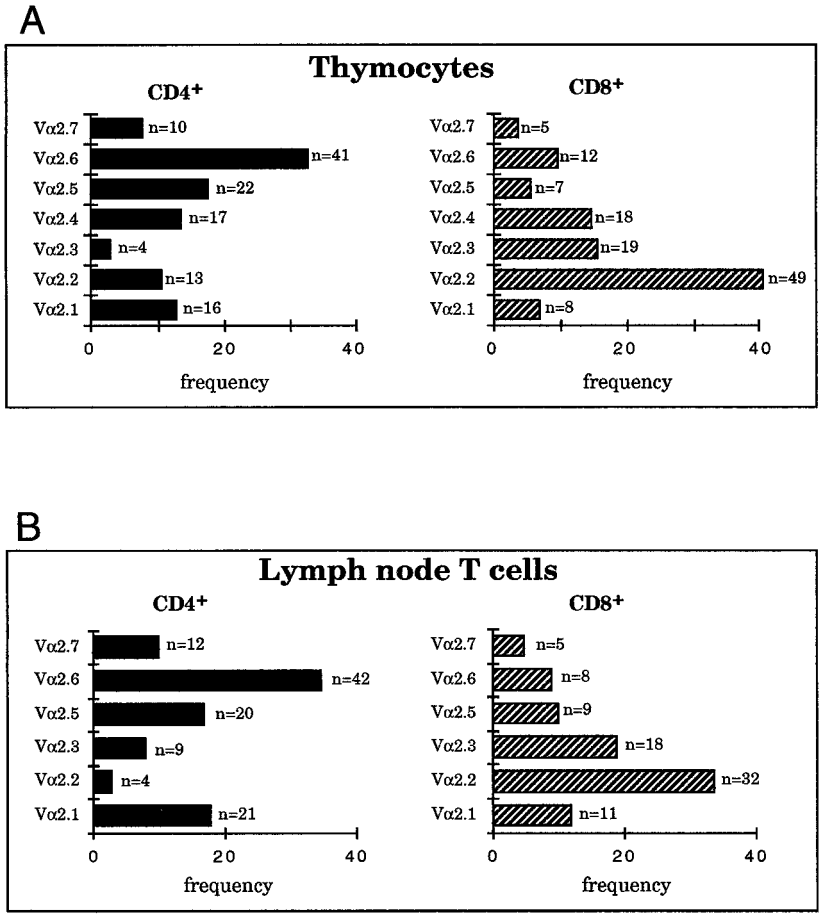



**FIGURE 1.** Frequency of different V$\alpha$2 family members in individual $CD4^+$ and $CD8^+$ T cells. LN T cells were sorted as $V\alpha2^+CD4^+CD8^-CD62L^{high}$ or $V\alpha2^+CD4^-CD8^+CD62L^{high}$, and thymocytes were sorted as $V\alpha2^+CD4^+CD8^-$ and $V\alpha2^+CD4^-CD8^+$. The actual V$\alpha$2 family member expressed by each cell was determined by single-cell RT-PCR and sequencing (or hybridization with member-specific oligonucleotides for cells from one mouse). Data are pooled from four TCR-V$\beta$5 transgenic mice heterozygous for the TCR$\alpha$ knockout mutation. The pattern of V$\alpha$2 family member usage was similar in all mice studied. The nomenclature follows that used by Gahery-Segard et al. for V$\alpha$2 family members of the B10.A TCR$\alpha$ haplotype (9), rather than the confusing nomenclature used by Arden et al. (22), which mixes isotypic and allelic variants. For those sequences that are present in the Arden et al. listing, V$\alpha$2.1 = AV2S1, V$\alpha$2.2 = AV2S4, V$\alpha$2.3 = AV2S7, V$\alpha$2.4 = AV2S8, and V$\alpha$2.7 = AV2S3.

The pattern of expression was quite different in the two compartments; one was almost a mirror image of the other. The family members showing the strongest bias were V$\alpha$2.6, found preferentially in mature $CD4^+$ single positives, and V$\alpha$2.2, which dominates the $CD8^+$ subset.

Are these biases maintained in the mature pools once they have exited from the thymus, or are they reinforced by the continued MHC engagement known to be required for survival in the periphery? To minimize artifacts due to expansion of particular clones during an ongoing immune response, we restricted our analysis of peripheral populations to naive T lymphocytes, selected on the basis of their $CD62L^{high}$ phenotype. V$\alpha$2 family member usage was determined by RT-PCR on single $CD4^+CD8^-V\alpha2^+$-$CD62L^{high}$ or $CD4^-CD8^+V\alpha2^+CD62L^{high}$ cells. The V$\alpha$2 family members expressed by 116 $CD4^+$ and 94 $CD8^+$ lymph node (LN) T cells are presented in Fig. 1*B*. The pattern of V$\alpha$2 usage was very similar to that found for thymocytes, with V$\alpha$2.2 and V$\alpha$2.6 again showing the most extreme skewing. Thus, the differential usage of V$\alpha$2 family members is imparted during thymocyte differentiation, and the pattern is subsequently conserved in the periphery.

