

- small,  $B_{\text{comet}} < 10^{-4}$  gauss.
25. Y. R. Fernandez *et al.*, in preparation.
  26. A. H. Minter and G. Langston, *Astrophys. J. Lett.* **467**, L37 (1996).
  27. M. Neugebauer, *Rev. Geophys.* **28**, 231 (1990).
  28. T. I. Giombosi, K. Lorencz, J. R. Jokipii, *J. Geophys. Res.* **94**, 15011 (1989).
  29. K. Dennerl *et al.*, *IAU Circular* **6404**, 6413 (1996); K. Dennerl *et al.*, in preparation.
  30. The authors are indebted to M. D. Desch, Y. R. Fernandez, J. Harrington, T. A. Livengood, D. A. Mendis, T. Northrup, H. U. Schmidt, S. L. Snowden, and D. K. Yeomans for their help and many useful

discussions. This work was supported in the U.S. by NASA's Planetary Astronomy Program Grant #NAGW188 and by a cooperative research grant between NASA Goddard Spaceflight Center and the University of Maryland. The ROSAT project is supported by the German Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF/DARA) and the Max-Planck-Gesellschaft. The UK contribution to the ROSAT Project is supported by the Particle Physics and Astronomy Research Council.

28 June 1996; accepted 16 September 1996

# An $\alpha\beta$ T Cell Receptor Structure at 2.5 Å and Its Orientation in the TCR-MHC Complex

K. Christopher Garcia, Massimo Degano, Robyn L. Stanfield, Anders Brunmark, Michael R. Jackson, Per A. Peterson, Luc Teyton, Ian A. Wilson\*

The central event in the cellular immune response to invading microorganisms is the specific recognition of foreign peptides bound to major histocompatibility complex (MHC) molecules by the  $\alpha\beta$  T cell receptor (TCR). The x-ray structure of the complete extracellular fragment of a glycosylated  $\alpha\beta$  TCR was determined at 2.5 angstroms, and its orientation bound to a class I MHC-peptide (pMHC) complex was elucidated from crystals of the TCR-pMHC complex. The TCR resembles an antibody in the variable  $V\alpha$  and  $V\beta$  domains but deviates in the constant  $C\alpha$  domain and in the interdomain pairing of  $C\alpha$  with  $C\beta$ . Four of seven possible asparagine-linked glycosylation sites have ordered carbohydrate moieties, one of which lies in the  $C\alpha$ - $C\beta$  interface. The TCR combining site is relatively flat except for a deep hydrophobic cavity between the hypervariable CDR3s (complementarity-determining regions) of the  $\alpha$  and  $\beta$  chains. The 2C TCR covers the class I MHC H-2K<sup>b</sup> binding groove so that the  $V\alpha$  CDRs 1 and 2 are positioned over the amino-terminal region of the bound dEV8 peptide, the  $V\beta$  chain CDRs 1 and 2 are over the carboxyl-terminal region of the peptide, and the  $V\alpha$  and  $V\beta$  CDR3s straddle the peptide between the helices around the central position of the peptide.

**T** lymphocytes respond to a wide variety of foreign antigens that are presented as peptides in the context of major histocompatibility molecules (MHC) (1). Specific recognition of peptide-MHC (pMHC) complexes is accomplished by a membrane-bound, multicomponent, cell surface glycoprotein termed the T cell receptor (TCR). The TCR complex consists of highly diverse, clonotypic  $\alpha\beta$  or  $\gamma\delta$  heterodimers and the  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$  chains of the invariant accessory protein CD3 (2). The  $\alpha$  and  $\beta$  chains participate in the interaction with the pMHC complex, whereas the CD3 chains partici-

pate in signal transduction. The genes encoding the TCR resemble immunoglobulin (Ig) genes not only in sequence, but in their assembly by somatic rearrangement of linked variable (V), diversity (D), joining (J), and constant (C) gene segments during lymphocyte development (3, 4). The formation of functional  $\alpha$  chain polypeptide requires the in-frame rearrangement of a V-region gene segment to a J-region gene segment, whereas functional  $\beta$  chain polypeptide is formed by two successive rearrangements of V-, D-, and J-region gene elements (5). The rearranged V-J and V-D-J regions are then attached to their respective C regions to assemble the mature  $\alpha$  and  $\beta$  chain gene products. A vast number of potential protein sequences can result from these recombinations (4), as with Igs. Although not as striking as in Igs, four regions of hypervariable amino acid sequence are found on both the  $\alpha$  and  $\beta$  chains, three of which are analogous to the

antibody complementarity-determining regions (CDRs) (6), which serve as the primary contact points between antibody and antigen (7). Both CDRs 1 and 2 are encoded within the V genes; CDR3 occurs at the V-J junction in the  $\alpha$  chain and at the V-D-J junction in the  $\beta$  chain (4).

The high degree of sequence identity between various V and C elements of TCRs and Igs (30 to 70 percent) suggested that TCR domains are folded into  $\beta$ -sheet sandwich structures (8, 9), resembling Ig domains (10), that would pair in a manner similar to the heavy (H) and light (L) chains of antibodies. However, the recognition requirements of a TCR are more restricted than that of an antibody. Antibodies can bind ligands of extensive chemical and structural diversity, as reflected in the different shapes of antibody combining sites, from flat surfaces to deep grooves (7). The function of the TCR is to discriminate among different peptide antigens embedded in the largely flat, undulating surfaces of MHC molecules, whose dimensions and shape are relatively constant. Therefore, a more conserved binding site topology among different TCRs could be expected. Limited sequence diversity of the CDRs 1 and 2 suggested that most of the peptide specificity of the TCR would reside in CDR3, which is the most variable because of the junctional diversity of the V(D)J recombination (4). Site-directed mutagenesis studies confirmed that substitutions in CDR3 can either alter antigen specificity or abolish the response (11). Various models have been proposed for TCR recognition of pMHC in which the CDR3s of  $V\alpha$  and  $V\beta$  contact the peptide, whereas CDRs 1 and 2 interact primarily with the MHC  $\alpha$  helices (4, 12, 13) or the ends of the peptide (14).

The x-ray structures of an individual TCR  $\beta$  chain from a T cell clone termed 14.3.d (15) and an isolated  $V\alpha$  fragment from a T cell clone termed 1934.4 (16) have confirmed that the TCR does indeed contain Ig-like domains. The  $V\alpha$  and  $V\beta$  domains resemble a v-type Ig fold (17), whereas  $C\beta$  more distantly resembles a c-type Ig fold (17). The monomeric  $\beta$  chain was proposed to be rather rigid because of extensive contacts between  $V\beta$  and  $C\beta$ , and a large protruding loop from  $C\beta$  that might limit "elbow" motion between  $V\beta$  and  $C\beta$  domains (15). Both CDRs 1 and 2 appear to be conformationally restricted by main-chain interactions with framework residues but, in the absence of the buttressing effect of their respective  $\alpha$  and  $\beta$  chains, the CDR3s fold away from the domain surface toward the solvent.

The absence of a TCR  $\alpha\beta$  heterodimer crystal structure until now has been due to difficulties in producing large quantities of

K. C. Garcia, M. Degano, R. L. Stanfield, and I. A. Wilson are at the Department of Molecular Biology and the Skaggs Institute of Chemical Biology, Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA. A. Brunmark, M. R. Jackson, P. A. Peterson, and L. Teyton are at the R. W. Johnson Pharmaceutical Research Institute-La Jolla, 3535 General Atomic Court, San Diego, CA 92121, USA.

\*To whom all correspondence should be addressed. E-mail: wilson@scripps.edu

soluble, homogeneous  $\alpha$  and  $\beta$  chains that are correctly paired and folded. Recombinant techniques, although effective for the expression of antibodies, have not worked as well for TCRs. Prokaryotic and eukaryotic systems have been used to express TCRs (18), TCR-variable domain constructs (TCR-Fv) (18), and chimeras with Ig C domains (19), membrane-bound lipid linkages (20), the CD3  $\zeta$  chain (21), and leucine zippers (22), but with low yields. Even for correctly paired  $\alpha\beta$  TCRs, sufficient purity for obtaining diffraction quality crystals has not been achieved, even with deletion of asparagine-linked (N-linked) glycosylation sites (23). The conclusion is that chain pairing and secretion is less efficient (24), and refolding from *Escherichia coli* does not reliably yield functional TCRs.

TCR 2C ( $V_{\alpha}3J_{\alpha}58C_{\alpha}$ ;  $V_{\beta}8.2D_{\beta}2J_{\beta}2.4C_{\beta}2$ ), which was one of the first TCRs cloned (3, 25), is a murine receptor of a cytotoxic T cell with specificity for MHC class I H-2K<sup>b</sup> molecules (syngeneic), but cross-reactive (allogeneic) with H-2L<sup>d</sup> in

association with a self-peptide (p2Ca) derived from a mitochondrial protein (25). 2C has a relatively high affinity for the H-2L<sup>d</sup>-p2Ca complex ( $K_d \sim 0.5 \times 10^{-6}$ M) (26), compared to H-2K<sup>b</sup>-dEV8 ( $K_d \sim 0.5 \times 10^{-5}$ M) (26).

Structural analysis of the TCR 2C has been made possible by crystallization of a fully glycosylated form of 2C expressed in *Drosophila melanogaster* cells. The three-dimensional structure determination at 2.5 Å provides insights into the architecture, assembly, and folding of a TCR. In addition, crystallization of a complex of 2C and H-2K<sup>b</sup>-dEV8 has allowed us to determine by molecular replacement the orientations of both the TCR and the pMHC in the TCR-pMHC complex.

**TCR expression, crystallization, and structure determination.** Expression of 2C from *D. melanogaster* cells (27) yielded a heterogeneous mixture of disulfide-bonded heterodimers and free monomers that, after purification, could be crystallized in bundles of needles that diffracted to low resolution.

The COOH-terminal 10 to 15 residues were then digested away by successive treatment with carboxypeptidases A and B so that 2C could be purified on a hydrophobic interaction chromatography (HIC) column (28). The correctly paired disulfide-linked heterodimers (28) could now be crystallized and macro-seeded to a size sufficient for x-ray data collection. Interestingly, 2C expressed without N-linked glycosylation, by including the glycosylation inhibitor tunicamycin in the growth media, is soluble and binds to pMHC and all conformation-specific mAbs available (28), but it did not produce crystals (29).

The crystal structure of 2C was determined by means of a combination of molecular replacement (MR) and multiple isomorphous replacement (MIR). An initial room temperature data set to 2.9 Å was collected from a monoclinic crystal (30), space group  $P2_1$  (Table 1). Solutions (MR) were obtained for the  $\beta$  chain and  $V_{\alpha}$  (Table 1) but could not be determined for  $C_{\alpha}$  even with a wide range of Ig-like do-

**Table 1.** Data collection, heavy atom, and refinement statistics. Monoclinic 2C crystals were grown in sitting drops from 2C (15 mg/ml), 1.8 M ammonium sulfate, 50 mM sodium acetate, and 50 mM sodium citrate (pH 6.4) at 22°C. The crystals were  $P2_1$ , with unit cell parameters  $a = 136.9$  Å,  $b = 76.9$  Å,  $c = 57.0$  Å, and  $\beta = 101.1^\circ$ , with two  $\alpha\beta$  TCR heterodimers in the asymmetric unit [ $V_m = 2.95$  Å<sup>3</sup> per dalton (31); solvent content = 58 percent]. The crystals have pronounced pseudo-face centering, with  $h + k = 2n + 1$  reflections being systematically weak but not absent. The frozen 2C crystal had slightly altered unit cell parameters  $a = 138.0$  Å,  $b = 72.9$  Å,  $c = 57.3$  Å, and  $\beta = 101.2^\circ$  that resulted in an exact C2 centering through a slight adjustment of one molecule (30). X-ray data were collected on a MAR Imaging Plate mounted on a Siemens rotating anode operating at 40 kV and 50 mA with Supper double-focusing long mirrors. Data were integrated, scaled, and reduced with DENZO and SCALEPACK (31). An MR solution was obtained with the free  $\beta$  chain (1 bec) as a search model (15) and the program AMoRe (31). The  $V_{\alpha}$  domain position was determined in X-PLOR with an REI light chain monomer (68) as a search model, after positioning the  $\beta$  chain. A set of 1156 reflections (5 percent) was excluded from refinement for cross-validation ( $R_{free}$ ) (31). Isomorphous heavy atom derivatives were obtained by soaking 2C crystals in 20  $\mu$ M ethyl mercury thiosalicylate (EMTS) or 500  $\mu$ M trimethyl lead acetate (TMLA) in 2.2 M ammonium sulfate, 50 mM sodium acetate (pH 7.0). Heavy atom sites were identified and refined with PHASES (31) and PHASIT (31). Phases to 3.2 Å were improved by solvent flattening and molecular averaging with DM (31) and by combining MIR and partial model phases by means of SIGMAA (31). Noncrystallographic symmetry restraints were imposed in simulated annealing refinement in X-PLOR (31). After one cycle of refinement and manual model building with O (31), SIGMAA-weighted  $F_o - F_c$  difference maps showed density for four N-linked glycans. After several cycles, the  $R_{cryst}$  and  $R_{free}$  were 0.20 and 0.37 for data with  $F > 2\sigma_f$ . The model was then placed in the new C2 unit cell with the MR procedure in AMoRe (31) and refined with an  $R_{free}$  data set of 1536 reflections (5.4 percent). The high quality of this electron density map allowed unambiguous tracing of  $C_{\alpha}$ . The  $R_{cryst}$  and  $R_{free}$  values are 0.208 and 0.285, respectively ( $F > 2\sigma_f$ ) with individual B values. The 2C model contains 437 amino acids [residues  $\alpha(1-213)$  and  $\beta(3-247)$ ] (34), ten carbohydrate moieties, 30 ordered solvent molecules, and three sulfate ions. Of the residues in the 2C model, 85 percent fall in the most favorable regions of a Ramachandran plot [generated from PROCHECK (31)] with no residues in a disallowed region. Analysis by 3D-PROFILE (31) and ERRAT (31) shows the 2C model is consistent with highly refined protein structures.

