Structural basis of tyrosine sulfation and V_H-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120

Chih-chin Huang*, Miro Venturi*, Shahzad Majeed*†, Michael J. Moore‡, Sanjay Phogat§, Mei-Yun Zhang§, Dimiter S. Dimitrov§, Wayne A. Hendrickson†¶, James Robinson‖, Joseph Sodroski††, Richard Wyatt*,**, Hyeryun Choe‡‡, Michael Farzan‡, and Peter D. Kwong*†§§**

*Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892; †Department of Biochemistry and Molecular Biophysics and ¶Howard Hughes Medical Institute, Columbia University, New York, NY 10032; ‡Brigham and Women's Hospital, and ‡‡Children's Hospital, Department of Pediatrics, Department of Medicine (Microbiology and Molecular Genetics), and **Department of Cancer Immunology and AIDS, Dana–Farber Cancer Institute, and Department of Pathology, Division of AIDS, Harvard Medical School, Boston, MA 02115; §National Cancer Institute, National Institutes of Health, Frederick, MD 21702; ‖Department of Pediatrics, Tulane University Medical Center, New Orleans, LA 70112; and ††Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA 02115

Contributed by Wayne A. Hendrickson, December 19, 2003

The conserved surface of the HIV-1 gp120 envelope glycoprotein that binds to the HIV-1 coreceptor is protected from humoral recognition by multiple layers of camouflage. Here we present sequence and genomic analyses for 12 antibodies that pierce these defenses and determine the crystal structures of 5. The data reveal mechanisms and atomic-level details for three unusual immune features: posttranslational mimicry of coreceptor by tyrosine sulfation of antibody, an alternative molecular mechanism controlling such sulfation, and highly selective V_H-gene usage. When confronted by extraordinary viral defenses, the immune system unveils novel adaptive capabilities, with tyrosine sulfation enhancing the vocabulary of antigen recognition.

The HIV type 1 (HIV-1) utilizes a variety of mechanisms to evade the humoral immune response. Chief among these mechanisms are those used by the exterior gp120 envelope glycoprotein, the principal component of the viral spike and the primary target of antibody recognition (1). Antigenic loop variation, oligomeric occlusion, conformational masking, and glycan cloaking (either by "self-camouflage" or through steric occlusion) (2–4), all enable viral escape from antibody-mediated neutralization, thus facilitating a persistent infection leading to host immune dysfunction and the development of acquired immunodeficiency syndrome (AIDS) (5, 6).

Despite these sophisticated mechanisms of evasion, limitations related to functional constraints of viral entry create opportunities for antibody recognition. Foremost among these involve virus–cell-surface receptor interactions. HIV-1 propagates in only a select subset of immune cells, identified by the primary viral receptor, CD4 (7, 8), and by a coreceptor, generally CCR5 or CXCR4 (reviewed in ref. 9). The coreceptors are chemokine receptors, seven-helix integral membrane proteins. Recognition by HIV-1 gp120 involves interactions primarily with their second extracellular loop as well as their N-termini (10–13), which are distinguished by a concentration of tyrosines, modified by posttranslational addition of sulfate (14). Tyrosine sulfation of coreceptor is critical for gp120 recognition (14).

The gp120 surface that interacts with the coreceptors overlaps the epitopes for an emerging group of antibodies, which were originally identified as being induced by CD4 binding and thus were labeled "CD4i" antibodies (15). The antibodies display extremely broad HIV-1 recognition, although neutralization potency may be restricted by adjacent variable loops and steric and conformational constraints (16), indicating that gp120 has evolved to protect this conserved surface. We previously determined the crystal structures for primary and laboratory-adapted core gp120 molecules in complex with the CD4 receptor and the archetype CD4i antibody, 17b (17, 18). In these structures, 17b showed a relatively small surface of interaction (≈500 Å^2), dominated by interactions involving a 19-residue heavy chain third complementarity-determining region (CDR H3). The protruding nature of the paratope suggested that 17b was accessing a sterically restricted surface.

In light of the multiple mechanisms that protect the coreceptor-binding surface on gp120, we asked what novel antibody features might be necessary for recognition of this highly protected site. Because the unusual features that we observed for 17b might be exclusive to this particular antibody, we examined a panel of CD4i antibodies, isolated from five different individuals and from two different phage display libraries. Here we describe sequence and genomic analyses for 12 of these CD4i antibodies, and we crystallize and determine the x-ray structures of 5. In a separate manuscript we report biochemical and mutagenic analyses (19). These studies reveal atomic-level details for immune mechanisms involving posttranslational mimicry and selective V_H-gene usage.