Unlike the sequences of all other family members, the majority of which had diverse J region usage and a unique sequence at the junctional region, V$\alpha$2.4 sequences from LN $CD8^+$ cells were highly skewed in the three mice analyzed: 75% (21 of 28) of the $CD8^+V\alpha2.4^+$ cells incorporated the J$\alpha$44 segment and had very related CDR3 regions (with typical LTGANTGKL or SXDT GANTGKL CDR3 motifs; sequences can be found at http://biblio-igbmc.u-strasbg.fr/cbdm). This phenomenon was not observed in thymocyte samples, in which the canonical sequence was seen only once in 18 sequences, and presumably reflected peripheral amplification of cells with a particular specificity, reactive to self or to a foreign Ag to which these mice were exposed. This amplification prevented a reliable analysis of the $CD4^+/CD8^+$ distribution of the V$\alpha$2.4 segment, and so the frequencies of this family member were not included in Fig. 1*B*.

The differences observed in the expression of V$\alpha$2 family members would seem to denote preferential interactions of individual

| | Jα | 6 | 7 | 8 | 10 | 11 | 12 | 14 | 15 | 16 | 18 | 21 | 23 | 24 | 25 | 26 | 27 | 30 | 32 | 33 | 34 | 35 | 37 | 38 | 39 | 40 | 41 | 42 | 44 | 47 | 48 | 49 | total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Vα2.2 | CD4 | . | . | . | . | . | . | . | . | . | 1 | 1 | . | 1 | . | . | 1 | 2 | . | 1 | . | . | . | . | . | . | 1 | . | . | 1 | . | . | 9 |
| | CD8 | . | 4 | 1 | 1 | . | 2 | . | . | 1 | 1 | . | 4 | 6 | . | . | 2 | 1 | 5 | . | . | 5 | 1 | 1 | . | 1 | 1 | . | 1 | . | 1 | 1 | 40 |
| Vα2.6 | CD4 | . | . | . | . | 1 | . | 1 | 2 | 1 | 1 | 8 | 2 | 4 | 3 | 2 | 3 | 4 | 3 | 5 | 3 | . | . | 1 | 4 | . | 1 | 5 | . | . | . | . | 54 |
| | CD8 | 1 | . | . | . | . | . | . | . | . | 1 | 1 | . | . | . | . | . | . | 1 | . | 2 | 8 | . | . | . | . | . | . | 1 | . | . | . | 15 |
| | total | 1 | 4 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 4 | 10 | 6 | 11 | 3 | 2 | 6 | 7 | 9 | 6 | 5 | 13 | 1 | 2 | 4 | 1 | 3 | 5 | 2 | 1 | 1 | 1 | 118 |

**FIGURE 2.** J$\alpha$ gene segments used by $V\alpha2.2^+$ and $V\alpha2.6^+$ T cells. The sequences were derived from thymocytes and lymphocytes from three mice (data from the experiments in Fig. 1). The J$\alpha$ segments are numbered as previously described (30). The number of occurrences of each segment is indicated.

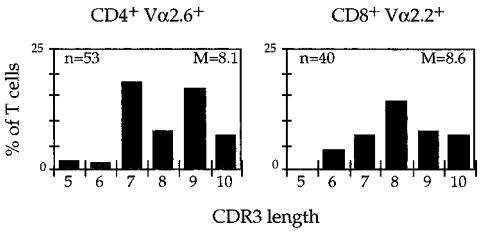



**FIGURE 3.** Length of the CDR3 segments in Vα2.2$^+$ and Vα2.6$^+$ sequences (data from the experiments in Fig. 1). N, number of sequences analyzed; M, average CDR3 length in each group.

family members with class I or class II MHC molecules. However, it could be argued that the influence is indirect. For example, a given Vα2 family member could rearrange preferentially with a particular Jα gene segment, this Jα encoding a CDR3 region with a favored interaction with either MHC class. Such a preferential mode of rearrangement could also yield particular CDR3 lengths or composition. The bias in Vα2 family member distribution would then only be secondary, and not reflect sequence variation in the V-coding region per se. To address this caveat, Fig. 2 depicts, for thymocytes and LN T cells, the Jα usage in Vα2.2$^+$ and Vα2.6$^+$ TCRs, the two Vα2 family members exhibiting opposite biases. A total of 31 Jα genes were identified among 116 TCRα sequences from 63 CD4$^+$ and 55 CD8$^+$ cells. There are some biases in Jα segment usage between CD4$^+$ and CD8$^+$ cells (e.g., Jα21 and Jα35), yet it is clear that these do not suffice to explain the fundamental differences between Vα2.2 and Vα2.6 usage. Similarly, there are no notable differences in the lengths (Fig. 3) or composition (Fig. 4) of the CDR3 regions of TCRs using the Vα2.2 and Vα2.6 segments. We did note some divergence in CDR3 composition between TCRs from CD4$^+$ and CD8$^+$ cells, such as a high frequency of acidic residues at position 2 for CD8$^+$ cells, but this is a general characteristic of TCRs in CD8$^+$ T cells and does not correlate with the Vα2 family member involved. Thus, the differential distribution of Vα2 family members between CD4$^+$ and CD8$^+$ T cells is a direct outcome of preferential interactions they engage in and cannot merely be attributed to indirect effects due to favored rearrangement patterns.