Data set	Resolution limit (Å)	$R_{sym}^*$ (last shell)	Completeness (last shell) (%)	$\langle I/\sigma(I) \rangle$ (last shell)	Sites (n)	$R_{iso}^\dagger$	$R_{Cullis}^\ddagger$	$R_{Kraut}^\S$	Phasing power $  $
Native-C2	2.5 (2.6-2.5)	4.8 (23.7)	99.7 (99.3)	18.3 (2.9)					
Native- $P2_1$	2.9 (3.0-2.9)	11.5 (54.0)	88.3 (80.2)	10.5 (1.1)					
EMTS- $P2_1$	3.2 (3.5-3.2)	9.8 (37.0)	73.0 (71.7)	10.0 (2.0)	2	8.2	57.6	11.6	1.54
TMLA- $P2_1$	3.2 (3.5-3.2)	11.1 (45.2)	80.0 (73.0)	9.8 (1.3)	2	9.8	56.1	13.1	1.73

Mean figure of merit = 0.340 (low because the EMTS and TMLA sites are almost identical)

#### Refinement statistics

Resolution (Å)	R value	$R_{free}$	rms from ideal values			Average B values (Å <sup>2</sup> )			
			Bond lengths	Bond angles	Improper	$V_{\alpha}$	$C_{\alpha}$	$V_{\beta}$	$C_{\beta}$
$\infty-2.5$ ( $F > 2\sigma$ )	0.224 (0.208)	0.304 (0.285)	0.011 Å	1.95°	1.45°	36.9	50.3	29.8	27.1

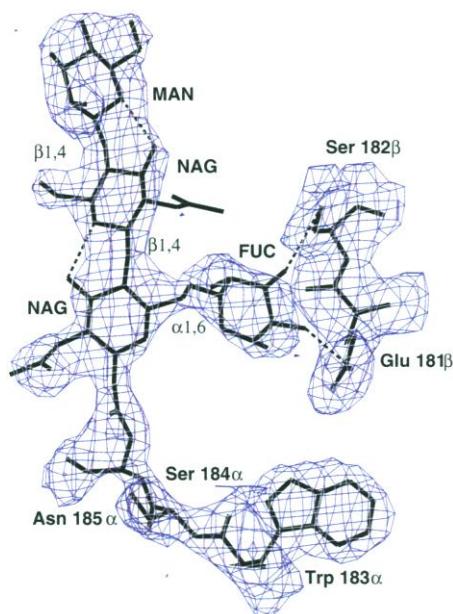
$R_{sym} = 100 \cdot \sum (|I_i(hkl) - \langle I(hkl) \rangle|) / \sum I_i(hkl)$ , where  $I_i(hkl)$  are the intensities of multiple measurements and  $\langle I(hkl) \rangle$  is the average of the measured intensities of the  $i$ th reflection, with the summation extended over all measured reflections;  $R_{iso}^\dagger = 100 \cdot \sum (|F_{PH} - F_P|) / \sum F_P$ , where  $F_{PH}$  is the derivative structure factor and  $F_P$  is the native structure factor and the sum is extended over all reflections common between the two data sets;  $R_{Cullis}^\ddagger = 100 \cdot \sum (|F_{PH}(hkl)| \pm \sum |F_P(hkl)|) / \sum (|F_{PH}(hkl)| \pm |F_P(hkl)|)$ , where the sum is extended over all centric reflections;  $R_{Kraut}^\S = 100 \cdot \sum (|F_{PH}(hkl)| - |F_H(hkl)|) / \sum |F_P(hkl)|$ , where the sum is extended over all acentric reflections;  $||$  Phasing power =  $\langle F_r \rangle / e$ , where  $e$  is the lack-of-closure error.

main. The  $V\alpha$ ,  $V\beta$ , and  $C\beta$  domains were refined by X-PLOR simulated annealing (31) to an  $R$  value of 0.25. The  $C\alpha$  region showed some continuous density only for three  $\beta$  strands of the back sheet of the classical Ig fold (17), but there was no interpretable density for a top sheet. Although  $C\alpha$  had very few crystal contacts and there was precedent for constant domain disorder in some Igs (32), extensive disorder in a single Ig domain was puzzling. An MIR structure determination was undertaken to eliminate the possibility of MR model bias (Table 1); MIR and partial structure factor phases from the refined coordinates of the other three domains were combined and used to retrace and refine the entire molecule. Four of the  $\beta$  strands of  $C\alpha$  could be traced in an unambiguous register, but the outer strands remained fragmented. Successive rounds of model building, refinement, and phase combination resulted in an  $R$  value of 0.21, but with an  $R_{\text{free}}$  (31) of 0.37 at 2.9 Å. In addition, the presence of the pseudo-centering caused half of all the observed reflections to be vanishingly weak. Finally, a crystal of 2C was successfully cryocooled to 100 K, and a complete data set with very high redundancy was collected to 2.5 Å. A small change in the crystal lattice on freezing (33) resulted in exact crystallographic centering and a C2 space

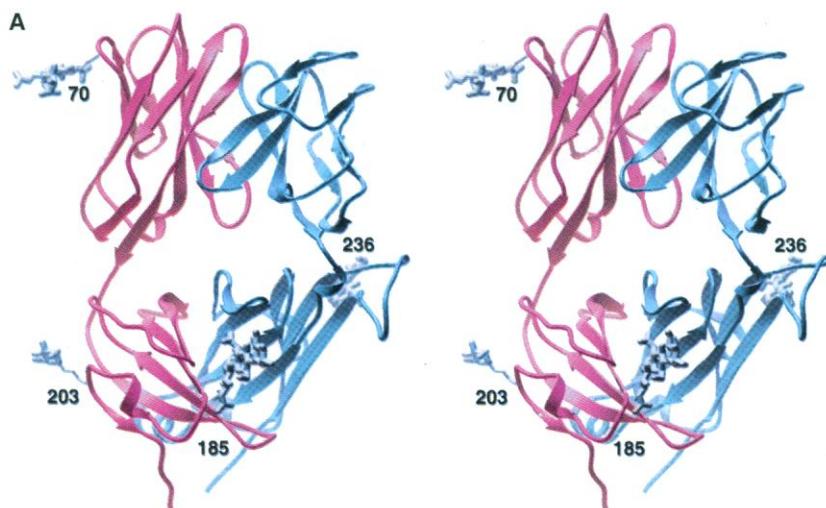
group (30). Refinement improved the electron density map (Fig. 1), especially in the  $C\alpha$  domain, where iterative cycles of model building and refinement allowed a complete trace, up to and including the COOH-terminal disulfide-linked cysteine residues ( $\text{Cys}^{213\alpha}$ - $\text{Cys}^{247\beta}$ ) (34), with good geometry and a substantial drop in  $R_{\text{free}}$  to 0.285 (Table 1). The carbohydrate density also improved, and at present ten sugar moieties have been located (Figs. 1 and 2A; Table 1).

**Overall structure.** The 2C  $\alpha$  and  $\beta$  chains [ $\alpha(1-213)$ ,  $\beta(3-247)$ ] fold into a quaternary structure that resembles the antigen binding

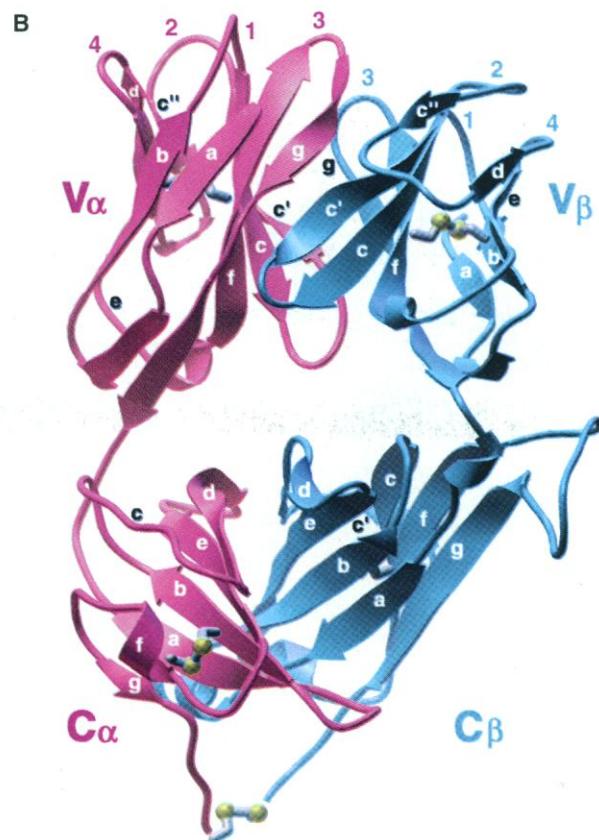
region (Fab) of antibodies (Fig. 2). Each chain contains two domains, three of which ( $V\alpha$ ,  $V\beta$ , and  $C\beta$ ) are related to Ig domains, and one of which ( $C\alpha$ ) substantially deviates from the standard Ig fold (17). The overall dimensions of 2C are 61 by 56 by 33 Å, compared to about 65 by 46 by 30 Å for a representative Fab (35), whose elbow angle and CDR lengths are similar to the 2C TCR. The 2C appears slightly compressed in its length and wider across its middle elbow region than does an Fab. A classical measurement made in Fabs is the angle between the vectors that represent the pseudo-dyad axes relating  $V_L$  to  $V_H$



**Fig. 1.** Electron density in the 2C  $C\alpha$ - $C\beta$  interface around  $\text{Asn}^{185\alpha}$ , showing an occupied N-linked glycosylation site ( $\text{Asn}^{185}$ - $\text{GlcNAc}1_{\beta 1,4}$ - $\text{GlcNAc}2_{\beta 1,4}$ - $\text{Man}$ ,  $\alpha 1,6$ - $\text{Fuc}1$ ), as well as an invariant Trp residue ( $\text{Trp}^{183\alpha}$ ). Close interaction with  $C\beta$  is shown by hydrogen bonds (dotted lines) of the fucose residue to side chains of  $C\beta$ . The electron density map was generated from SIGMAA-weighted  $2F_o - F_c$  coefficients (31) and contoured at  $1.0\sigma$  with the program TOM (69). NAG, GlcNAc.



**Fig. 2.** Three-dimensional structure of an  $\alpha\beta$  TCR. (A) Stereoview of the backbone ribbon representation of 2C, showing four N-linked glycosylation sites that have ordered carbohydrate moieties as represented by white ball-and-stick models. (B) Backbone ribbon representation of the 2C TCR. The  $\alpha$  chain is in pink (residues 1-213), and the  $\beta$  chain is in blue (residues 3-247). The  $\beta$  strands are represented as arrows and labeled according to the standard convention used for Ig folds (17). The disulfide bonds (yellow balls for sulfur atoms) are shown within each domain and for the COOH-terminal interchain disulfide (Table 1). The CDRs are numerically labeled (1 to 4) for each chain. Figures 2 through 6 and 8 through 9 were produced with AVS (69).



and  $C_L$  to  $C_{H1}$ , otherwise known as the elbow angle, which at present ranges from  $133^\circ$  to  $220^\circ$  (7). For 2C, the elbow angle is  $148.6^\circ$  (36), within the normal range for an Fab, but about  $6^\circ$  as small as the 14.3.d  $\beta$  chain structure, indicating some segmental flexibility between the V and C domains. The pseudo-dyad relating  $C_\alpha$  to  $C_\beta$  is shifted toward the  $\alpha$  chain by about  $4.6 \text{ \AA}$  relative to the  $V_\alpha$ - $V_\beta$

pseudo-dyad, giving the TCR a somewhat more asymmetric appearance.

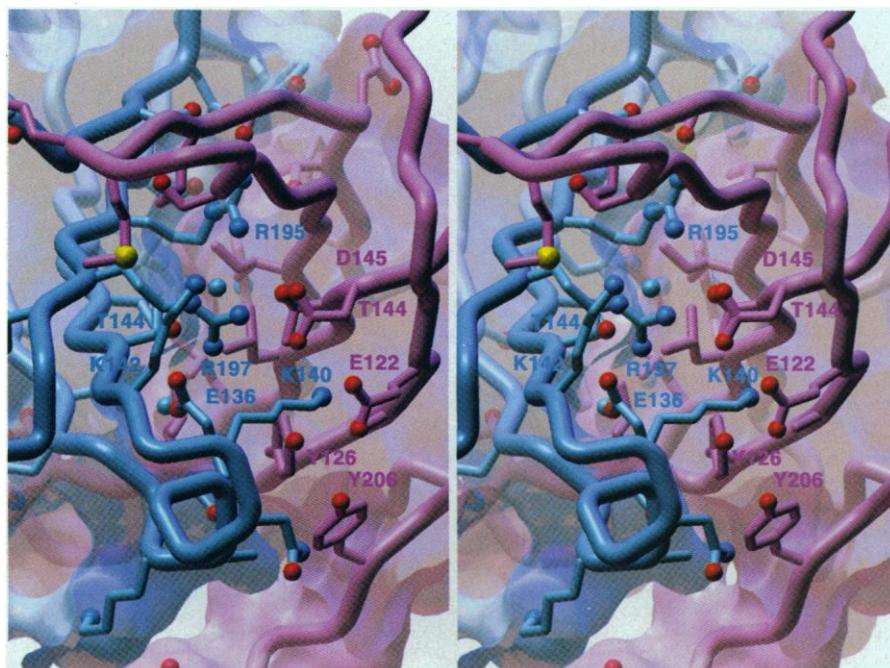
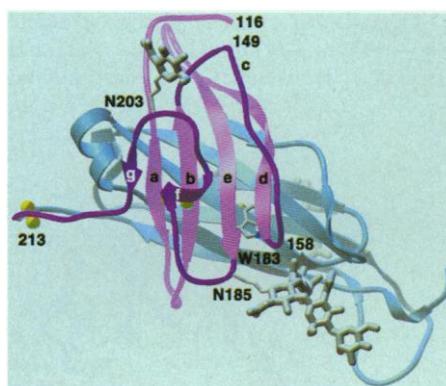
In the  $V_\alpha$  crystal structure (16), the  $V_\alpha$  forms homodimers that themselves dimerize, leading to a proposal (16) that this arrangement is spatially compatible with the dimer seen in the class II MHC crystal structure (37) and could therefore be part of a dimeric TCR-MHC signaling assembly.