Materials and Methods

CD4i Antibody Origin. Peripheral blood B cells from HIV-1 infected subjects were transformed with Epstein–Barr virus to generate stable B cell lines, and antibodies were selected for gp120 reactivity. Antibody 17b from asymptomatic subject N70 and antibody 48d from asymptomatic subject Y76 have been described previously (15). Antibodies isolated from subjects undergoing structured treatment interruption (20) included 47e, 412d, and E51 from subject AC-01 and 16c from subject AC-13. Monoclonal antibodies 23e and 411g were derived from a long-term nonprogressor, subject AD19, who had not been treated with antiretroviral drugs (21). Antibodies were also isolated from phage display libraries prepared from the bone marrow RNA of HIV-1 infected individuals. Antibodies C12, Sb1, and X5 were isolated from a library prepared from subject FDA-2, whose serum exhibited potent broadly neutralizing anti-HIV activity (22). Antibody m16 was isolated from a library

Abbreviations: CDR, complementarity-determining region; Fab, antigen-binding antibody fragment; HIV-1, HIV type 1.

Data deposition: The atomic coordinates of the CD4i antibodies 48d, 17b, 47e, 412d, and E51 (PDB ID codes 1rz7, 1rz8, 1rzi, 1rzg, and 1rzf, respectively), as well as the refined coordinates of the YU2 and HXBc2 ternary complexes (PDB ID codes 1rzk and 1rzj, respectively) have been deposited in the Protein Data Bank, www.pdb.org. The nucleotide sequences reported in this paper have been deposited in the GenBank database (accession nos. AY515002–AY515013, AY539808, and AY539809).

§§To whom correspondence should be addressed at: Vaccine Research Center, NIAID/NIH; 40 Convent Drive; Building 40, Room 4508, Bethesda, MD 20892-3027. E-mail: pdkwong@nih.gov.

prepared from RNA of three long-term nonprogressors, whose sera exhibited the broadest and most potent HIV-neutralization among 37 HIV-infected individuals (23).

CD4i Antibody Sequencing and Genomic Analysis. The sequences of 16c, 411g, 23e, 47e, 412d, and E51 were determined by using methods described previously for 17b and 48d (3). The sequencing of C12 and Sb1 from phage libraries is reported in a companion manuscript (19); m16 and X5 were sequenced by following the same procedure. Genomic analysis was carried out with nucleotide sequences submitted to IGBLAST (24) or IMGT (25). Details are contained in *Supporting Experimental Procedures*, which is published as supporting information on the PNAS web site.

Crystallography. The antigen-binding fragment (Fab) of monoclonal antibodies 48d, 17b, 47e, 412d, and E51 was made and purified by using methods similar to those described previously (26), except that with the sulfated antibodies, endoproteinase Lys-C was used for digestion and anion-exchange chromatography was used to purify the fully sulfated Fab. Details on crystallization, data collection, structure solution, and molecular modeling and docking are presented in *Supporting Experimental Procedures*. Final refinement parameters are presented in Table 2, which is published as supporting information on the PNAS web site.

Biochemical Analysis of Sulfation. Labeling and immunoprecipitation of antibodies (X5, m16, and b12) and modulation of sulfation by RNA interference for Fab (X5, m16, and b12) and single-chain antibody variable fragment (scFv; C12 and Sb1) were performed as described previously (19).

Figures. All superpositions were carried out with the program LSQKAB in the CCP4 package (27). Figures were made with GRASP (28), XTALVIEW (29), or RASTER3D (30).

Results and Discussion

Structural and Genomic Analysis of 17b. To more fully understand the recognition of gp120 by the archetype CD4i antibody, 17b, we determined the structure of the uncomplexed 17b Fab at 2.2-Å resolution (Fig. 1). The uncomplexed 17b crystallized with two molecules in the asymmetric unit, with the protruding acidic CDR H3 forming 204 Å^2 of the noncrystallographic twofold interface. The entire CDR H3 was ordered, with an average *B* factor of 25.2 Å^2, which was lower than the average *B* factor of 28.2 Å^2 for the entire Fab.

The large crystal lattice contact in the uncomplexed 17b structure might be expected to influence the CDR H3 orientation, especially when compared to the gp120-complexed 17b, where 243 Å^2 of the CDR H3 surface is buried in the gp120 interface (18). However, superposition of free and complexed 17b showed little CDR H3 change upon complex formation (Fig. 1*A*). For the gp120 contact surface, root-mean-square deviations (rmsd) after superposition of the Fab variable domains for the two free and two complexed 17b showed main-chain rmsd of 0.39–0.53 Å and side-chain rmsd of 1.1–1.2 Å. By contrast, elbow angle differences between free and complexed of 11.6°–27.1° showed that crystal lattice packing could strongly influence the relative orientation of the variable and constant portions of the Fab, with superpositions of the entire Fab yielding main-chain rmsd of up to 3.3 Å. In light of the observed CDR H3 rigidity, we were surprised to observe only three internal backbone–backbone hydrogen bonds in the uncomplexed 19-residue loop. Loop rigidity depended primarily on interactions with side chains from other CDR loops.