### *Other TCR elements influence Vα2 family member distribution between CD4$^+$ and CD8$^+$ cells*

Usage of the Vα2 family members in the CD4$^+$ vs CD8$^+$ T cell compartments is skewed, but not in an absolute fashion; even those members showing the strongest bias toward one subset can be found in the other. Since the TCR β-chain is fixed in our experimental system, it must be that the precise composition and organization of the junctional (CDR3) region reverse the natural predilection in these instances. A close examination of the CDR3 sequences did not reveal any particularly striking feature in the "revertant" CDR3s (not shown).

We wondered how general these observations might be and whether the Vα2 family member bias would still be present if the β-chain were allowed to vary. These questions were addressed by sorting CD4$^+$ or CD8$^+$ cells expressing diverse, randomly rearranged, Vβ5$^+$ TCR β-chains from nontransgenic littermates of the TCRβ tg line described above (also heterozygous for the TCRα null mutation). The Ab used to select Vβ5$^+$ cells (MR9-4) recognizes both the Vβ5.1 and Vβ5.2 variable regions, which are 82% identical at the amino acid level. As in the experiments described above, single-cell RT-PCR was performed to amplify Vα2 chains expressed in sorted CD4$^+$CD8$^-$Vβ5$^+$Vα2$^+$ and CD4$^-$CD8$^+$Vβ5$^+$Vα2$^+$ thymocytes (115 and 88 cells, respectively). The pattern of Vα2 genes used (Fig. 5) is quite different from the one described above (cf., Fig. 1). Vα2.6 is no longer the most frequent member in the CD4$^+$ population, nor is Vα2.2 preferentially used by CD8$^+$ T cells. In this context, most family members do not show any preferential distribution, with the exception of Vα2.7, which is significantly more frequent in CD4$^+$ cells. Therefore, the skewed pattern of Vα2 family member usage observed in Vβ5$^+$ TCR tg mice is dependent on the transgene-encoded β-chain, in particular on the CDR3 region. Since much of the data in Fig. 5 come from TCRs using the same Vβ5.2 gene segment as the transgene-encoded receptors, that there is no trace of the skewing in the TCR tg mice implies that variations in the β-chain CDR3 exert a strong effect in abolishing the bias.

CD4$^-$CD8$^+$

| Vα2 | CDR3 |
|---|---|
| 1 | SPYAQGL |
| 1 | SANYXVL |
| 1 | REYSNNRL |
| 1 | SERGGSAKL |
| 2 | SSNNRI |
| 2 | SANNRI |
| 2 | SETGKL |
| 2 | SGSNNRI |
| 2 | SPNNYAP |
| 2 | SNNNNAP |
| 2 | SYNAGAKL |
| 2 | SDNAGAKL |
| 2 | NYQGGRAL |
| 2 | SEGGGYKV |
| 2 | TSNMGYKL |
| 2 | HDTNAYKV |
| 2 | SDNNNNAP |
| 2 | SANYNQGKL |
| 2 | SAGSNYYVL |
| 2 | SRQGGSAKL |
| 2 | SDLGSGGKL |
| 2 | SRSSNTNKV |
| 2 | SDPTGYQNF |
| 2 | SAQGTNAYKV |
| 2 | SPPSTNAYKV |
| 2 | SDTGANTGKL |
| 3 | SALSNNRI |
| 3 | RDNAGAKL |
| 3 | SEGGGYKV |
| 3 | SENYNQGKL |
| 3 | SGGYSNYRL |
| 3 | SDNTNTGKL |
| 3 | SARVNYAQGL |
| 3 | SDPGTGSYXL |
| 3 | SAASGGNYKP |
| 3 | SDSGGSNAKL |
| 4 | REGTGSKL |
| 4 | LTGANTGKL |
| 4 | RAHNNAGXXL |
| 4 | RGYNNAGAKL |
| 4 | RPNQGGSAKL |
| 4 | LDSGGSNAKL |
| 4 | SPHTGANTXL |
| 5 | SPNNQGKL |
| 6 | SDNNNAP |
| 6 | SDNNNAP |
| 6 | GDNNNNAP |
| 6 | SEGNNNAP |
| 6 | SEPNNNAP |
| 6 | SSNAGAKL |
| 6 | SRRHNNNAP |
| 6 | RYSGGSNAKL |
| 6 | SDRNYNQGKL |
| 6 | SSDTGANTGXL |