The 2C TCR, however, is monomeric both in solution and in the crystals (38).

The amino acid sequences (39) of mouse and human TCR C regions suggested structural homology with Ig C domains (4) although the sequence similarity is substantially higher for  $C_\beta$  than for  $C_\alpha$ . About 30 percent sequence identity is seen for up to 100  $C_\beta$  residues in antibody C domains  $C_L$  and  $C_{H1}$ , with similar spacings between the intrachain disulfide cysteine residues, and preservation of key structural residues such as Trp<sup>161</sup> in the c strand. The  $C_\beta$  contains an unusual insertion between residues 219 $\beta$  and 232 $\beta$  (Fig. 2B) that otherwise is found only in TCR  $C_\gamma$  sequences (39, 40). The  $C_\beta 2$  structure of 2C is indeed similar (41) to the C domains of both heavy and light chains of Igs ( $2.0$  to  $2.2 \text{ \AA}$  root mean square (rms) deviation for 89 matched residues for  $C_L$  and  $C_{H1}$ ), but it is more similar to the  $C_{H3}$  domain of the human Fc fragment of an Ig ( $1.6 \text{ \AA}$  rms deviation for 89 residues). The 2C  $C_\beta 2$  gene sequence is almost identical to the 14.3.d  $C_\beta 1$ , and their domain structures overlap with small differences ( $1.6 \text{ \AA}$  rms deviation for all residues) except in two major areas. Residues 217 $\beta$  to 236 $\beta$ , which comprise the large, solvent-exposed insertion in  $C_\beta$ , has a different loop conformation from that of 14.3.d ( $2.0 \text{ \AA}$  rms deviation for 20  $\alpha$  carbons), indicating some conformational flexibility in this region (42). The other difference occurs for the a-b loop (residues 133 $\beta$  to 144 $\beta$ ) that folds into a short  $\alpha$  helix in 2C in the interface with  $C_\alpha$ . In Ig  $C_L$ - $C_{H1}$  and  $C_{H3}$ - $C_{H3}$  pairings, the analogous residues also form a short  $\alpha$  helix that is close to the a, b, e, d interface  $\beta$  sheet. In 14.3.d, these residues drift away from the  $\beta$  sheet in the absence of an opposing  $C_\alpha$  chain.

Four of the seven possible N-linked glycosylation consensus sequences (43) have ordered carbohydrates (Figs. 1, 2A, and 3). In  $V_\alpha$ , three carbohydrate moieties from Asn<sup>70 $\alpha$</sup>  (GlcNAc1  $\beta 1,4$  GlcNAc2,  $\alpha 1,6$  Fuc1) (43) extend out into a large solvent channel in the crystal, with the amide nitrogen of the first N-acetylglucosamine moiety (GlcNAc1) H-bonded back to the side-chain carbonyl of Asn<sup>70 $\alpha$</sup>  (Fig. 2A). Asn<sup>236 $\beta$</sup>  has a visible disaccharide (GlcNAc1  $\beta 1,4$  GlcNAc2) in which the O6 of GlcNAc1 is in a water-mediated H bond to Arg<sup>211 $\beta$</sup>  of  $C_\beta$ . In  $C_\alpha$ , density for one GlcNAc moiety extends from Asn<sup>203 $\alpha$</sup>  toward the solvent. The largest ordered carbohydrate stretch extends from Asn<sup>185 $\alpha$</sup>  (GlcNAc1  $\beta 1,4$  GlcNAc2  $\beta 1,4$  Man,  $\alpha 1,6$  Fuc1) (Figs. 1 and 3) and is involved both in crystal contacts and  $C_\alpha$ - $C_\beta$  interactions. The fucose moiety (43) is H-bonded to Glu<sup>181 $\beta$</sup>  and Ser<sup>182 $\beta$</sup>  (Fig. 1). A water molecule bridges the fucose exocyclic oxygen and

**Fig. 3.** Backbone ribbon representation of  $C_\alpha$ . The  $C_\alpha$  domain (pink) begins at residue 116 $\alpha$  and ends at the COOH-terminal residue 213 $\alpha$ , and is shown lying on top of the  $C_\beta$  domain (blue). The  $\beta$  strands of the back  $\beta$  sheet in  $C_\alpha$ , in lighter pink, are depicted as arrows, with the unusual  $C_\alpha$  "top strands" in darker pink. The strands are labeled according to the designations used for standard Ig folds (17). The back  $\beta$  sheet actually comprises strands g, a, b, e, and d with strand g H-bonding in a parallel arrangement with strand a. Strand f contains a short helical segment that is disulfide bonded (partially hidden yellow balls represent the sulfurs) to the back b strand. The other side of the back  $\beta$  sheet of  $C_\alpha$  (light pink) forms the interface with the back  $\beta$  sheet of  $C_\beta$ . The carbohydrate moieties of the two N-linked glycosylation sites (Asn<sup>185 $\alpha$</sup>  and Asn<sup>203 $\alpha$</sup> ) in  $C_\alpha$  are also shown. The Trp<sup>183 $\alpha$</sup>  in the  $C_\alpha$ - $C_\beta$  interface is shown as a white ball-and-stick representation. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.



**Fig. 4.** The  $C_\alpha$ - $C_\beta$  interface is highly polar. A stereoview of the interacting  $\beta$  sheets of  $C_\alpha$  (pink) and  $C_\beta$  (blue) is shown along with side chains making either salt bridges, H bonds, or van der Waals contacts across the interface [H bonds calculated with HBPLUS (52) and van der Waals and salt bridges calculated with CONTACTSYM (52)]. The complementarity of the  $C_\alpha$ - $C_\beta$  interface is emphasized by the superposition of a cross-section of the interacting surfaces as a translucent ghost. A predominance of basic side chains lies on the  $C_\beta$  side of the interface, and acidic side chains on the  $C_\alpha$  side, along with many other polar amino acids. Asp<sup>145 $\alpha$</sup>  makes salt bridges with Arg<sup>197 $\beta$</sup>  and Lys<sup>140 $\beta$</sup> . Glu<sup>122 $\alpha$</sup>  and Tyr<sup>126 $\alpha$</sup>  H-bond to Lys<sup>140 $\beta$</sup>  and Thr<sup>144 $\alpha$</sup>  to Arg<sup>195 $\beta$</sup> . Met<sup>170 $\alpha$</sup>  is in van der Waals contact with Lys<sup>142 $\beta$</sup> . The buried water molecules within the interface are shown. The molecular surfaces were calculated by means of AVS (69) with a  $1.4 \text{ \AA}$  probe and the PQMS program (69).

the side chain of Arg<sup>150 $\beta$</sup> . The GlcNAc2 moiety also H-bonds to a symmetry-related molecule. This N-linked sugar possibly strengthens association of C $\alpha$  with C $\beta$  and may also influence the crystal packing of 2C, which perhaps explains why non-glycosylated 2C TCR does not crystallize on its own or why it cannot be seeded with glycosylated crystals. The other carbohydrate moieties in 2C are presumably disordered or heterogeneous, or the N-linked sites are not occupied (43) although all the Asn residues are on exposed strands or  $\beta$  turns.

**The C $\alpha$  structure.** The TCR C $\alpha$  and C $\delta$  domains are the most divergent from known Ig sequences (4), showing only 12 to 18 percent sequence identities with Ig C domains. C $\alpha$  was predicted to be an Ig fold because of the presence of the two characteristic cysteine residues that normally link the top (g, f, c) and bottom (a, b, e, d)  $\beta$  sheets. Many other structurally conserved residues in Igs are missing, such as the Trp residue that normally occurs in the c strand 14 or 15 residues after the first cysteine of the intrachain disulfide. The C $\alpha$  has a relatively short span (50 amino acids) between its two cysteine residues compared to Ig C domains (60 to 65 amino acids), indicating a deletion of about 12 to 15 residues whose locations are not obvious from sequence alignments.

The C $\alpha$  structure shows that the a, b, e, and d strands comprising the back  $\beta$  sheet are highly similar to the corresponding back sheet of Ig C domains (Fig. 3). An overlap of this  $\beta$  sheet with the structurally equivalent residues of Ig C domains gives an rms deviation of  $\sim 1.3$  Å for 29 C $\alpha$  residues (41). However, large differences emerge for the c, f, and g strands, which would normally make up the top  $\beta$  sheet in a standard Ig-fold (Fig. 3). Instead, these strands are loosely packed against the bottom sheet (44) and are too far apart to form main-chain hydrogen bonds with each other. Gaps between these top elements result in exposure of some of the hydrophobic interior of the bottom sheet to solvent. The c strand connects directly from the b to the d strand and eliminates about six residues of the c' strand found in the classical c-type Ig set (17). Only 14 residues connect the c and e strands, compared to 21 in c-type folds and would place the C $\alpha$  in the h- or s-type of Ig folds (17). However, classification is complicated by the absence of an outer  $\beta$  sheet that represents a major structural deviation from all other Ig-type folds. The c strand is loosely tethered to the bottom sheet, as reflected in the higher thermal parameters of this strand ( $\sim 69$  Å<sup>2</sup>) (44), by interactions between small hydrophobic residues (Pro, Val, Ile), as compared to a standard Ig-fold in which a conserved Trp residue (Trp<sup>148L</sup>, Trp<sup>157H</sup>) anchors this

strand to the bottom sheet.

Another divergence from c-type folds comes in the loop connecting the e and f strands. Normally, this loop is large (about ten residues) and often adopts a minihelix containing a second conserved Trp residue that occupies the internal core of the loop. In C $\alpha$ , the connection consists of a three-residue turn (Fig. 3). Trp<sup>183 $\alpha$</sup>  does occur at the carboxyl end of the e strand (Figs. 1 and 3), just before the turn, but it is on the opposite side of the strand, rather than in its usual interior  $\beta$ -sheet position and is buried in a hydrophobic pocket at the C $\alpha$ -C $\beta$  interface. The e-f turn contains a conserved N-linked glycosylation site (Asn<sup>185 $\alpha$</sup> ) for which clear electron density exists for four carbohydrate moieties adjacent to the C $\alpha$ -C $\beta$  interface (Fig. 1). The deletion of about six residues between the c and d strands, and five in the turn connecting the f and g strands, accounts for the 10 to 15 fewer amino acids between the C $\alpha$  cysteines (50 residues) than for antibody C or V domains (60 to 65 residues).

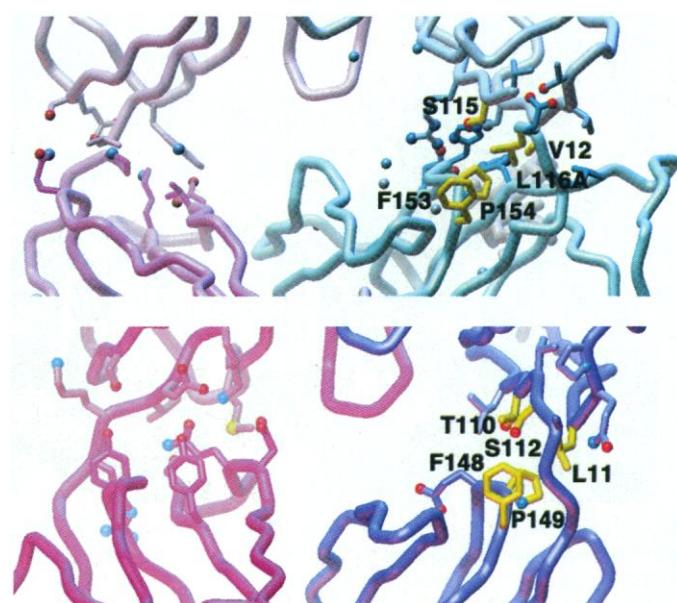
After the e-f turn, the C $\alpha$  again deviates from a canonical Ig fold (Fig. 3). The f strand is also not H-bonded to a neighboring strand, although the second cysteine (Cys<sup>191 $\alpha$</sup> ) remains disulfide linked to Cys<sup>141 $\alpha$</sup>  in an ideal disulfide geometry, as in Igs (45). Surprisingly, Cys<sup>191 $\alpha$</sup>  is in a short helical turn (residues 191 to 195) that places the side chain of Phe<sup>195 $\alpha$</sup>  into a hydrophobic pocket created by Phe<sup>143 $\alpha$</sup>  and Phe<sup>145 $\alpha$</sup>  of strand b of the back  $\beta$  sheet. This short  $\alpha$ -helical stretch (Gln<sup>192 $\alpha$</sup> -Asp<sup>193 $\alpha$</sup> -Ile<sup>194 $\alpha$</sup> -Phe<sup>195 $\alpha$</sup> ) does not appear to be chemically compatible with the normal hydrophobic-hydrophilic alternation of resi-

dues in a surface-exposed  $\beta$  strand. After the minihelix, the chain adopts a short turn from which another carbohydrate moiety extends from Asn<sup>203 $\alpha$</sup>  at the tip of the loop (Fig. 3). Three residues of the final g strand (Ala<sup>204 $\alpha$</sup>  to Tyr<sup>206 $\alpha$</sup> ) do, in fact, form a  $\beta$  strand, but, surprisingly, H-bond with the a strand of the bottom sheet, a variation not seen in Ig C domains or other Ig folds (46). Pro<sup>207 $\alpha$</sup>  then kinks the  $\beta$  strand toward the  $\beta$  chain so that its C-terminal Cys<sup>213 $\alpha$</sup>  can make the interchain disulfide with C $\beta$  Cys<sup>247 $\beta$</sup>  (28).