Genomic analysis identified the 17b heavy chain precursor to be VH1-69, with 20 of its 98 residues altered during affinity

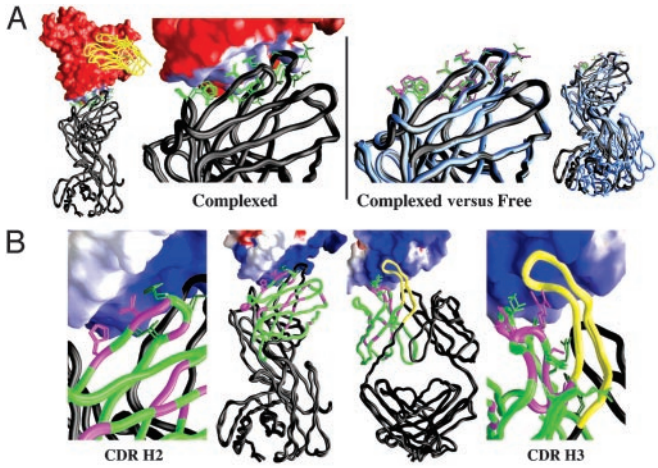

Fig. 1. Structure of the archetype CD4i antibody, 17b. (*A*) Complexed versus free structure of 17b. The *Left* two structures show the rerefined YU2 and HXBc2 ternary complexes after superposition of the 17b V_H framework, with the two complexed Fab 17b in black C^α worm, interacting 17b side chains in green, the N-terminal domain of CD4 in yellow, and the molecular surface of YU2 core gp120 in red, except for the surface within 3.5 Å of 17b, which is blue. In this orientation, the viral membrane would be positioned toward the top of the page. The *Right* two structures show the two independent copies of free 17b from the $P2_12_12_1$ crystals superimposed on the complexed structures. The color and orientation for the complexed structures are the same as in *Left*, with the free 17b structures shown in blue with magenta interactive residues. The *Far Right* shows the entire Fab, including the constant portion. Whereas the variable domains are quite similar, considerable differences are seen in the constant portions, especially between the two free structures. (*B*) Details of gp120–17b interaction at CDR H2 and CDR H3. The electrostatic potential of gp120 is shown at the molecular surface colored blue for electropositive, red for acidic, and white for apolar. The *Left* two structures show 17b in the same orientation as *A*. The portion corresponding to the V_H gene, VH1-69, has been colored green, except for residues altered by somatic mutation, which are colored magenta. The five side chains of the CDR H2 that interact with gp120 are shown: I52, I53, L54, V56, and H58. The *Right* two structures show an ≈90° view, adjusted so that the pseudo twofold axes of the Fab are aligned with the edges of the page. In this view, the acidic CDR H3 loop (yellow C^α worm) can be seen reaching up to contact a basic gp120 surface. Side chains of VH1-69 that interact with the CDR H3 loop are shown.

maturation. These included 5 of the 16 side chains that support the CDR H3, including a Ser-to-Arg change at residue 31, the side chain of which buries 74 Å^2 of the CDR H3 surface (Fig. 1*B*). These changes suggested that the CDR H3 before affinity maturation might have been more flexible, and that increasing its rigidity enhanced its affinity with gp120.

Analysis of the interface between 17b and gp120 found a number of direct contacts with the CDR H2. Two of the five residues in the CDR H2 were altered by somatic mutation (Fig. 1*B*) to enhance hydrophobic interactions (Thr to Val) and to optimize hydrogen-bonding geometry (Asn to His, with the His $N^{\varepsilon 2}$ making a hydrogen bond to Thr-202 $O^{\gamma 1}$ of gp120). These interactions suggested that in addition to loop rigidification, direct contacts with less protruding portions of the epitope were important for high-affinity interaction.

The structure and genomic analysis of 17b thus identified a number of specific attributes that might permit recognition of the highly cloaked chemokine receptor-binding site. These included a highly acidic CDR loop, a relatively small surface of interaction, and a protruding paratope, rigidified by somatic mutation. To see how general these features were for recognition, we examined these traits in a panel of CD4i antibodies.