CD4$^+$CD8$^-$

| Vα2 | CDR3 |
|---|---|
| 1 | TVAGNKL |
| 1 | YQGGRAL |
| 1 | THSNNRI |
| 1 | IDGSGGKL |
| 1 | SEANTGKL |
| 1 | THSNXRIF |
| 1 | SAGNMGYKL |
| 1 | TPNSAGNKL |
| 1 | SAGNMGYKL |
| 1 | SSGNMGYKL |
| 2 | TGNTGKL |
| 2 | SANTNKV |
| 2 | SNTGNYKY |
| 2 | SSNSNNRI |
| 2 | FPGNTGKL |
| 2 | RDPTGGNNKL |
| 2 | SRRNTGYQNF |
| 3 | SAGTNTGKX |
| 3 | KDQGGSAKL |
| 4 | SPGYQNF |
| 4 | EGADRL |
| 4 | SANSNNRI |
| 4 | RDTGNYKY |
| 4 | SGNTGYQXF |
| 4 | MATGGNXKL |
| 4 | KDQGGSAKL |
| 4 | CTGANTXKL |
| 5 | SANSNNRI |
| 5 | STGNYKY |
| 5 | SGGGNYKY |
| 5 | MNYNQGKL |
| 5 | SANSNXRX |
| 5 | INYNQGKL |
| 5 | SAFSNTYKV |
| 6 | SGSNNRI |
| 6 | STGNYKY |
| 6 | QDYANKM |
| 6 | SPNTNKV |
| 6 | SGNTGKL |
| 6 | YTNTGKL |
| 6 | SPGYQNF |
| 6 | SVNSGTYQ |
| 6 | LNYNQGKL |
| 6 | SDTNTGKL |
| 6 | NNNAGAKL |
| 6 | RNNNAGAKL |
| 6 | SDNSAGNKL |
| 6 | SVASSFSKL |
| 6 | SVSSSFSKL |
| 6 | SLNTGNYKY |
| 6 | SATGNTGKL |
| 6 | SGDTNTGKL |
| 6 | SGDTNTGKL |
| 6 | SGSGGSNAKL |
| 6 | SGSGGSNAKL |
| 7 | SYGGYQNF |
| 7 | SGTGYQNF |
| 7 | SEGNMGYKL |

**FIGURE 4.** CDR3 amino acid sequences for thymocytes expressing Vα2$^+$ α-chains, grouped according to length (representative data from one mouse in the experiments of Fig. 1). Charged amino acids are highlighted. Sequences from LN cells from two other mice showed a similar pattern.

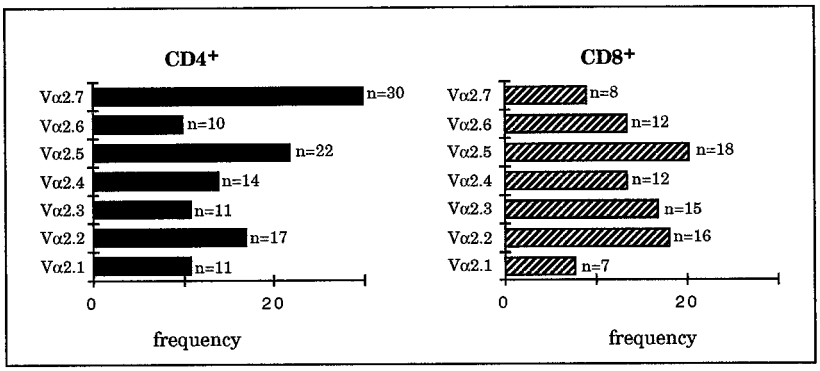



**FIGURE 5.** Vα2 family member usage by thymocytes from nontransgenic mice. Data are pooled from four mice heterozygous for the TCRα knockout mutation (nontransgenic littermates of the animals analyzed in Fig. 1). Thymocytes were sorted as $V\alpha2^+CD4^+CD8^-$ and $V\alpha2^+CD4^-CD8^+$. Single-cell RT-PCR was performed, and the Vα2 PCR products were sequenced.