Studies on TCR biosynthesis and signaling allow us to consider the functional relevance of the unusual top strand topology and possible flexibility. The  $\alpha$  chain is unusually unstable and susceptible to degradation in the endoplasmic reticulum (47), that limits high-level expression in transfected eukaryotic cells (21). The instability of the  $\alpha$  chain may be partly related to a less-compact fold of the C $\alpha$  domain. Several site-directed mutagenesis studies have implicated residues in the top elements of the C $\alpha$  domain as being critical for interaction with the CD3  $\zeta$  chain (48). Some flexibility in these outer elements of C $\alpha$  may provide a mechanism for adaptability in binding CD3.

**The C $\alpha$ -C $\beta$  interface.** The C $\alpha$ -C $\beta$  association is more like that of C<sub>H</sub>3-C<sub>H</sub>3 than it is C<sub>H</sub>1-C<sub>L</sub>-like with respect to the geometry of the domain interaction, the buried surface area, and a predominance of charged interactions in the C $\alpha$ -C $\beta$  interface (Fig. 4). The total buried surface area (49) between C $\alpha$  and C $\beta$  is 2049 Å<sup>2</sup>, as compared to an average of 1700 Å<sup>2</sup> for C<sub>H</sub>1-C<sub>L</sub> (the range is 1500 to 2100 Å<sup>2</sup>). The equivalent

**Fig. 5.** The ball-and-socket joint of antibodies is conserved in the TCR. The elbow region between variable and constant domains for (top) 2C ( $\alpha$ , pink;  $\beta$ , blue) and (bottom) Fab NC41 (L chain, pink; H chain, blue) is shown for these two structures with similar elbow angles (148°) (35). The amino acid residues for the ball-and-socket joint are colored yellow for 2C residues Val<sup>12 $\beta$</sup> , Ser<sup>115 $\beta$</sup> , Leu<sup>116A $\beta$</sup> , Phe<sup>153 $\beta$</sup> , and Pro<sup>154 $\beta$</sup>  and for Fab-NC41 (35) residues Leu<sup>11H</sup>, Thr<sup>110H</sup>, Ser<sup>112H</sup>, Phe<sup>148H</sup>, and Pro<sup>149H</sup>. Other interdomain (V $\alpha$ -C $\alpha$ , V $\beta$ -C $\beta$ , V<sub>L</sub>-C<sub>L</sub>, V<sub>H</sub>-C<sub>H</sub>1) contact residues, as determined by CONTACSYM (52), are shown to highlight the comparative number of residues in the V-C interface of 2C TCR and an Fab.



$C_H3$  dimerization in Fc buries about 2200  $\text{\AA}^2$ . The crossing angle between opposing  $\beta$  sheets is  $-41^\circ$ , as opposed to around  $-53^\circ$  for most  $C_H1-C_L$  associations, and  $-45^\circ$  for  $C_H3-C_H3$  (50). The more parallel  $C\alpha-C\beta$  intersection angle gives the TCR structure its squat appearance (Fig. 2). Although the  $C\alpha$  and  $C\beta$  are related by a  $180^\circ$  dyad (36), similar to  $C_H1-C_L$ , a pairwise superposition also reveals a higher degree of similarity to

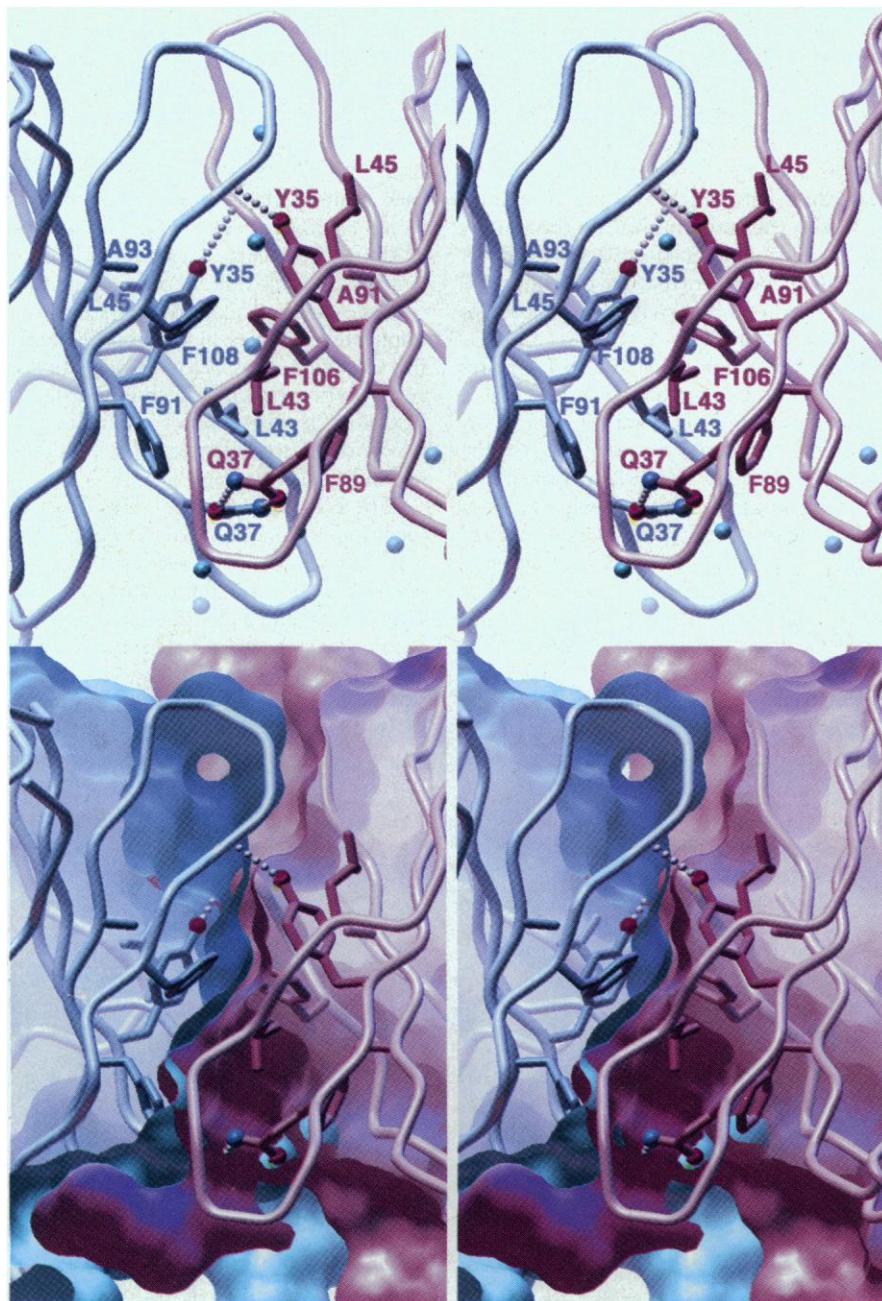
$C_H3$  domain interactions (51). When a  $C_H1$  domain is superimposed onto the  $C\beta$ , a  $15^\circ$  rotation is required to bring  $C_L$  to optimal superposition with  $C\alpha$ . However, when a domain of  $C_H3$  is superimposed onto  $C\beta$ , only a  $3^\circ$  rotation is required to overlap the other  $C_H3$  domain with  $C\alpha$ .

The  $C\alpha-C\beta$  interface is highly polar, with a skewed distribution of acidic residues on  $C\alpha$  and basic residues on  $C\beta$  (Fig. 4). A

highly conserved salt bridge, found in all  $C_H1-C_L$  (Glu<sup>123L</sup>-Lys<sup>213H</sup>) and  $C_H3-C_H3$  (Glu<sup>357H</sup>-Lys<sup>370H</sup>) interfaces, is retained in a similar location as a constellation of charged residues at one end of the  $C\alpha-C\beta$  interface (Fig. 4). This cluster of charges includes acidic  $C\alpha$  residues Asp<sup>145 $\alpha$</sup>  and Glu<sup>122 $\alpha$</sup>  that form multiple salt bridges and H-bond interactions (52) with  $C\beta$  basic residues Lys<sup>140 $\beta$</sup> , Arg<sup>195 $\beta$</sup> , and Arg<sup>197 $\beta$</sup> . In addition, Glu<sup>136 $\beta$</sup>  and two buried water molecules provide polar contacts along with a number of other side-chain H bonds. The other end of the interface is largely hydrophobic involving Trp<sup>183 $\alpha$</sup> , Phe<sup>161 $\alpha$</sup> , Tyr<sup>179 $\beta$</sup> , and Leu<sup>148 $\beta$</sup> . Interaction of  $C\beta$  with a carbohydrate extending from  $C\alpha$  Asn<sup>185 $\alpha$</sup>  is discussed above.

**Elbow region.** An extensive interaction occurs between the V and C domains (Fig. 5), with about 700  $\text{\AA}^2$  of buried surface between  $V\beta$  and  $C\beta$ , and 650  $\text{\AA}^2$  between  $V\alpha$  and  $C\alpha$  (49). The  $V\beta-C\beta$  interaction area is highly polar, with a conserved salt bridge (Glu<sup>158 $\beta$</sup> -Arg<sup>113 $\beta$</sup> ) and eight interdomain H bonds, most of which are between conserved  $\beta$  chain residues. The  $V\alpha-C\alpha$  interface, on the other hand, is composed primarily of van der Waals interactions, with only one H bond between the Ser<sup>148 $\alpha$</sup>  hydroxyl and the main-chain carbonyl oxygen of Pro<sup>116 $\alpha$</sup> . The  $V\beta-C\beta$  interface is substantially larger than a corresponding  $V_H-C_H1$  interface (390  $\text{\AA}^2$ ) in an Fab with a similar elbow angle (35). The  $V\alpha-C\alpha$  interface is also larger but differs less from corresponding  $V_L-C_L$  interfaces ( $\sim 500 \text{\AA}^2$ ). The conserved "ball-and-socket joint" in the  $V_H-C_H1$  interface of antibodies is present (Fig. 5) in the 2C  $\beta$  chain (Val<sup>12 $\beta$</sup> , Ser<sup>115 $\beta$</sup> , Leu<sup>116A $\beta$</sup> , Phe<sup>153 $\beta$</sup> , and Pro<sup>154 $\beta$</sup> ), as predicted (9). A superposition of 2C and 14.3.d shows that 2C  $V\beta-C\beta$  has rotated around this ball-and-socket joint relative to 14.3.d so that the elbow angle in 2C is  $148.6^\circ$  (36), compared to about  $155^\circ$  in 14.3.d. The conservation of the ball-and-socket joint suggests some possible flexibility in the TCR elbow angle, contrary to some hypotheses of TCR signal transduction that invoke rigidity of the  $\beta$  chain (40).

**The variable regions.** As predicted from sequence alignments (4), the 2C variable domains resemble antibody variable domains (Fig. 2B). Both  $V\alpha$  and  $V\beta$  have residues that are conserved across all mouse and human TCRs and many of these are similarly conserved in Fab sequences. Their locations in Fab x-ray structures correspond to structurally important locations, such as the  $V_L-V_H$  interface or the  $\beta$ -sheet framework. The TCR has its own subset of conserved residues, which are not found in Fab sequences and, hence, indicates residues or regions that relate directly to their own



**Fig. 6.** Stereoview of the  $V\alpha-V\beta$  interface. (**Top**) Backbone representation including interacting conserved interdomain TCR side chains ( $\alpha$ , pink;  $\beta$ , blue). The labeled side chains are conserved across all mouse and human TCRs and are also conserved in antibody  $V_H-V_L$  interfaces. Water molecules (small blue spheres) are present along the edges of the interface. (**Bottom**) The  $V\alpha-V\beta$  complementarity is shown by this molecular surface cross-section. The cluster of aromatic side chains between  $V\alpha$  and  $V\beta$  accounts for most of the buried surface area in the interface.

specific function.

In 2C,  $V\alpha$  and  $V\beta$  associate closely in a pseudo-dyad arrangement of  $174^\circ$ , well within the normal range for  $V_H$ - $V_L$  associations (7, 53, 54). Within the  $V\alpha$ - $V\beta$  interface (Fig. 6),  $1160 \text{ \AA}^2$  ( $610 \text{ \AA}^2$  for  $V\alpha$ ;  $550 \text{ \AA}^2$  for  $V\beta$ ) of surface area (49) is buried, as opposed to an average value of  $1420 \text{ \AA}^2$  (7, 54) found in Fabs (ranges from  $1068$  to  $1700 \text{ \AA}^2$ ). About 33 percent is contributed by CDR residues (mostly CDR3s), which is again smaller than the 40 to 50 percent average in Fabs (54). The  $\beta$ -sheet crossing angle (50) between the inside sheets of  $V\alpha$  and  $V\beta$  is  $-62^\circ$ , which is larger than for any Fab (ranges from  $-38^\circ$  to  $-60^\circ$ ), and  $10^\circ$  larger than for the  $V\alpha$  homodimer 1934.4 (16).

General conclusions about trends in the  $V\alpha$ - $V\beta$  association must await other TCR structures. Notwithstanding, the 2C  $V\alpha$ - $V\beta$  domain buried surface area value is at the extreme end of the range seen in antibodies. The only Fab determined thus far whose  $V_L$ - $V_H$  buried surface area is as small as 2C  $V\alpha$ - $V\beta$  is Fab 50.1 ( $1063 \text{ \AA}^2$ ), of which  $\sim 35$  percent is contributed by the CDR loops (54). Fab 50.1 experiences a very large  $V_L$ - $V_H$  rotational displacement of  $16.3^\circ$  on binding its peptide antigen, which may be the largest interdomain rotation observed so far between free and bound Fab structures (54). Smaller  $V_L$ - $V_H$  contact surface areas appear to correlate with an ability to use changes in domain association for better complementarity of fit with antigen (54, 55).