Isolation, Sequencing, and Genomic Analysis of a Panel of CD4i Antibodies. The sequences of 48d and 17b have recently been published (3). We were able to obtain sequences for six addi-

BIOCHEMISTRY

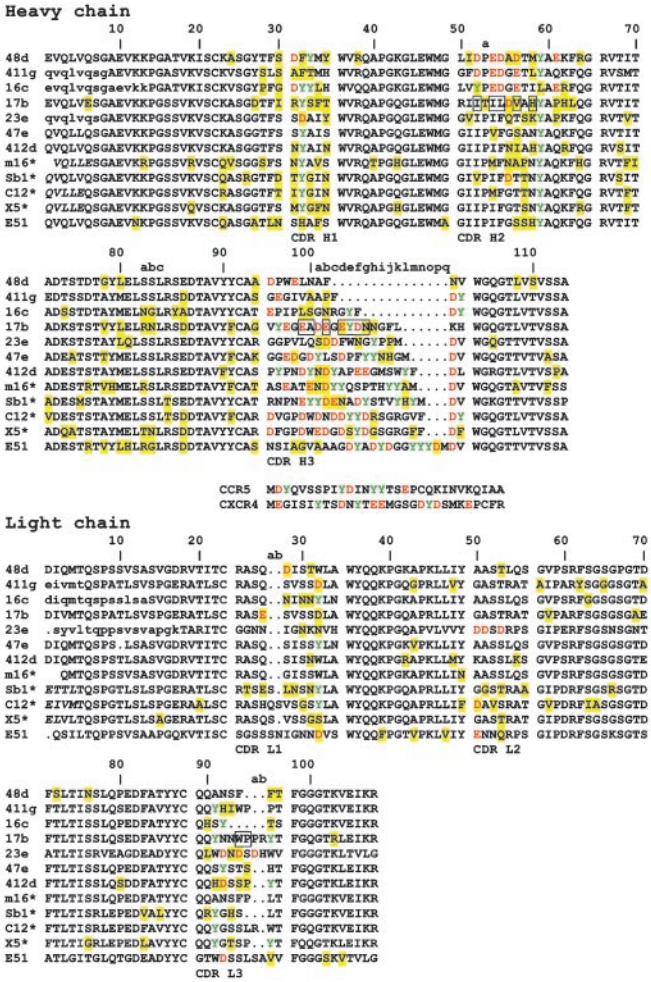

Fig. 2. CD4i antibody variable domain sequences. A multiple sequence alignment of CD4i antibodies is shown with CDRs labeled. Both Kabat CDR definitions and numbering are used. Sequences have been ordered according to CDR H3 length, which varies from 10 for 48d to 25 for E51. Antibodies isolated from phage display are labeled with an asterisk. N-terminal sequences influenced by sequencing primers are shown in lowercase; those influenced by phage library construction are shown in italics. Somatic mutations are highlighted with yellow background, acidic CDR residues with red, and CDR tyrosines with green. Boxed residues of 17b contact gp120. Because of allelic differences, somatic mutations were determined by using the closest genomic progenitor. The N termini of CCR5 and CXCR4 are shown for reference.

tional CD4i antibodies isolated from immortalized B cells, 16c, 23e, 47e, 412d, 411g, and E51. Screening and sequencing of phage display libraries derived from HIV-1-infected individuals produced four additional sequences, C12, m16, Sb1, and X5 (Fig. 2). Genomic analysis showed extraordinarily selective V_H-gene usage, with 9 of the 12 antibodies using VH1-69, and the other 3 using VH1-24 (Table 1). A higher usage of A27 was also observed in the phage display light chains, although this was probably related to particularities of the phage library. A range of affinity maturation changes was observed, from a low of 6.3% for 47e, to a high of 14.3% for Sb1 and E51. In general, antibodies obtained from phage display libraries had levels of changes (12.8 ± 1.2%) higher than those obtained from immortalized B cells (10.6 ± 2.5%), but the difference was not distinctive (Fig. 2, Table 1). Analysis of the CDR loops showed a variety of CDR H3 lengths, ranging from 10 (for 48d) to 25 (for E51). CDR H3 net charge as defined on the unmodified amino acids also varied from −1 (for 16c) to −6 (for X5). These

Table 1. V and J gene usage in gp120-reactive antibodies

	H chain		L chain		V and J
Antibody	V gene	J gene	V gene	J gene	mutations*
CD4i					
48d	VH1–24†	JH3	L19	JK4	23
411g	VH1–24	JH4	L2	JK1	17
16c	VH1–24†	JH4	O2/O12‡	JK4	17
17b	VH1–69	JH1	L2	JK2	27
23e	VH1–69	JH6	V2–14	JL3	17
47e	VH1–69	JH6	O2/O12‡	JK2	11
412d	VH1–69	JH2	L12	JK2	18
m16	VH1–69	JH6	L5	JK4	26
Sb1	VH1–69	JH6	A27	JK4	27
C12	VH1–69	JH4	A27	JK1	20
X5	VH1–69	JH4	A27	JK2	23
E51	VH1–69	JH6	V1–19	J3/JL2‡	28
CD4BS					
b12	VH1–3	JH6	A27	JK2	45
15e	VH4–39	JH6	O2	JK3	26
21h	VH3–11	JH5	V2–14	JL2	20
GP13	VH5–51§	JH6	B3	JK2	20
GP44	VH1–46	JH4	V1–2	JL2	22
GP68	VH1–69	JH4	L12	JK3	27
F105	VH4–59	JH4	A27a	JK2	13
F91	VH3–21	JH4	V3–3	JL3	33
V3 loop					
MN215	VH3–21	JH4	V1–22	JL2	13
D0142–10	VH1–24†	JH4	B3	JK3	33
ITH5–2	VH3–11	JH4	L15	JK5	24
447–52D	VH3–15	JH6	V1–19	JL3	14
Glycan					
2G12	VH3–21	JH3	L12	JK2	51

Nucleotide and protein sequences were submitted to ImMunoGeneTics (IMGT) and IGBLAST for analysis (24). Nomenclature from IGBLAST is used throughout this table.