## Discussion

*Basis of the bias*

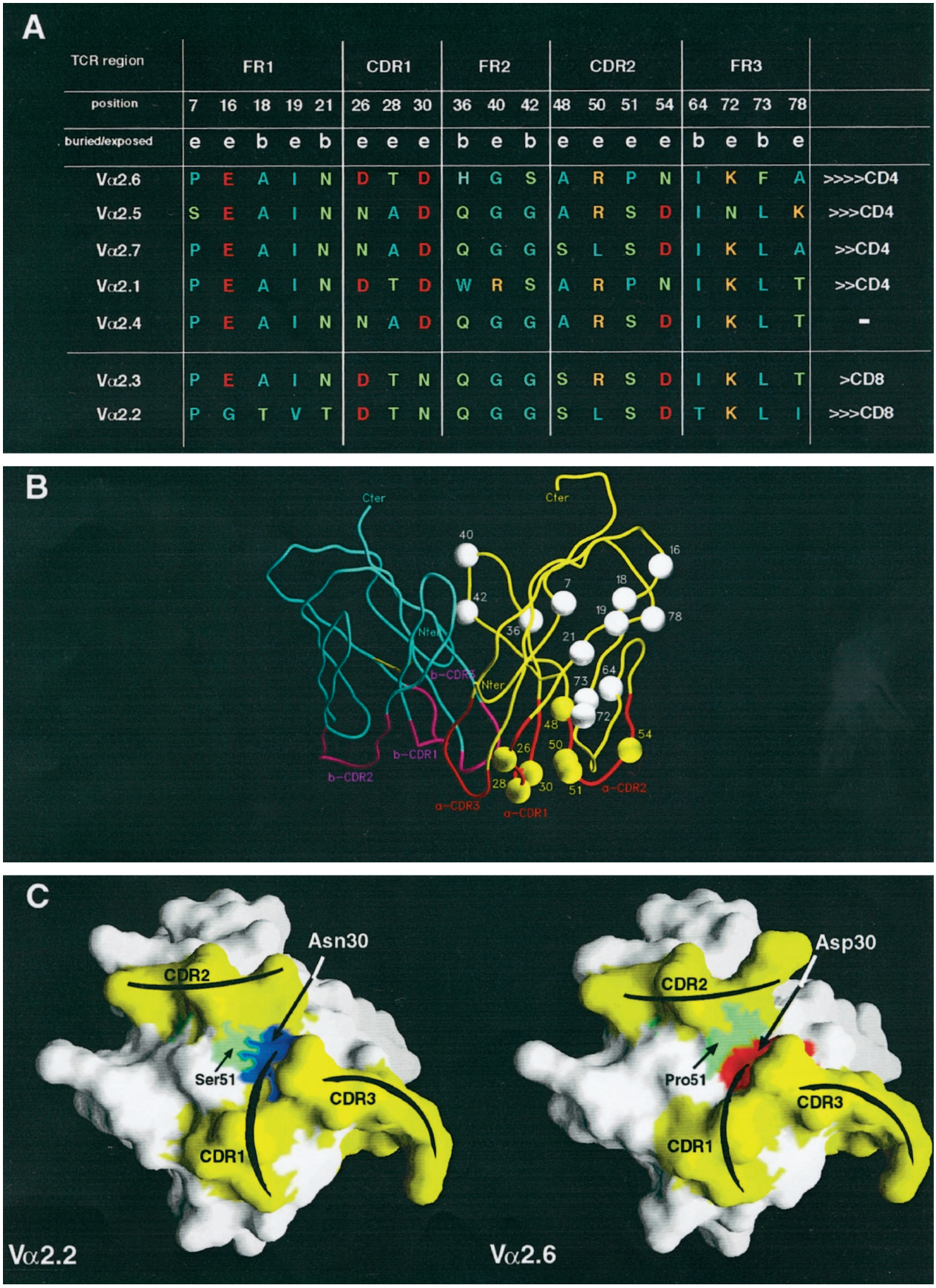
Our results indicate that the various Vα2 family members are differentially employed by $CD4^+$ and $CD8^+$ T cells as a result of thymocyte selection, but that this tendency can be influenced by sequences in the CDR3 regions of both the TCR α- and TCR β-chains. Useful clues to the molecular underpinnings of these biases are provided by the crystallographic structure of the Vα2.3 region of the KB5-C20 TCR (10) (it was unfortunately not possible to use directly the two other structures of Vα2 TCRs that have been reported recently (24, 25), as they correspond to allelic variants not present in the $TCR\alpha^b$ haplotype). We have performed multistep homology modeling of other Vα2 family members on the basis of this structure using the Modeller package (homology modeling by distance restraint algorithm, refined by probability density function of individual features) (16). Apparently, none of the variable positions has a marked impact on the disposition of the α-carbon backbone, in keeping with the comparison of Hare et al. (24) (Fig. 6*A* indicates the identities and positions of these variable amino acids, depicted on the structure in Fig. 6*B*). At most positions, the particular amino acids are not correlated with T cell subset skewing (Fig. 6*A*). The residues at positions 16 and 19 in FR1 are, however, and are solvent exposed, but Vα2.2 is the only family member with variant amino acids at these positions. On the other hand, position 30 within CDR1 does show a good correlation between the nature of the amino acid, aspartic acid or asparagine, and preferential representation in the $CD4^+$ or $CD8^+$ populations. Its importance in determining the subset skewing is further substantiated by its position in the structure: exposed on the surface of the CDR1 loop, on the face of the TCR predicted to contact the peptide/MHC ligand (Fig. 6*C*). The side-chain of N-30 does not interact with other amino acids of TCRα in the KB5-C20 structure, indicating that its influence should be direct, via differential interactions with peptide/MHC. The CDR1 region of the Vα2 TCR D10, restricted by the $A^k$ class II molecule, also has Asp at α30, contributing to the formation of a negatively charged pocket (24). A correlate could also be made with the A6 TCR/Tax/HLA-A2 complex (5), in which α30 (also Asn) engages in several contacts with the α1 helix of the MHC class I molecule as well as with the peptide. Interestingly, in the B7 TCR structure, complexed with the same Tax/HLA-A2 ligand, α30 is Asp, as in Vα2.6 and consorts, and no longer contacts the MHC molecule, but only with the peptide (6). Thus, one might speculate that an Asn at α30 confers a generic propensity for interaction with MHC class I molecules, while charged residues are preferred in interactions with class II. Yet the influence of α30 is probably modulated by residues at other positions; for example, Vα2.6 shows a stronger bias for the $CD4^+$ compartment than Vα2.1, which has the same CDR1 composition. No clear explanation for such differences emerges from the alignment shown in Fig. 6*A*; in particular, no recognizable sequence patterns in the CDR2 region are evident.