The interdomain contact residues between  $V\alpha$  and  $V\beta$  are conserved in TCRs (Fig. 6) and in the corresponding  $V_L$ - $V_H$  interface in antibodies (9). These residues form a twofold symmetric core with an invariant pair of side-chain to side-chain H bonds between  $\text{Gln}^{37\alpha}$  and  $\text{Gln}^{37\beta}$  at the base of the interface. Both  $\text{Tyr}^{35\alpha}$  and  $\text{Tyr}^{35\beta}$  H-bond to the opposing backbone of CDR3 residues  $\text{Leu}^{106\beta}$  and  $\text{Leu}^{104\alpha}$ , respectively. Three interdomain H bonds form along the edge of the inner  $\beta$  sheets, which are water mediated ( $\text{Leu}^{45\alpha}$ -water- $\text{Leu}^{106\beta}$ ,  $\text{Phe}^{106\alpha}$ -water- $\text{Leu}^{43\beta}$ ,  $\text{Ser}^{108\alpha}$ -OH-water- $\text{Gly}^{40\beta}$ ). The other conserved residues ( $\text{Leu}^{43\alpha}$ ,  $\text{Leu}^{45\alpha}$ ,  $\text{Phe}^{89\alpha}$ ,  $\text{Ala}^{91\alpha}$ ,  $\text{Leu}^{43\beta}$ ,  $\text{Leu}^{45\beta}$ ,  $\text{Phe}^{91\beta}$ , and  $\text{Ala}^{93\beta}$ ) form a symmetrical hydrophobic core between the  $V\alpha$ - $V\beta$  inner  $\beta$  sheets, as predicted (9).

Within each  $V\alpha$  and  $V\beta$ , about 24 residues form the core of the  $\beta$ -sheet sandwich and are conserved in mouse and human TCRs, as well as in Igs (8, 9). When  $V\beta$  is superimposed onto six representative Fab  $V_H$ 's (56) by overlapping 36 structurally equivalent framework residues, an average  $1.6 \text{ \AA}$  rms deviation is obtained for up to all 70  $\beta$ -sheet framework residues (56). A sim-

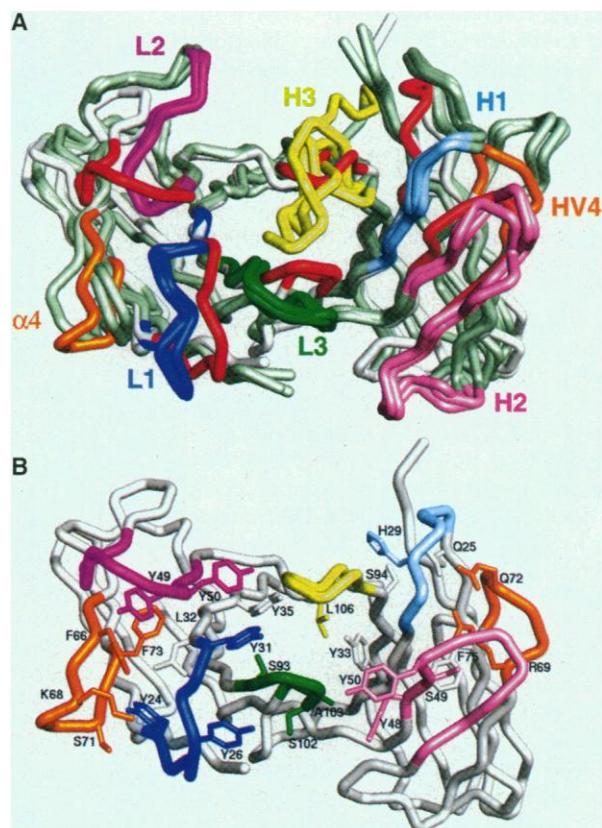
ilar overlap of  $V\alpha$  onto six  $V_L$ 's yields an average rms deviation of  $1.2 \text{ \AA}$  for 76 structurally equivalent  $V\alpha$  residues, indicating that  $V\alpha$  is more similar to  $V_L$  than  $V\beta$  is to  $V_H$  (56). Superimposition of  $V\beta$  (57) is equally good with all  $V_H$ 's and  $V_L$ 's in the Protein Data Bank ( $\sim 2.0 \text{ \AA}$  rms deviation), but  $V\alpha$  appears to superimpose better with  $V_L$  than with  $V_H$  ( $1.6 \text{ \AA}$  versus  $2.0 \text{ \AA}$  rms deviation).

In the 2C  $V\alpha$  domain, as in the 1934.4  $V\alpha$  structure (16), the  $c'$  strand is H-bonded to the d strand of the outer  $\beta$  sheet as opposed to the  $c'$  strand, as in v-type Ig domains (17). Framework residue  $\text{Lys}^{77\alpha}$ , which is highly conserved in  $V\alpha$  subgroup I but not in Ig frameworks, is at the carboxyl end of the e strand and H-bonds to the main chain of the d strand, possibly keeping the solvent-exposed strand-switched  $c''$ -d loop pinned next to the core of the  $V\alpha$  domain. The  $V\beta$  domain folds into a canonical v-type Ig set (17). An invariant residue in the framework of subgroup II  $\beta$

chains, but not in Igs, is  $\text{His}^{47\beta}$ , which H-bonds to the d strand and helps tether the  $c''$ -d loop, known as HV4, to the main body of  $V\beta$ . HV4 is an additional region of hypervariability observed in TCRs thought to be involved in superantigen ( $V\beta$ ) or accessory molecule ( $V\alpha$ ) interaction (58). Since  $\text{His}^{47\alpha}$  is not conserved in any other mouse or human subgroup, the position of  $V\beta$  HV4 may differ in other TCRs.

The relative positions of the 2C CDRs, with the exception of  $V\alpha$  CDR2, are close to, but not exactly coincident with, corresponding CDRs in antibodies (Fig. 7A). The  $V\alpha$  CDR1 is eight residues long and is stabilized by hydrophobic interactions between highly conserved residues  $\text{Tyr}^{24\alpha}$ ,  $\text{Leu}^{32\alpha}$ , and  $\text{Phe}^{66\alpha}$  (Fig. 7B). An H bond between the  $\text{Tyr}^{26\alpha}$  and  $\text{Ser}^{102\alpha}$  hydroxyls serves to anchor both CDR1 and CDR3 (Fig. 7B). As the TCR CDR1 does not vary in length by more than three residues, its conformation is probably rather similar in most TCRs. However, CDR2, is more vari-

**Fig. 7.** The antigen binding site of a TCR. **(A)** Comparison of TCR and antibody combining sites. An overlap of 2C  $V\alpha$  and  $V\beta$  is shown with three representative Fab  $V_L$  and  $V_H$  domains (56). The relative dispositions of the CDRs can then be compared.  $V\alpha$  was optimally superimposed with three Fab  $V_L$ 's, and  $V\beta$  was separately superimposed with three Fab  $V_H$ 's with the program OVLAP (36). The 2C CDRs are colored in red (except HV4, which is orange, and its equivalent in  $\alpha$  chains, labeled  $\alpha 4$ ). The Fab CDRs are labeled (L1, L2, L3, H1, H2, and H3) and are in different colors. The  $\alpha$  chain is to the left and CDRs  $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ , and  $\alpha 4$  correspond to the Fab CDRs L1, L2, L3, and L4 (only exists in TCRs, so is not colored in the Fabs), respectively. The  $\beta$  chain is to the right and CDRs  $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ , and HV4 correspond to the Fab CDRs H1, H2, and H3, with the fourth  $V\beta$  CDR again with no corresponding hypervariability in Fabs. **(B)** CDR stabilizing residues in the 2C TCR. The view into the combining site of 2C shows side chains that are important for CDR main-chain conformations and positions. The  $\alpha$  chain is to the left,  $\beta$  chain to the right. Backbone representations of the 2C CDRs are as follows:  $\alpha 1$ , navy;  $\alpha 2$ , magenta;  $\alpha 3$ , green;  $\alpha 4$ , orange;  $\beta 1$ , cyan;  $\beta 2$ , pink;  $\beta 3$ , yellow; and HV4, orange. The assignment of structurally important residues is based on contacts [van der Waals, H bond, or salt bridge as determined by CONTACTSYM and HBPLUS (52)] of framework (colored white) or CDR (colored by CDR code) residues with main-chain or side-chain residues of a CDR. These residues may affect the conformation of the CDR itself or regions outside of the CDR that are critical to the positioning of the CDR within the combining site. [Figure produced with MidasPlus version 2.0 (69)]



able, as reflected by its different conformation in 2C versus 1934.4 (16). The top of the loop bulges away from the center of the combining site, with the phenolic ring of Tyr<sup>49α</sup> taking the place of the main-chain path in CDR2 of 1934.4. Tyr<sup>49α</sup> provides most of the specific interactions for this loop (Fig. 7B), through an H bond from its hydroxyl to the main-chain carbonyl of residue 52α. Hydrophobic interactions of the Tyr<sup>49α</sup> aromatic ring with Phe<sup>66α</sup> and Phe<sup>73α</sup> pack the CDR2 (c'-c'' strands) against the Vα core (Fig. 7B). The main chain of Vα CDR2 is almost perpendicular to that of CDR2 in antibodies (Fig. 7A) because of the c'' strand switch. Vα CDR3 is short (eight residues long) and has a type-II β turn conformation, with Phe<sup>100</sup>, protruding from the surface of the binding site (Fig. 8B). Two H bonds from the strand leading to the CDR1 turn (Ser<sup>93α</sup>-OH-Tyr<sup>31α</sup>-OH and Ser<sup>102α</sup>-OH-Tyr<sup>26α</sup>-OH) connect to residues in CDR3 (Fig. 7B). Ala<sup>103α</sup> in CDR3 is within the Vα-Vβ interface and buried in a hydrophobic pocket composed entirely of Vβ aromatic residues (Fig. 7B).

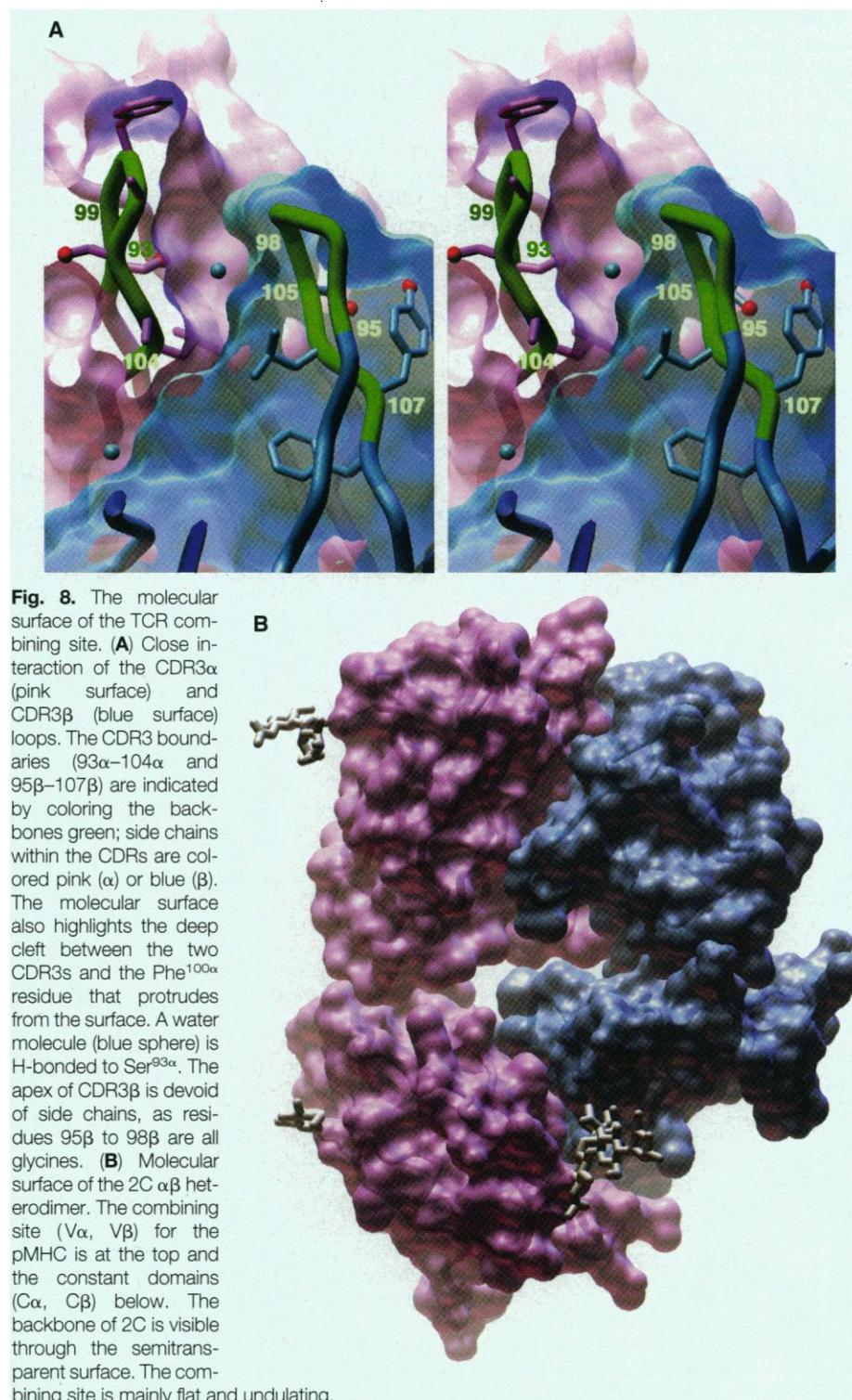
The conformation of Vβ CDR1 is similar to 14.3.d CDR1 (15) and is strongly constrained by main-chain H bonds with the side chain of Gln<sup>25β</sup> (Fig. 7B). CDR2 also has the identical hairpin conformation as 14.3.d and maintains a similar set of H bonds between Ser<sup>49β</sup> and Arg<sup>69β</sup>. HV4, which is entirely solvent accessible, has its conformation maintained by intraloop H bonds from the Arg<sup>69β</sup> guanidinium group to the main-chain atoms of residues 71 and 72. The tip of HV4 is held toward the center of the binding site by H bonds between Gln<sup>72β</sup> and side-chain and main-chain atoms of residues 28β and 29β at the tip of CDR1 (Fig. 7B). CDR3 has a classic gamma-type turn at its tip (59) with an unusual repeat of four consecutive glycine residues (Fig. 8A). The Vβ CDR3 interacts with Vα CDR1 through an H bond from Tyr<sup>33α</sup> to the main-chain carbonyl of Vβ Leu<sup>106β</sup> (Fig. 7B). Additional stabilization of the loop position may be derived from a water-mediated H bond from the main chain of Gly<sup>98β</sup> to the Ser<sup>98α</sup>-OH in CDR3. The Vβ CDR3 moieties of many TCRs are glycine-rich, which is the consequence of TCR D regions coding for a high proportion of glycines in all three reading frames (60).