*Because somatic mutations are difficult to discern in regions of V-D-J recombination, especially of N-addition and P-addition, and because of variable CDR H3 length, we list only mutations for V and J regions.

†IMGT aligned these sequences most closely with VH1-f, which has not been located in the genome.

‡These genomic sequences were equally homologous.

§IMGT aligned this sequence most closely with HV5-a, which has not been located in the genome.

observations suggested that the various traits identified as being unusual to 17b were not traits that were general to all CD4i antibodies. Segregating the antibodies into two groups, distinguished by CDR H3 length, allowed correlative properties to appear.

In the short group (CDR H3 length from 10 to 14), the CDR H3s were not very acidic (net average charge of −1.6 ± 0.5), whereas the CDR H2s were, displaying an unusual concentration of three or four acidic residues in a short six-residue stretch at the loop tip. In this short group, only VH1-24 was used (Fig. 2, Table 1).

In the long group (CDR H3 length from 19 to 25), the CDR H3s were acidic (net average charge of −4.9 ± 0.9), and in contrast to the short group, the CDR H2s were primarily electroneutral. For all nine antibodies in the long group, VHl-69 was the genomic progenitor. In addition, a preponderance of tyrosines was found in many of the long acidic CDR H3s (Fig. 2).

Tyrosine Sulfation of CD4i Antibodies. Because negative charge and tyrosine usage are often associated with posttranslational modification of tyrosine by O-sulfation, and because coreceptor sulfation is critical for gp120 recognition (14), we analyzed the

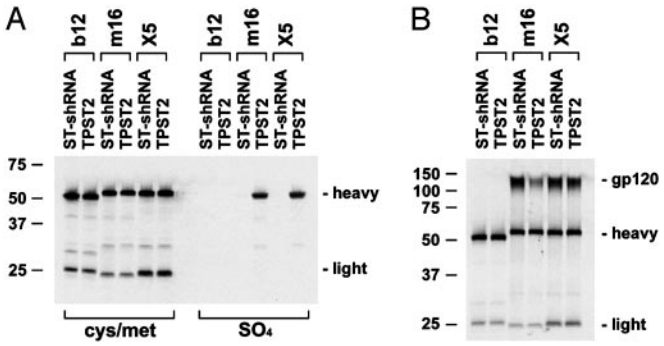

Fig. 3. Tyrosine sulfation and HIV-1 gp120 recognition for m16 and X5. (*A*) Modulation of antibody sulfation with sulfotransferase and small hairpin RNA (shRNA). Initially 293T cells were transfected with plasmid encoding shRNA complementary to the message of the two known tyrosyl protein sulfotransferases (ST-shRNA) or to an irrelevant message. Two days later, cells were transfected with the indicated antibody in the presence (TPST2) or absence of tyrosine sulfotransferases. Cells were divided and labeled with [^{35}S]cysteine and [^{35}S]methionine or [^{35}S]sulfate. Cell supernatants were immunoprecipitated with staphylococcal protein A-Sepharose and analyzed by SDS/PAGE. CD4 binding-site antibody b12 is shown as a control. (*B*) Sulfation does not contribute to gp120 association. Antibody from assay shown in *A* was incubated with radiolabeled gp120 of the CCR5-using isolate ADA and an excess of unlabeled CD4. Antibody and gp120/CD4 complexes were immunoprecipitated with protein A-Sepharose and analyzed by SDS/PAGE. With the functionally sulfated E51, such ST-shRNA treatment reduces binding by 6- to 8-fold (19).