This correlative analysis suggests that variability in contacts with the peptide/MHC ligand dictates preferential selection of some Vα2 family members into the $CD4^+$ or $CD8^+$ population. We have seen no indication of an influence of positions proposed to interact with the CD8 coreceptor (26); K56 is invariant in the Vα2 family, and the pattern of variability of amino acids in its immediate vicinity (α54, α64) does not correlate with the T cell subset bias.

Other elements of the TCR V regions do seem to modulate MHC class preference. First, the CDR3 region of the TCR α-chain must be influential, since the Vα family member repartition is not absolute in our system; particular CDR3α sequences allow selection into the less-favored compartment. Interestingly, as illustrated in Fig. 6*C*, exposed CDR3α amino acids are located very close to α30 in the Vα CDR1, which appears to be in the alignment of the CDR3α loop. Very small changes in the CDR3 region seem to be capable of a profound impact. Indeed, we have found (in TCRs analyzed in another project) that single amino acid replacements in the CDR3α region of otherwise identical TCRs suffice to switch the restriction of the TCR from class I to class II molecules (M. Correia-Neves, unpublished observations). A second modulating element is the TCR β-chain, in particular the CDR3β region. The bias observed in $V\beta5.2^+$ TCR tg mice was not seen when we analyzed broader populations expressing diverse Vβ5 chains in nontransgenic mice. This is consistent with the fact that quite a few of the interactions between TCR and MHC molecules that have been previously identified involve CDR3 residues on both TCR chains (4, 5).

*Generality?*

There seem to be a few common rules concerning class I/class II molecule discrimination by the TCR. First, the discriminating Vα positions one can identify are very dependent on the broader molecular context; the influence of the α30 residue detected in a context of limited TCR variability disappeared when the TCR β-chain was no longer monomorphic. Second, the highlighted positions vary between studies; the analyses of Sim et al. clearly established the roles of α27 and α51 for Vα3 family members' preferential interactions with MHC class I or II molecules (8, 27). The influence of α51 of Vα3 was also manifest in the results of Andersen et al. (28); a serine at this position was required for stabilization of a superantigen/TCR/MHC class II complex. An influence of α27 would not have been detected in our study because it is not polymorphic in the Vα2 family; on the other hand, an influence of α51,




| TCR region | FR1 | | | | | CDR1 | | | FR2 | | | CDR2 | | | | FR3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| position | 7 | 16 | 18 | 19 | 21 | 26 | 28 | 30 | 36 | 40 | 42 | 48 | 50 | 51 | 54 | 64 | 72 | 73 | 78 | |
| buried/exposed | e | e | b | e | b | e | e | e | b | e | b | e | e | e | e | b | e | b | e | |
| Vα2.6 | P | E | A | I | N | D | T | D | H | G | S | A | R | P | N | I | K | F | A | >>>>CD4 |
| Vα2.5 | S | E | A | I | N | N | A | D | Q | G | G | A | R | S | D | I | N | L | K | >>>CD4 |
| Vα2.7 | P | E | A | I | N | N | A | D | Q | G | G | S | L | S | D | I | K | L | A | >>CD4 |
| Vα2.1 | P | E | A | I | N | D | T | D | W | R | S | A | R | P | N | I | K | L | T | >>CD4 |
| Vα2.4 | P | E | A | I | N | N | A | D | Q | G | G | A | R | S | D | I | K | L | T | - |
| Vα2.3 | P | E | A | I | N | D | T | N | Q | G | G | S | R | S | D | I | K | L | T | >CD8 |
| Vα2.2 | P | G | T | V | T | D | T | N | Q | G | G | S | L | S | D | T | K | L | I | >>>CD8 |