The structural elucidation of the CDRs of 2C enables us to address the issue of CDR boundaries. The CDR boundaries in TCRs have been chosen by analogy to the Kabat hypervariable regions in antibodies but are not as clearly delineated (6, 39). A structural definition of a CDR as those residues within the hypervariable loops that are actually accessible to antigen reveals that the structural CDRs can be shorter than the sequence-

defined CDRs (61). Using this accessibility criterion, the CDR ranges for 2C would be as follows: CDR1α, 24-31; CDR2α, 48-55; CDR3α, 93-104; CDR1β, 26-31; CDR2β, 48-55; and CDR3β, 95-107.

Five of the six CDRs in antibodies fall into a small number of main-chain structural subclasses (61) that can now be predicted on the basis of loop size, composition, and

the presence of certain conserved residues. These commonly occurring conformations of the CDRs, called canonical structures (61), are universal for antibodies and are not specific to individual Kabat subgroups (39). Whether canonical CDR structures exist for TCRs has not been clearly identifiable from modeling of TCR sequences on antibodies (40). The 2C TCR structure shows that



**Fig. 8.** The molecular surface of the TCR combining site. **(A)** Close interaction of the CDR3α (pink surface) and CDR3β (blue surface) loops. The CDR3 boundaries (93α-104α and 95β-107β) are indicated by coloring the backbones green; side chains within the CDRs are colored pink (α) or blue (β). The molecular surface also highlights the deep cleft between the two CDR3s and the Phe<sup>100α</sup> residue that protrudes from the surface. A water molecule (blue sphere) is H-bonded to Ser<sup>93α</sup>. The apex of CDR3β is devoid of side chains, as residues 95β to 98β are all glycines. **(B)** Molecular surface of the 2C αβ heterodimer. The combining site (Vα, Vβ) for the pMHC is at the top and the constant domains (Cα, Cβ) below. The backbone of 2C is visible through the semitransparent surface. The combining site is mainly flat and undulating.

canonical CDR conformations would most likely exist only within Kabat subgroups, but not across all TCRs. Many of the CDR-stabilizing residues in 2C are conserved only within  $\beta$  subgroup II and  $\alpha$  subgroup I. For instance, Gln<sup>25 $\beta$</sup>  appears to play a major role in stabilizing CDR1 $\beta$  but is invariant only in subgroup II of the mouse and human, whereas in subgroup I it is Pro. Ser<sup>49 $\beta$</sup>  and Arg<sup>69 $\beta$</sup> , H-bonded to CDR2 $\beta$  (Fig. 7B), are

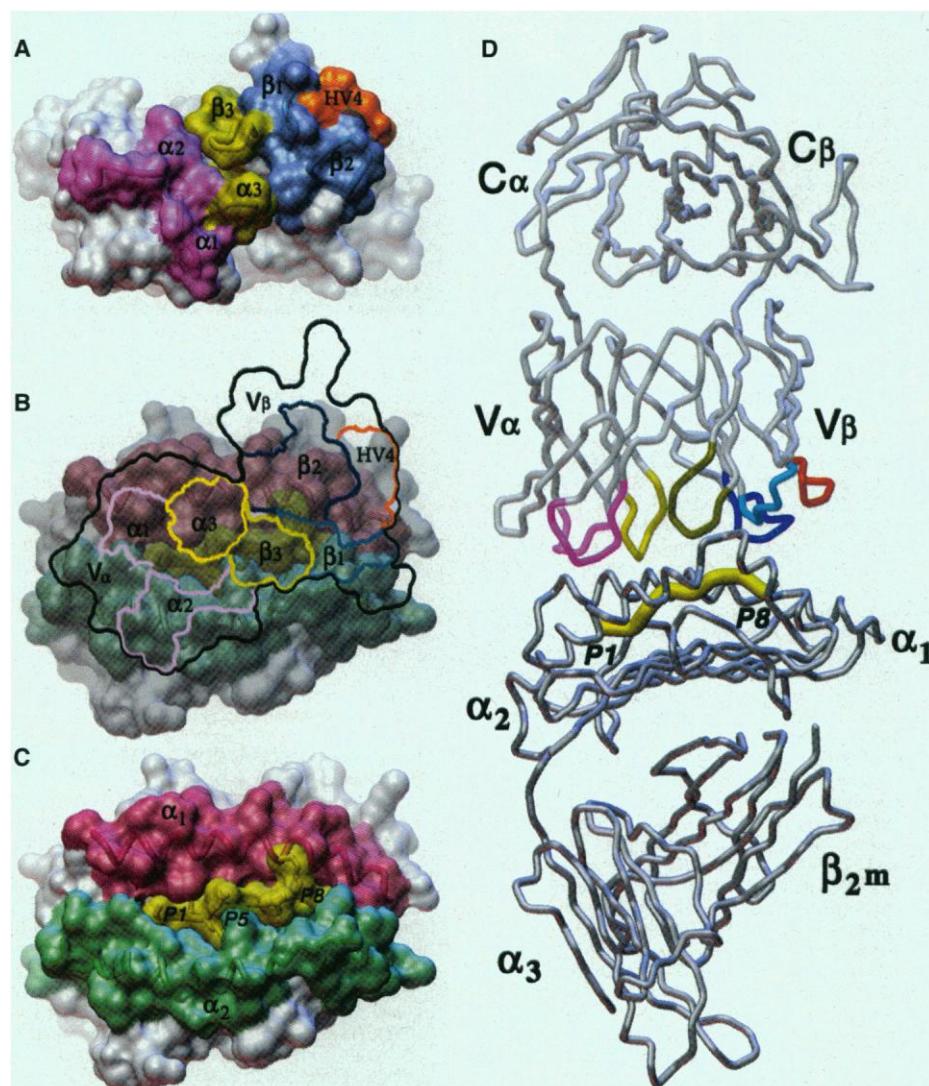
invariant in  $\beta$  subgroup II, but in other subgroups can be Phe (human subgroup I) or Phe(His) (mouse subgroup I). The CDR2 $\alpha$  loop conformation and position appear to require a hydrophobic interaction between Tyr<sup>49 $\alpha$</sup>  and Phe<sup>66 $\alpha$</sup> , which are conserved in  $\alpha$  subgroup Ia, but are small hydrophobic residues in other subgroups. Subgroup distinctions in TCRs then appear to be important determinants of antigen recognition and in-

dividual TCR functions (62).

The overall binding surface of 2C is relatively flat (Figs. 2B and 8B). With the exception of a hydrophobic pocket between CDR3 $\alpha$  and CDR3 $\beta$  (Fig. 8A), the absence of large protruding loops or deep clefts is characteristic of the undulating surfaces of antibodies to proteins (7). The CDR surface is neither exceptionally hydrophobic or polar, but rather presents an even distribution of residues of both chemical types. One feature of the binding surface is the paucity of large side chains protruding from the tips of the CDRs. Many glycine and other small residues are at the apices of the CDRs (Ala<sup>28 $\alpha$</sup> , Gly<sup>52 $\alpha$</sup> , Gly<sup>99 $\alpha$</sup> , Ala<sup>101 $\alpha$</sup> , Ala<sup>103 $\alpha$</sup> , Gly<sup>51 $\beta$</sup> , Gly<sup>53 $\beta$</sup> , Gly<sup>95 $\beta$</sup> , Gly<sup>96 $\beta$</sup> , Gly<sup>97 $\beta$</sup> , and Gly<sup>98 $\beta$</sup> ) in positions that would seem likely to be within TCR-pMHC interface. The absence of protruding side chains could allow the TCR to approach the pMHC surface more closely so that a smaller subset of side chains could supply the main binding energy for specific TCR-pMHC interactions.

Several models have been proposed to describe the relative orientation of a TCR with respect to a pMHC (4, 12–14). The CDR3s of the  $\alpha$  and  $\beta$  chains are proposed to interact predominantly with the bound peptide, whereas the locations of the less diverse CDRs 1 and 2 are thought to lie over the  $\alpha$  helices of the MHC (4, 12, 13) or partially interact with both helices and peptide (14). The approximate overall dimension (63) of the 2C binding surface is 36 Å by 20 Å (Fig. 8B), as compared to the H-2K<sup>b</sup> binding surface of 32 Å by 22 Å. The tips of CDR3 $\alpha$  (101 $\alpha$ ) and CDR3 $\beta$  (97 $\beta$ ) are 7.3 Å apart (Fig. 8A). A hydrophobic pocket between the two CDR3s (~10 Å deep) appears ideally situated to accommodate one large side chain (64) that would presumably come from the peptide component of the pMHC complex (Fig. 8A). The exposed peptide surface in H-2K<sup>b</sup> is ~12 Å wide; that could allow both CDR3 $\alpha$  and CDR3 $\beta$  to fit between the MHC  $\alpha$  helices and hence interact with the peptide side chains. The  $\alpha$  helices of the pMHC are about 18 Å apart from each other, nearly the same as the separation between CDRs 1 and 2 of the  $\alpha$  chain and CDRs 1 and 2 of the  $\beta$  chain.

**Orientation of the TCR-pMHC Complex.** Crystals that diffract to beyond 3 Å have been grown of the 2C TCR in complex with the class I MHC molecule H-2K<sup>b</sup> bound to the self-peptide dEV8 (26, 65). A complete x-ray data set has been collected to 3.4 Å, and the positions of the TCR and pMHC molecules have been determined by molecular replacement (65) with the use of the refined crystal structures of the 2C TCR reported here and the 2.5 Å H-2K<sup>b</sup>-OVA class I MHC crystal structure reported previously (66).



**Fig. 9.** The TCR-pMHC interaction. (A) Molecular surface of the 2C binding site viewed from above the CDRs. The surface and loop trace of the V $\alpha$  and V $\beta$  CDRs  $\alpha_1$  and  $\alpha_2$  are magenta; CDRs  $\beta_1$  and  $\beta_2$ , blue; V $\alpha$  and V $\beta$  CDR3s, yellow; and V $\beta$  HV4, orange. (B) Footprint of the 2C binding site molecular surface on the H-2K<sup>b</sup>-peptide molecular surface. The outline and footprint of the TCR molecule was obtained by slicing through the TCR surface in the vicinity of the combining site and projecting that surface onto the pMHC molecule. The view is looking directly onto the H-2K<sup>b</sup>-peptide surfaces as in (C) through the 2C TCR in the complex. The TCR molecular surface is outlined in black and the borders delineating CDR subregions ( $\alpha_1$ -3,  $\beta_1$ -3, and HV4) are indicated. The pink lines separate the molecular surfaces of the CDRs  $\alpha_1$  and  $\alpha_2$ ; blue, the CDRs  $\beta_1$  and  $\beta_2$ ; and yellow, the CDR3s. (C) Molecular surface of the H-2K<sup>b</sup>-OVA model used to determine the position of the H-2K<sup>b</sup>-dEV8 in the crystals of the 2C-H-2K<sup>b</sup>-dEV8 complex. The peptide in the figure still retains the OVA sequence of the search model (66). (D) Backbone tube representation of the oriented 2C-H-2K<sup>b</sup>-peptide complex from the molecular replacement solution. The pMHC is below with the octamer peptide (P1-P8) shown as a large tube in yellow. The 2C TCR is above with the  $\alpha_1$  and  $\alpha_2$  CDRs colored pink;  $\alpha$ HV4, white; CDRs  $\beta_1$  and  $\beta_2$ , blue;  $\beta$ HV4, orange; and the CDR3s, yellow.