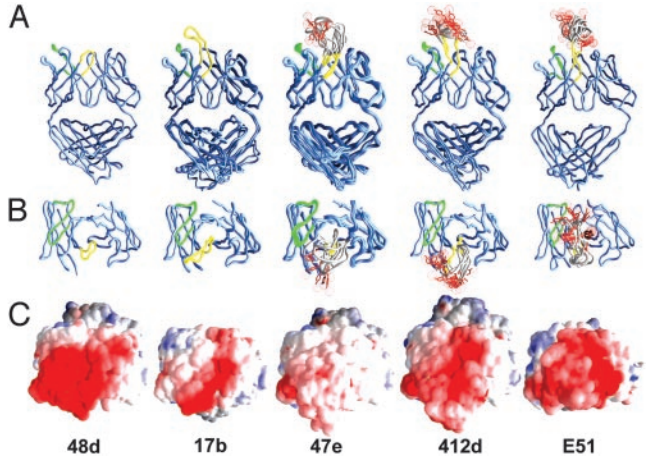

Fig. 4. Structure and electrostatic surface of 48d, 17b, 47e, 412d, and E51. CD4i antibodies were aligned by superposition of their V_H framework. The two independent copies of both 17b and 412d as well as the eight independent copies of 47e are shown. The disordered portions of the CDR H3 loops for 47e, 412d, and E51 were modeled and subjected to molecular dynamics. Shown for the disordered regions are 8 models of 47e and 10 for 412d and E51. (*A*) Fab structures. Blue C^{α} worms are shown with CDR H2 green and CDR H3 yellow. The modeled sections, subjected to molecular dynamics, are shown with C^{α} worm in gray and sulfated tyrosines in red. The orientation of this figure is the same as the *Right* two structures of Fig. 1*B*. (*B*) Perpendicular view of *A*, showing only the variable fragment (Fv) portion of the Fab for clarity. (*C*) Electrostatic surface. The electrostatic potential is displayed at the molecular surface, with electropositive surface in blue, electronegative surface in red, and apolar surface in white. The Fabs are in the same orientation as in *B*. Although diluted somewhat by the disorder of 47e, 412d, and E51, an acidic surface can be seen on all of the CD4i antibodies. The alterations in shape of the overall surface are due primarily to variations in the positions of the constant domains.

sequences for this modification. The sequence motif that specifies sulfation is only partially defined, with a dominant characteristic being three or four acidic amino acids within five residues of the sulfotyrosine (reviewed in ref. 31). Nonetheless, computational analysis (32) suggested that the sequences of at least 412d, m16, and Sb1 had high probabilities of being sulfated in their CDR H3.

We used metabolic labeling to define further the sulfation of these antibodies. Much of this characterization is reported in a separate manuscript (19), where we show that 47e, 412d, Sb1, C12, and E51 are sulfated in their CDR H3 and that tyrosine sulfation of 47e, 412d, and E51 contributes to recognition of some strains of gp120, especially the CCR5-using primary isolate ADA. Here we characterize several additional antibodies. We used [^{35}S]sulfate labeling to show that the antibodies m16 and X5 were tyrosine sulfated (Fig. 3). However, in contrast to our previous findings, interfering RNA inhibition of sulfotransferase activity did not significantly alter recognition of ADA gp120 by C12, m16, Sb1, or X5 (Fig. 3 and Fig. 6, which is published as supporting information on the PNAS web site). Although it is possible that autologous virus may bind differently than the ADA strain used in our study, the ADA strain was the one most sensitive to sulfation for 47e, 412d, and E51 (19).

Thus the CD4i antibodies showed two unusual traits: highly selective V_H-gene usage and tyrosine sulfation. The tyrosine sulfation appeared to be of two types: functional, where it contributed to recognition, and inadvertent, where recognition appeared independent of this posttranslational modification. As functional posttranslational modifications have not been previously observed with antibodies, we sought to use x-ray crystallography to reveal mechanistic and atomic-level details.

Structural Analysis of the CD4i Antibodies: 48d, 47e, 412d, and E51. For the short CDR H3 group, we crystallized 48d in space group $P2_12_12_1$, one molecule per asymmetric unit, and determined the structure at 2.0-Å resolution (Fig. 4). All of the CDR loops were well ordered, with average *B* factors of 36.2 Å^2, similar to the average *B* factor 29.3 Å^2 for the entire structure. In contrast to 17b, no protruding loop was observed. However, electrostatic analysis showed the unusual concentration of acidic residues in the CDR H2 sequence to indeed be highly electronegative, even more electronegative than the acidic CDR H3 of 17b (Fig. 4*C*).

For the functionally sulfated CDR H3 antibodies, we crystallized and determined the structures of all three: 47e (singly sulfated), 412d (doubly sulfated), and E51 (triply sulfated), at resolutions of 2.9, 2.0, and 1.7 Å, respectively. The 47e crystals were in space group *P*1, with eight molecules per asymmetric unit. In contrast to the ordered CDR loops of 17b and 48d, in all eight 47e molecules, the CDR H3 was disordered. For the 20 residues in the 47e CDR H3, an average of 13 were disordered, with residues 97–100i disordered in all copies (Fig. 4*A*).

The 412d crystals were in space group $P2_1$ with 2 molecules per asymmetric unit. In both 412d molecules, residues 100a–100g of the CDR H3 were disordered (Fig. 4*A*). In the ordered portions of the CDR H3, the two crystallographically distinct molecules assumed similar loop configurations. The sulfated tyrosine at residue 100 could be seen in both crystallographically distinct molecules, although it was better ordered in one (Fig. 5).