**FIGURE 6.** *A*, Identities of variable amino acids in each of the Vα2 family members, numbered as previously described (9). *B*, Model of the Vβ5-Vα2 TCR, deduced by β-chain replacement from the structure described by Housset et al. (10); variable positions localized in FRs are represented in white, and those in the CDRs are shown in yellow. *C*, Solvent exclusion representation of the Vα2.2 and Vα2.6 surfaces facing the peptide/MHC complex. Vα position 30 is represented in magenta and red for Vα2.2 and Vα2.6, respectively. The CDRs are represented in yellow, and position 51, identified as important in class I/class II discrimination in the Vα3 family (8), is shown in dark green.

which is variable in V$\alpha$2 family members, could have been observed. Yet there appeared to be no such influence, as the amino acid distribution at position 51 did not correlate with MHC class preference. In fact, the proline residue at $\alpha$51 that favored interaction with class I molecules in V$\alpha$3 (8, 27) is precisely the amino acid found in V$\alpha$2.6 that had the strongest association with class II molecule preference. Thus, $\alpha$51P cannot be a generally applicable determinant of class I/class II discrimination, as proposed (29). However, it is interesting that position 51 seems to occupy a region next to residue 30 in the tertiary structure (Fig. 6*C*). Third, it is becoming clear from comparative analyses of the TCR/peptide/MHC crystal structures that TCR:MHC molecule contacts are not the same in different complexes. In the studies of Wiley and colleagues (6), comparison of the A6 and B7 TCRs, both of which recognize a Tax/HLA-A2 ligand, revealed that many of the TCR residues that contact the MHC molecule in one structure also contact it in the other, but that the nature and direction of these contacts can be different; overall, only 1 in 17 contacts was shared between the two structures. There cannot, in this context, be strongly dominant rules that guide class I or class II preference, and it may ultimately be impossible to draw general rules from primary sequence comparisons.

Together the data indicate that MHC restriction by class I vs class II molecules does not depend solely on the recognition of particular MHC residues by a few specific TCR$\alpha$ CDR1/2 amino acids. Rather, MHC class preference results from a combination of inputs, from particular residues in CDR1 and CDR2, but also from the randomly generated TCR$\alpha$ and TCR$\beta$ CDR3 regions.

## Acknowledgments

We thank F. Carbone for useful discussions, G. Maza and B. Malissen for the structural data and useful discussion, P. Gerber and C. Ebel for assistance, M. Gendron and F. Fischer for following the mice, and S. Vicaire and D. Sommer-Stephan for the automatic sequencing.