In the 2C-H-2K<sup>b</sup>-dEV8 complex (Fig. 9), the TCR combining site interacts with the  $\alpha$  helices from  $\alpha_1$  and  $\alpha_2$  and the exposed residues of the bound peptide of the pMHC. A footprint of the TCR interaction surface on the pMHC molecule (Fig. 9, A to C) indicates that the long dimensions of the TCR and pMHC binding surfaces are not quite parallel to each other, but tilted about 20° to 30° toward the diagonal. The footprint (Fig. 9B) reveals that V $\alpha$  CDR $\alpha$ 1 and  $\alpha$ 2 form a patch covering the peptide NH<sub>2</sub>-terminal residues and parts of both pMHC  $\alpha_1$  and  $\alpha_2$  helices. The CDR $\alpha$ 1 lies over both the NH<sub>2</sub>-terminal region of the pMHC  $\alpha_1$  helix and the three NH<sub>2</sub>-terminal residues of the peptide, whereas CDR $\alpha$ 2 lies mostly over the COOH-terminal end of the pMHC  $\alpha_2$  helix. Conversely, V $\beta$  CDR $\beta$ 2 mostly covers the COOH-terminal end of the pMHC  $\alpha_1$  helix with some slight coverage of the COOH-terminal end of the peptide. CDR $\beta$ 1 is mostly over the NH<sub>2</sub>-terminal end of the  $\alpha_2$  helix, but also appears to have some interaction with the COOH-terminal residues of the peptide. The V $\alpha$  and V $\beta$  HV4s, even though they lie over the pMHC in the footprint (Fig. 9B), do not appear close enough to the pMHC surface (Fig. 9D) to take part in significant interatomic contacts. The V $\alpha$  and V $\beta$  CDR3s appear to play the prominent role in peptide interaction and straddle the central region of the octameric (P1 to P8) peptide within the pMHC  $\alpha_1$  and  $\alpha_2$  helices, with the P4 position of the peptide lying between the two CDR3s. The footprint of CDR $\beta$ 3 extends from the P4 position of the peptide all the way to P7. Refinement of the structure should yield further, detailed information (67).

The orientation of 2C to H-2K<sup>b</sup>-dEV8 resembles the model (14) for a TCR-MHC class II complex (D10-IA<sup>K</sup>), in which the CDRs  $\alpha$ 1 and  $\alpha$ 2 lie over the NH<sub>2</sub>-terminal region of the peptide, CDR3 $\alpha$  interacts with the central P5 position of the peptide, and CDR3 $\beta$  has interactions extending from P5 to P8. These restrictions would place the CDRs  $\beta$ 1 and  $\beta$ 2 over the  $\beta_1$  and  $\alpha_1$  helices, respectively. The 2C-H-2K<sup>b</sup>-dEV8 orientation differs from the model (4, 12), in which the TCR is rotated about 90° relative to the pMHC surface. In that model, the CDRs 1 and 2 of both the  $\alpha$  and  $\beta$  chain interact predominantly with the  $\alpha$  helices of the  $\alpha$  and  $\beta$  chains of the pMHC. However, both models (4, 12-14) were derived for class II TCR-MHC systems, and the absence of similar structure-function data for class I MHC interactions with TCR has precluded proposals for a more detailed model until now. In our TCR-pMHC orientation, the TCR and pMHC surfaces have substantial overlap that provides a structural framework for the central role of the V $\alpha$  and V $\beta$  CDR3s in peptide recognition; CDRs 1

and 2 contact both peptide and the  $\alpha$ -helical segments of the MHC molecule.

## REFERENCES AND NOTES

- D. H. Katz, T. Hamaoka, B. Benacerraf, *J. Exp. Med.* **137**, 1405 (1973); E. M. Shevach and A. S. Rosenthal, *ibid.* **138**, 1213 (1973); P. C. Doherty, R. V. Blanden, R. M. Zinkernagel, *Transplant. Rev.* **29**, 89 (1976).
- H. Clevers, B. Alarcon, T. Wileman, C. A. Terhorst, *Annu. Rev. Immunol.* **6**, 629 (1988).
- H. Saito *et al.*, *Nature* **312**, 36 (1984) [cloning of 2C  $\alpha$  chain]; S. M. Hedrick, D. I. Cohen, E. A. Nielsen, M. M. Davis, *ibid.* **308**, 149 (1984); S. M. Hedrick, E. A. Nielsen, J. Kavalier, D. I. Cohen, M. M. Davis, *ibid.*, p. 153; Y. Yangi *et al.*, *ibid.*, p. 145; Y. Chien *et al.*, *ibid.* **312**, 31 (1984).
- M. M. Davis and P. J. Bjorkman, *ibid.* **334**, 395 (1988).
- S. Tonegawa, *ibid.* **302**, 575 (1981); M. Kronenberg *et al.*, *Annu. Rev. Immunol.* **4**, 529 (1986).
- P. Patten *et al.*, *Nature* **312**, 40 (1984).
- D. R. Davies, E. A. Padlan, S. Sheriff, *Annu. Rev. Biochem.* **59**, 439 (1990); I. A. Wilson and R. L. Stanfield, *Curr. Opin. Struct. Biol.* **4**, 857 (1994); I. A. Wilson *et al.*, *Methods Enzymol.* **203**, 153 (1991); I. A. Wilson *et al.*, *Ciba Found. Symp.* **5**, 139, (1991); E. A. Padlan, *Mol. Immunol.* **31**, 169 (1994).
- J. Novotny, S. Tonegawa, H. Saito, D. M. Kranz, H. N. Eisen, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 742 (1986).
- C. Chothia, D. R. Bosswell, A. M. Lesk, *EMBO J.* **7**, 3745 (1988).
- L. M. Amzel and R. J. Poljak, *Annu. Rev. Biochem.* **48**, 961 (1979).
- I. Engel and S. M. Hedrick, *Cell* **54**, 473 (1988); C. D. Katayama *et al.*, *EMBO J.* **14**, 927 (1995).
- J. L. Jorgensen *et al.*, *Nature* **355**, 224 (1992).
- J.-M. Claverie, A. Prochnika-Chalufour, L. Bougueret, *Immunol. Today* **10**, 10 (1989).
- D. B. Sant'Angelo *et al.*, *Immunity* **4**, 367 (1996).
- G. A. Bentley, G. Boulot, K. Karjalainen, R. A. Mariuzza, *Science* **267**, 1984 (1995).
- B. A. Fields *et al.*, *ibid.* **270**, 1821 (1995). Coordinates for 1934.4 V $\alpha$  were provided by R. A. Mariuzza. The V $\alpha$  domain has an Ig-fold topology, but with an unexpected strand switch in which the c' strand is H-bonded to the d strand of the outer  $\beta$  sheet, in contrast to the inner c' strand, as in Ig variable domains.
- Four distinct Ig structural subtypes have been defined on the basis of the number of strands and the spacing between c' and e strands: (i) c type, seven-stranded, as in C domains, with ~21 amino acids spacing; (ii) v type, nine-stranded, as in V domains, with ~25 amino acids spacing; (iii) s type, seven-stranded, strand-switched with ~9 amino acids spacing; and (iv) h type, a hybrid between c and s types as described by P. Bork, L. Holm, and C. Sander (*J. Mol. Biol.* **242**, 309 (1994)).
- I. Engel, T. H. Ottenhoff, R. D. Klausner, *Science* **256**, 1318 (1992); K. L. Hilyard *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9057 (1994); Y. Reiter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 281 (1994); W. F. Soo Hoo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 4759 (1993); J. Novotny *et al.*, *ibid.* **88**, 8646 (1991).
- S. Weber *et al.*, *Nature* **356**, 793 (1992); C. Gregoire *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8077 (1991).
- A. Y. Lin *et al.*, *Science* **249**, 677 (1990).
- S. Chung *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12654 (1994).
- H. C. Chang *et al.*, *ibid.*, p. 11408.
- R. K. Strong *et al.*, *J. Immunol.* **153**, 4111 (1994).
- E. S. Ward *et al.*, *Scand. J. Immunol.* **34**, 215 (1991).
- W. C. Sha *et al.*, *Nature* **335**, 271 (1994).
- For H-2K<sup>b</sup>-dEV8: M. D. Talquist, T. J. Yan, L. R. Pease, *J. Exp. Med.*, **184**, 1017 (1996); L. Teyton *et al.*, in preparation. For H-2L<sup>d</sup>-p2Ca: Y. Sykulev *et al.*, *Immunity* **1**, 15 (1994); M. Corr *et al.*, *Science* **265**, 946 (1994).
- M. R. Jackson, E. S. Song, Y. Yang, P. A. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 12117 (1992); M. Matsumura *et al.*, *J. Biol. Chem.* **267**, 23589 (1992).
- Carboxypeptidases A and B (Boehringer Mannheim) were used at a ratio of enzyme to TCR of 1:200 (w/w) at 23°C (O/N) in phosphate buffered saline (PBS) (pH 7.2) to remove any free, disordered COOH-termini that might inhibit crystallization. The interchain disulfide (Cys<sup>213</sup> $\alpha$ -Cys<sup>247</sup> $\beta$ ) provided an absolute stop point for the enzyme. No more than two to four residues remained after the disulfide, as estimated from mass spectrometric molecular mass determination, but no electron density was visible beyond the disulfide. To confirm that the TCR in the crystals is the disulfide-bonded heterodimer, we washed the crystals exhaustively in protein-free mother liquor and analyzed them. In the absence of reducing agent, the dissolved crystals showed on SDS-polyacrylamide gel electrophoresis a homogeneous molecular mass of 58 kD, which on reduction gave two bands of 30 kD ( $\beta$  chain) and 28 kD ( $\alpha$  chain). Monomeric  $\alpha$  and  $\beta$  chains, or nondisulfide-linked dimers, could not be detected in the crystal. Two-dimensional gel electrophoresis revealed a 1:1 ratio of both the acidic ( $\alpha$ ) and basic ( $\beta$ ) chains. Sequencing of the residues NH<sub>2</sub>-terminal confirmed the 1:1 ratio of both chains in the nonreduced 58-kD band. In addition, in BIAcore binding experiments, the expressed 2C material specifically bound both its pMHCs [H-2L<sup>d</sup>-p2Ca (26) and H-2K<sup>b</sup>-dEV8 (26)], and the conformation-specific mAbs-clonotypic 1B2 [D. M. Kranz, S. Tonegawa, H. N. Eisen, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7922 (1984)] and V $\beta$ 8-specific F23.1 [U. D. Stear *et al.*, *J. Immunol.* **134**, 3994 (1985)].
- For other TCRs expressed and purified in our laboratory, elimination or reduction of carbohydrate content has not so far been a successful strategy for improving crystal quality.
- Although this crystal was initially assigned to the C2 space group (with one molecule in the asymmetric unit), further examination of the data integrated as P2 (with two molecules in the asymmetric unit) revealed very weak but above background intensities for  $h + k = 2n + 1$  reflections, which gradually increased at higher resolutions, an indication of strong pseudo-centering of the cell.
- X-ray crystallographic programs and methods: (i) X-PLOR: A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987); A. T. Brünger, X-PLOR, Version 3.1: A System for X-ray and NMR (Yale Univ. Press, New Haven, CT, 1992); (ii) R-free: A. T. Brünger, *Nature* **355**, 472 (1992); (iii) PHASES: W. J. Furey, A program package and analysis of diffraction data from macromolecules (American Crystallographic Association Fortieth Anniversary Meeting, New Orleans, 1990); (iv) CCP4: Collaborative Computational Project No. 4, Daresbury, UK (1994); *Acta Crystallogr. D* **50**, 760 (1994); (v) O: T. A. Jones, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991); (vi) SIGMA: R. J. Read, *ibid.* **42**, 140 (1986); (vii) AMoRe: J. Navaza, *Acta Crystallogr. A* **50**, 157 (1994); (viii) V<sub>n</sub>: B. W. Matthews, *J. Mol. Biol.* **33**, 491 (1968); (ix) DENZO and SCALEPACK: Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Bailey, Eds. (SERC Daresbury Laboratory, Warrington, UK, 1993), p. 56; (x) PROCHECK: R. A. Laskowski, M. W. MacArthur, S. D. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993); (xi) 3D-PROFILE: R. Luthy *et al.*, *Nature* **356**, 83 (1992); and (xii) ERRAT: C. Colovos and T. O. Yeates, *Protein Sci.* **2**, 1511 (1993).
- In the crystal structure of an Fc fragment of an Ig [J. Deisenhofer, *Biochemistry* **20**, 2361 (1981)], regions with disorder and high B values are present in both the C<sub>H</sub>2 and C<sub>H</sub>3 domains. The entire constant region of Fab NC10 was disordered in the crystal structure of the NC10-neuraminidase complex [R. L. Malby *et al.*, *Structure* **2**, 733 (1994)].
- Cryocooling is not easily reproducible with the 2C crystals. After screening many cryoprotectants, we obtained moderate success by the gradual introduction of glycerol at a rate of 5 percent per hour to a final concentration of 25 percent. Still, this procedure most often resulted in corrupted diffraction patterns (split reflections, loss of resolution, smearing, or high mosaicity). Eventually, after more than 50 attempts with identical soaking and cryoconditions, one crystal was successfully frozen, which diffracted to higher resolution (2.45 Å versus 2.9 Å) with lower mosa-