The E51 crystals were in space group $P2_12_12_1$ with one molecule per asymmetric unit. By chance, the E51 crystals were virtually isomorphous with the 48d crystals. Despite this similarity, the CDR H3 in E51 was almost completely disordered, with 12 of its 25 residues, residues 99–100j, too flexible to be observed in the 1.7-Å structure (Fig. 4*A*). For the first two sulfated tyrosines in E51, the entire residue was disordered; for the third sulfated tyrosine, the main-chain atoms could be discerned, but the side chain, including the entire phenol ring, could not.

Thus, in contrast to the 17b and 48d structures, the CDR H3s for all of the functionally sulfated antibodies were disordered. Of

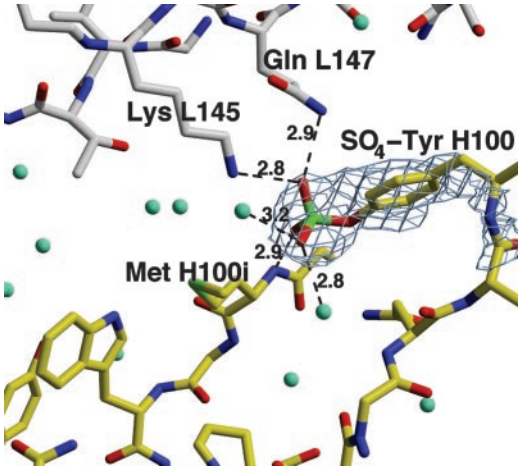

Fig. 5. Atomic-level details of antibody sulfation. The sulfated tyrosine at position H100 of 412d is shown. Two of the five coordinating ligands (Lys-145 and Gln-147) are from the light chain of a symmetry-related molecule. Electron density ($2F_o - F_c$) is shown at 0.5σ.

the 15 crystallographically distinct sulfated tyrosines, only 2 could be observed, both at the same position in 412d. Irrespective of this order/disorder, the overall electrostatics as well as the general shape of the combining surface were similar for the entire group of long-CDR H3 CD4i antibodies, with the paratopes of 17b, 47e, 412d, and E51 dominated by a protruding acidic CDR H3 (Fig. 4).

V_H-Gene Usage. We asked if we could explain the highly selective V_H-gene usage in light of the structures. For the short-CDR H3 group of antibodies, conservation of the highly electronegative paratope seen with 48d may explain the selective usage of VH1-24 and the homologous VH1-f, as these are the most acidic of the V_H-genomic progenitors. For the long-CDR H3 antibodies, the VH1-69 specifies the sequences for CDR H1 and H2, but not H3. It is involved in three interactions: with the CDR H3, with the light chain, and with gp120. In 17b, the primary residues of VH1-69 that buttress the protruding CDR H3 are Arg-31 (which buries 101 Å^2 of the CDR H3 surface), Tyr-32 (56 Å^2), and Arg-50 (44 Å^2). Examination of the CD4i antibodies shows that both Arg residues are unique to 17b. The light chain interactions, on the other hand, are highly conserved and do not distinguish VH1-69 from other V_H genes. Thus the VH1-69 conservation likely reflects some aspect of a direct interaction with gp120. Examination of the 17b–gp120 complex shows that the CDR H3 contributes roughly 50% of the buried surface, the CDR H2 35%, and the CDR L3 the remaining 15%. Thus if 17b were representative the only direct contacts that VH1-69 can make would involve the CDR H2, at residues 52, 53, 54, 56, or 58 (Fig. 1*B*).

We analyzed the conservation of these CDR H2 residues in the long-CDR H3 group of CD4i antibodies. Somatic mutation of all five residues was observed (Fig. 2), suggesting a mechanism distinct from that ascribed to germ-line carbohydrate recognition (33), which preserves a precise molecular interaction. The most distinctive conserved CDR H2 feature was the protruding hydrophobic residues at positions 53 and 54. These two residues account for 50% of the CDR H2 interactive surface. Their position at the loop tip allows them to protrude almost to the same extent as the long-CDR H3, where they form a hydrophobic anchor for CD4i antibody recognition. Analysis of genomic sequences shows that VH1-69 is the only V_H gene with hydrophobic residues at 53 and 54 (Fig. 7, which is published as supporting information on the PNAS web site). Thus, a precise molecular interaction is probably not preserved. Rather, the unusual VH1-69 gene usage likely reflects limitations of the VH-gene repertoire coupled to the need to preserve a protruding hydrophobic CDR H2 to anchor the long CDR H3 of the VH1-69-using CD4i antibodies.