## References

1. Zinkernagel, R. M., and P. C. Doherty. 1997. The discovery of MHC restriction. *Immunol. Today 18:14.*
2. Sim, B. C., D. Lo, and N. R. Gascoigne. 1998. Preferential expression of TCR V$\alpha$ regions in CD4/CD8 subsets: class discrimination or co-receptor recognition? *Immunol. Today 19:276.*
3. Sim, B. C., N. Aftahi, C. Reilly, B. Bogen, R. H. Schwartz, N. R. Gascoigne, and D. Lo. 1998. Thymic skewing of the CD4/CD8 ratio maps with the T-cell receptor $\alpha$-chain locus. *Curr. Biol. 8:701.*
4. Garcia, K. C., M. Degano, R. L. Stanfiled, A. Brunmark, M. R. Jackson, P. A. Peterson, T. Teyton, and I. A. Wilson. 1996. An $\alpha\beta$ T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. *Science 274:209.*
5. Garboczi, D. N., P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. 1996. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature 384:134.*
6. Ding, Y. H., K. J. Smith, D. N. Garboczi, U. Utz, W. E. Biddison, and D. C. Wiley. 1998. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity 8:403.*
7. Teng, M.-K., A. Smolyar, A. G. D. Tse, J.-H. Liu, J. Liu, R. E. Hussey, S. G. Natherson, H.-C. Chang, E. L. Reinherz, and J.-H. Wang. 1998. Identification of a common docking topology with substantial variation among different TCR-peptide-MHC complexes. *Curr. Biol. 8:409.*
8. Sim, B. C., L. Zerva, M. I. Greene, and N. R. J. Gascoigne. 1996. Control of MHC restriction by TCR V$\alpha$ CDR1 and CDR2. *Science 273:963.*
9. Gahery-Segard, H., E. Jouvin-Marche, A. Six, C. Gris-Liebe, M. Malissen, B. Malissen, P. A. Cazenave, and P. N. Marche. 1996. Germline genomic structure of the B10.A mouse Tcra-V2 gene subfamily. *Immunogenetics 44:298.*
10. Housset, D., G. Mazza, C. Gregoire, C. Piras, B. Malissen, and J. C. Fontecilla-Camps. 1997. The three-dimensional structure of a T-cell antigen receptor V$\alpha$V$\beta$ heterodimer reveals a novel arrangement of the V$\beta$ domain. *EMBO J. 16:4205.*
11. Pircher, H., N. Rebai, M. Groettrup, C. Grégoire, D. E. Speiser, M. P. Happ, E. Palmer, R. M. Zinkernagel, H. Hengartner, and B. Malissen. 1992. Preferential positive selection of V$\alpha$2$^+$CD8$^+$ T cells in mouse strains expressing both H-2$^k$ and T cell receptor V$\alpha^a$ haplotypes: determination with a V$\alpha$2-specific monoclonal antibody. *Eur. J. Immunol. 22:399.*
12. Kelly, J. M., S. J. Sterry, S. Cose, S. J. Turner, J. Fecondo, S. Rodda, P. J. Fink, and F. R. Carbone. 1993. Identification of conserved T cell receptor CDR3 residues contacting known exposed peptide side chains from a major histocompatibility complex class I-bound determinant. *Eur. J. Immunol. 23:3318.*
13. Kouskoff, V., K. Signorelli, C. Benoist, and D. Mathis. 1995. Cassette vectors directing expression of T cell receptor genes in transgenic mice. *J. Immunol. Methods 180:273.*
14. Philpott, K. L., J. L. Viney, G. Kay, S. Rastan, E. M. Gardiner, S. Chae, A. C. Hayday, and M. J. Owen. 1992. Lymphoid development in mice congenitally lacking T cell receptor $\alpha\beta$-expressing cells. *Science 256:1448.*
15. Chang, H. L., M. H. Zaroukian, and W. J. Esselman. 1989. T200 alternate exon use in murine lymphoid cells determined by reverse transcription-polymerase chain reaction. *J. Immunol. 143:315.*
16. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol. 234:779.*
17. Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins 11:281.*
18. Jorgensen, J. L., U. Esser, B. Fazekas de St. Groth, P. A. Reay, and M. M. Davis. 1992. Mapping T-cell receptor-peptide contacts by variant peptide immunization of single-chain transgenics. *Nature 355:224.*
19. Sant'Angelo, D. B., B. Lucas, P. G. Waterbury, B. Cohen, T. Brabb, J. Goverman, R. N. Germain, and C. A. Janeway, Jr. 1998. A molecular map of T cell development. *Immunity 9:179.*
20. Padovan, E., G. Casorati, P. Dellabona, S. Meyer, M. Brockhaus, and A. Lanzavecchia. 1993. Expression of two T cell receptor $\alpha$ chain: dual receptor T cells. *Science 262:422.*
21. Gascoigne, N. R. J. 1995. Genomic organization of T cell receptor genes in the mouse. In *T Cell Receptors*. J. I. Bell, M. J. Owen, and E. Simpson, eds. Oxford University Press, New York, p. 288.
22. Arden, B., S. P. Clark, D. Kabelitz, and T. W. Mak. 1995. Mouse T-cell receptor variable gene segment families. *Immunogen 42:501.*
23. Chou, H. S., S. J. Anderson, M. C. Louie, S. A. Godambe, M. R. Pozzi, M. A. Behlke, K. Huppi, and D. Y. Loh. 1987. Tandem linkage and unusual RNA splicing of the T-cell receptor $\beta$-chain variable-region genes. *Proc. Natl. Acad. Sci. USA 84:1992.*
24. Hare, B. J., D. F. Wyss, M. S. Osburne, P. S. Kern, E. L. Reinherz, and G. Wagner. 1999. Structure, specificity and CDR mobility of a class II restricted single-chain T-cell receptor. *Nat. Struct. Biol. 6:574.*
25. Plaksin, D., S. Chacko, J. Navaza, D. H. Margulies, and E. A. Padlan. 1999. The x-ray crystal structure of a V$\alpha$2.6.J$\alpha$38 mouse T cell receptor domain at 2.5 A resolution: alternate modes of dimerization and crystal packing. *J. Mol. Biol. 289:1153.*
26. Arden, B. 1998. Conserved motifs in T-cell receptor CDR1 and CDR2: implications for ligand and CD8 co-receptor binding. *Curr. Opin. Immunol. 10:74.*
27. Sim, B. C., J. L. Wung, and N. R. Gascoigne. 1998. Polymorphism within a TCRAV family influences the repertoire through class I/II restriction. *J. Immunol. 160:1204.*
28. Andersen, P. S., P. M. Lavoie, R.-P. Sekaly, H. Churchill, D. M. Franz, P. M. Schlievert, K. Karjalainen, and R. A. Mariuzza. 1999. Role of the T cell receptor $\alpha$ chain in stabilizing TCR-superantigen-MHC class II complexes. *Immunity 10:473.*
29. Sim, B. C., and R. J. Gascoigne. 1999. Reciprocal expression in CD4 or CD8 subsets of different members of the V$\alpha$11 gene family correlates with sequence polymorphism. *J. Immunol. 162:3153.*
30. Koop, B. F., R. K. Wilson, K. Wang, B. Vernooij, D. Zallwer, C. L. Kuo, D. Seto, M. Toda, and L. Hood. 1992. Organization, structure, and function of 95kb of DNA spanning the murine T-cell receptor C$\alpha$/C$\delta$ region. *Genomics 13:1209.*