- icity (0.4°) than did room temperature crystals (0.6°).
34. Electron density was traceable as far as the COOH-terminal cysteines in both chains; the  $C_{\alpha}$ - $C_{\alpha}$  distance between the cysteine residues was appropriate for a disulfide bond, but its electron density was weak.
  35. The NC41 Fab from an antibody to neuraminidase has a similar elbow angle of 148° [P. M. Colman *et al.*, *Nature* **326**, 358 (1987)].
  36. The elbow angle of the TCR, defined as the angle between the pseudo-twofold rotation axes relating the  $V_{\alpha}$  and  $V_{\beta}$  domains to the  $C_{\alpha}$  and  $C_{\beta}$  domains, was determined with the programs OVRLAP [M. G. Rossman and P. Argos, *J. Biol. Chem.* **250**, 7525 (1975)] and ELBOW (D. H. Fremont and T. O. Yeates, personal communication). To determine the positions of the pseudo-twofold axes, we superimposed the  $\alpha$  chain domains with the use of a pairwise method (57) onto the corresponding  $\beta$  chain domains with the same conserved  $C_{\alpha}$  atoms that were used for overlaps of the individual TCR domains with antibodies (56).
  37. J. H. Brown *et al.*, *Nature* **364**, 33 (1994).
  38. K. C. Garcia *et al.*, in preparation.
  39. All TCR and Fab numbering in this article follow E. A. Kabat *et al.*, Eds., *Sequences of Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD, ed. 5, 1991).
  40. G. A. Bentley and R. A. Mariuzza, *Annu. Rev. Immunol.* **14**, 563 (1996).
  41. With DALI, version 2.0, L. Holm and C. J. Sander, *J. Mol. Biol.* **233**, 123 (1993).
  42.  $C_{\beta}$  residues 218 to 220 are involved in extensive van der Waals contacts with a symmetry-related molecule in the 2C crystal lattice.
  43. *Drosophila melanogaster* cells are insect cells that correctly recognize N-linked glycosylation signals, Asn-X-Thr (Ser) [M. J. Fraser, *In Vitro Cell. Dev. Biol.* **25**, 225 (1989)]. However, the N-linked oligosaccharide structures produced are usually smaller than those produced in mammalian cells [K. Kuroda *et al.*, *Virology* **174**, 418 (1990)]. The fucose moiety has an  $\alpha$ 1,6 linkage to the first GlcNAc that is directly connected to the side-chain amide nitrogen of the Asn-X-Ser (Thr) sequence. The carbohydrate content was estimated by mass spectrometry of glycosylated 2C (non-carboxypeptidase digested) and unglycosylated 2C (non-carboxypeptidase digested) prepared by inducing expression in the presence of tunicamycin (1  $\mu$ g/ml). The glycosylated 2C was 58,829 daltons and the unglycosylated 2C was 53,965 daltons. On the assumption that each N-linked glycosylation site has 1 to 1.5 kD of carbohydrate, then only four or five possible N-linked sites are occupied. Alternatively, N-linked glycosylation may not have been completely inhibited.
  44. The average  $B$  values indicate that the  $C_{\alpha}$  domain has higher thermal parameters (Table 1). The average  $B$  value for the  $C_{\alpha}$  top strands (c, f, and g) is 69  $\text{\AA}^2$ , and for the bottom strands (a, b, e, and d) is 41  $\text{\AA}^2$ . However, the density is clear and unambiguous (Fig. 1). The  $C_{\alpha}$  domain has no crystal lattice packing interactions that probably contribute to the higher thermal parameters.
  45. The immunoglobulin disulfide has a  $\chi_2$  of  $\sim 180^\circ$  [J. S. Richardson, *Adv. Prot. Chem.* **34**, 167 (1981)].
  46. The program DALI (41) was used to search for structurally similar proteins or domains among 642 structures in the Protein Data Bank. The highest Z scores (4.0 to 5.0, which are very low values) were obtained with Ig fold-containing molecules based on alignments with the back  $\beta$  sheet (and part of strand c) of the  $C_{\alpha}$  (2.8 to 3.2  $\text{\AA}$  rms deviation over 68 to 71 residues). No reasonable alignments were found in which the top strands were included in the superposition.
  47. K. P. Kearse *et al.*, *EMBO J.* **13**, 4504 (1994).
  48. C. Geisler *et al.*, *J. Immunol.* **148**, 3469 (1992); S. Caspar-Bauguil *et al.*, *Scand. J. Immunol.* **40**, 323 (1994).
  49. All molecular surface areas buried by interaction were calculated with the program MS [M. L. Connolly, *J. Appl. Crystallogr.* **16**, 439 (1983)] with a 1.7  $\text{\AA}$  probe sphere and standard atomic radii, as described in (7).
  50. To define the  $\beta$ -sheet crossing angle, vectors were fit to the  $C_{\alpha}$  atoms in structurally equivalent, but opposing,  $\beta$  strands in the interface of each of the interacting domains. The crossing angle was then the angle between these two vectors. The structurally equivalent residues chosen in opposing  $\beta$  strands were as follows: (i)  $V_{\alpha}$ (30 to 38, 103 to 108)- $V_{\beta}$ (32 to 39, 105 to 112); (ii)  $C_{\alpha}$ (138 to 145, 175 to 185)- $C_{\beta}$ (143 to 151, 190 to 198); (iii)  $C_{H_1}$ 3 (363 to 370, 403 to 411)- $C_{H_2}$ 3(363 to 370, 403 to 411); (iv)  $C_L$  (132 to 138, 173 to 180)- $C_{H_1}$ 1 (140 to 145, 186 to 192); (v)  $V_L$  (30 to 38, 103 to 108)- $V_{\alpha}$ (32 to 39, 105 to 112); and (vi)  $V_{\alpha}$  (31 to 37, 103 to 108)- $V_{\alpha}$  (31 to 37, 103 to 108).
  51. For pairwise comparisons, an  $\alpha$  chain TCR domain was first superimposed on the corresponding Fab L chain domain (or  $C_{H_1}$ 3 domain). The same transformation was then applied to the  $\beta$  chain domains to superimpose on the H chain (or  $C_{H_2}$ 3) domain. The additional rotation and translation required to optimize the  $\beta$  chain-H chain superposition describes the domain shift. This procedure is similar to that described in (35) with superpositions specified in (56).
  52. Hydrogen bond interactions were based on both distance (3.5  $\text{\AA}$  cutoff) and geometrical considerations with HBPLUS [J. K. McDonald and J. M. Thornton, *J. Mol. Biol.* **238**, 777 (1994)]. The van der Waals and salt-bridge interactions were calculated with the program CONTACTSYM [S. Sheriff, W. A. Hendrickson, J. L. Smith, *ibid.* **197**, 273 (1987)].
  53. A pairwise superposition of all  $V_H$ - $V_L$  heterodimers from available Fab x-ray structures in the Protein Data Bank [F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977)] with the  $V_{\alpha}$ - $V_{\beta}$  of 2C shows a range from an essentially identical domain-domain rotational relationship with many antibodies to up to 17° differences with others.
  54. R. L. Stanfield *et al.*, *Structure* **1**, 83 (1993).
  55. P. M. Colman, *Adv. Immunol.* **43**, 99 (1988).
  56. The six Fabs used for comparison represent a range of isotypes. The Fabs were KOL [M. Marquart *et al.*, *J. Mol. Biol.* **141**, 369 (1980)]; 3D6 [X. M. He *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7154 (1992)]; Sel55-4 [M. Cygler *et al.*, *Science* **253**, 442 (1991)]; BV04 [J. N. Heron *et al.*, *Prot. Struct. Func. Genet.* **11**, 159 (1991)]; J539 [T. N. Bhat *et al.*, *ibid.* **1**, 74 (1986)]; 17/9 [J. M. Rini *et al.*, *Science* **255**, 959 (1992)]. For calculating rms deviations for the TCR and Fabs, the models were superimposed by means of OVRLAP (36). Conserved, structurally equivalent  $\beta$ -sheet framework residues in both TCR and Fabs were chosen for the superposition. For 2C  $V_{\alpha}$ , residues 3 to 5, 19 to 24, 32 to 37, 44 to 47, 63 to 65, 72 to 76, 88 to 91, and 110 to 112 were superimposed on Fab  $V_L$  residues 4 to 6, 20 to 25, 33 to 38, 45 to 48, 63 to 65, 70 to 74, 86 to 89, and 102 to 104, and an rms deviation was calculated for the resulting 77 structurally equivalent framework residues. For 2C  $V_{\beta}$ , residues 4 to 6, 20 to 25, 30 to 31, 34 to 37, 44 to 47, 66 to 68, 74 to 78, 90 to 93, and 112 to 114 were superimposed on  $V_{H_1}$  residues 4 to 6, 19 to 24, 34 to 35, 36 to 39, 46 to 49, 68 to 70, 77 to 81, 90 to 93, and 107 to 109, and the rms deviation was calculated for 70 structurally equivalent framework residues. For 2C  $C_{\beta}$ , framework residues 126 to 129, 144 to 149, 158, 160 to 161, 172 to 174, 192 to 197, and 212 to 215 were superimposed on  $C_{H_1}$ 1 residues 120 to 123, 139 to 144, 153, 156 to 157, 171 to 173, 187 to 192, and 208 to 211. For 2C  $C_{\alpha}$ , 29 framework residues in the back  $\beta$  sheet were chosen, 123 to 127, 129, 139 to 143, 145, 147, 159 to 162, 164, 179 to 185, 190, 204, 206 to 207 for overlap with  $C_L$  residues 113 to 117, 119, 132 to 136, 138, 140, 158 to 161, 163, 174 to 180, 193, 205, and 207 to 208. The Fabs in Fig. 7A were superimposed with 2C as described above and were F9.13.7 [J. Lescar, H. Souchen, P. Alzari, *Protein Sci.* **3**, 788 (1994)]; D11.15 [V. Chitarra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7711 (1993)]; and CNJ206 [R. Zemel *et al.*, *Mol. Immunol.* **31**, 127 (1994)].
  57. Each domain of 2C was submitted to the DALI (41) site. The average rms deviation for overlapping about 100 residues of each TCRV domain with other antibody V domains (L and H) of Fabs was  $\sim 2.0$   $\text{\AA}$  for  $V_{\beta}$  (Z values of 14 and 15) and 1.6  $\text{\AA}$  for  $V_{\alpha}$  (Z values of 14 and 15), with a slight skewing toward lower rms deviations of  $V_{\alpha}$  with  $V_L$  domains.
  58. Y. Choi *et al.*, *Nature* **346**, 471 (1990) [ $V_{\beta}$  interaction with superantigen]; A. V. Chernovsky *et al.*, The 9th International Congress of Immunology, Abstract 2955 [ $V_{\alpha}$  interaction with the accessory CD4].
  59. Despite the high degree of flexibility expected for a tetraglycine stretch of amino acids in a solvent exposed loop, the electron density for CDR3 is excellent. The  $\phi$  and  $\psi$  angles of CDR3 $\beta$  fall within accepted ranges for a classic gamma-type turn [J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981)].
  60. C. Abergall and J.-M. Claverie, *Eur. J. Immunol.* **21**, 3021 (1991).
  61. C. Chothia and A. M. Lesk, *J. Mol. Biol.* **196**, 901 (1987); C. Chothia *et al.*, *Nature* **342**, 877 (1989).
  62. M. Schiffer, E. A. Kabat, T. T. Wu, *Immunogenetics* **35**, 224 (1992).
  63. Length was measured from the tip of CDR1 $\alpha$  to CDR2 $\beta$  and width was measured from the furthest ends of CDR1 $\alpha$  to CDR2 $\alpha$  or CDR1 $\beta$  to CDR2 $\beta$ .
  64. A proposal for a TCR pocket that specifically binds a particular side chain is made for the class II I-E<sup>K</sup> system in [G. J. Kersh and P. M. Allen, *J. Exp. Med.*, in press].
  65. The 2C-H-2K<sup>b</sup>-dEV8 complex was prepared by mixing a 1:1 stoichiometric ratio of carboxypeptidase-treated and HIC-purified 2C with *D. melanogaster*-expressed H-2K<sup>b</sup> in the presence of a 1.5-fold excess of the dEV8 peptide (EQYKFYSV) (26). Crystals were grown by vapor diffusion in the presence of 10 percent PEG-6000, 0.2M tris-acetate (pH 7.2). Repeated macroseeding resulted in large crystals that could be cooled to 100K by the slow addition of ethylene glycol to a concentration of 22 percent. Although diffraction from these crystals now extends to beyond 3  $\text{\AA}$ , an initial data set was collected to 3.4  $\text{\AA}$  on a MAR imaging plate at the UCSD X-ray facility. The data were indexed, integrated, and scaled with DENZO and SCALEPACK (31). The space group was determined to be  $P2_12_12_1$ , with cell dimensions  $a = 296.7$   $\text{\AA}$ ,  $b = 89.73$   $\text{\AA}$ , and  $c = 84.5$   $\text{\AA}$ . The overall  $R_{\text{sym}}$  is 10.5 percent on  $I$ 's with a completeness of 85 percent to 3.4  $\text{\AA}$  (80 percent completeness and  $1/\sigma$  of 4.5 in the outer 3.4 to 3.5  $\text{\AA}$  shell). The 2C MR solution was determined with normalized structure factors ( $F'$ s) in AMoRe (31). The 2C coordinates were then placed in the cell and the rotation and translation functions were determined with H-2K<sup>b</sup>-OVA (66) as a model. One round of rigid body refinement that was followed by positional refinement in X-PLOR (37) lowered the  $R$  value to 34.5 percent. Inspection of electron density maps at this point indicated clear electron density for both 2C and H-2K<sup>b</sup>-dEV8.
  66. D. H. Fremont, E. A. Stura, M. Matsumura, P. A. Peterson, I. A. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2479 (1995).
  67. K. C. Garcia, M. Degano, P. A. Peterson, L. Teyton, I. A. Wilson, unpublished data.
  68. O. Epp *et al.*, *Eur. J. Biochem.* **45**, 513 (1974).
  69. Graphics Programs: (i) The program TOM was compiled by Christian Cambillau (1987); (ii) AVS: C. Upton *et al.*, *IEEE Computer Graphics and Application* **9**, 30 (1989); (iii) MidasPlus 2.0: T. E. Ferrin, C. C. Huang, L. E. Jarvis, R. J. Langridge, *J. Mol. Graph.* **6**, 13 (1988); and (iv) PQMS: M. L. Connolly, *ibid.* **11**, 139 (1993).
  70. We thank E. Stura for advice and help with data collection and heavy atom screening; T. Yeates, L. Pease, D. Kranz, H. Eisen, C. Scott, and R. Stefanko for discussions, materials, and assistance; N. H. Xuong for use of the UCSD X-ray facility; R. Lerner for constant encouragement and support; and M. Pique for production of figures. Supported by NIH RO1 CA58896 (I.A.W.) and NIH postdoctoral training grant T32-A107244 (K.C.G.) This is publication 10295-M from The Scripps Research Institute. The coordinates for 2C have been deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, NY) with accession code 1TCR.

23 August 1996; accepted 15 September 1996