Posttranslational Mimicry. Sequence and structural analyses show all of the CD4i antibodies to have a localized spot of high acidity in their antigen-combining sites. In the 17b–gp120 complex structures, this acidic patch interacts with gp120 residues shown to be important in binding the CCR5 N terminus. This interaction suggests a mimicry of the acidic N terminus of the viral coreceptor, CCR5 or CXCR4.

To understand this mimicry in greater detail, we examined the specific manner by which gp120 binds CCR5. It has been observed previously that if the CCR5 N-terminal sulfates are replaced with phosphates, binding is lost (34). Such sulfate/phosphate discrimination is rare, generally requiring full phosphate/sulfate coordination (35). Thus the observed gp120 discrimination suggests that gp120 likely fully coordinates a site of sulfate recognition. Such coordination would impose rigid stereochemical requirements on tyrosine-sulfate recognition.

In marked contrast to the ordered CDR H3 loops of the nonfunctionally sulfated CD4i antibodies, 48d and 17b, we observed CDR H3 disorder in all three of the functionally sulfated antibodies. Disorder is not a prerequisite for sulfation, as the CDR H3 loop of X5 is ordered (X. Ji and D.S.D., unpublished data). Rather, the disorder and associated flexibility may be the easiest way to provide the specific stereochemical orientation required for high-affinity tyrosine-sulfate recognition.

The CDR H3 is the only CDR with enough length and diversity to be a substrate for tyrosine sulfation. Examination of the CDR H3 for the CD4i antibodies shows virtually all of the sulfated tyrosines to result from recombination, encoded directly by D and J gene segments (Fig. 8, which is published as supporting information on the PNAS web site).

The data suggest the following requirements for functional sulfation of the CD4i antibodies: (*i*) VH1-69 CDR H2 binding to hydrophobic portions of the coreceptor binding site; (*ii*) tyrosine sulfation on a long acidic CDR H3 loop, which mimics the acidic N terminus of the coreceptor; and (*iii*) flexibility for the sulfated tyrosine, allowing for precise stereochemical recognition by gp120.

We envision that the combination of a CDR H2 hydrophobic anchor (in VH1-69) and the specific recognition of a CDR H3 sulfated tyrosine creates sufficient affinity to initiate antibody selection against the highly protected coreceptor binding site on gp120. Antibodies generated by this mechanism would incorporate posttranslational mimicry, with gp120 recognizing tyrosine sulfation on antibody in a manner similar to its natural recognition of the same posttranslational modification on coreceptor.

Inadvertent Posttranslational Modification: Substrate Mimicry. In addition to functional sulfation, our biochemical analysis suggests that inadvertent sulfation on the CDR H3 loop also occurs (Fig. 3). Such sulfation is clearly more than accidental: of the six long CDR H3 antibodies that were not functionally sulfated, we found four, or 67%, to be inadvertently sulfated. In contrast, none of the >250 deposited antibody structures in the Protein Data Bank is sulfated.

What is the molecular mechanism by which such inadvertent sulfation occurs? The long acidic CDR H3 loops of VH1-69-using antibodies imitate the N terminus of the coreceptor. As with 17b, this imitation may relate to the electrostatic character of the acidic N terminus. Electrostatic mimicry would be accomplished by Asp and Glu residues, which coupled to the natural CDR H3 preponderance for tyrosine, would combine to

produce sequences characteristic of the substrate appropriate for sulfation.

In this case, it is substrate mimicry that is the basis of the antibody selection, not affinity for the sulfotyrosine moiety itself. We would expect substrate mimicry to occur in situations where the posttranslational recognition sequence is relatively extensive, encoding elements related to the character of the epitope itself. The extensive recognition sequence of the tyrosine sulfotransferases fits this criterion, with the sulfate itself contributing to the electronegative character of the O-sulfation motif.

Conclusions. In the struggle between HIV-1 and the host immune system, the virus has a number of advantages, including an estimated evolutionary rate 10^6 times faster than the host (reviewed in ref. 36). But the immune system contains extraordinary mechanisms of adaptive recognition. Structural analyses of the few broadly neutralizing and highly potent antibodies against HIV-1 have revealed surprising features, such as a long protruding loop (b12) and variable domain swapping (2G12) (37, 38). When a panel of CD4i antibodies was investigated, the immune system did not fail to surprise, showing both highly selective V_H-gene usage and tyrosine sulfation.

The tyrosine sulfation extends the vocabulary of antigen recognition beyond the standard 20 amino acids. Tyrosine sulfation is not a rare event, occurring on an estimated 7% of mammalian proteins (31). In addition to selection mechanisms described here, our results showing that CDR H3 sulfation may be directly encoded by recombination may explain in part its prevalence among CD4i antibodies. In contrast, we found that the broadly neutralizing b12 and 2G12 showed at least 44 and 51 somatic mutations, respectively, double the average number of somatic mutations (22 ± 6) observed for other gp120-reactive antibodies (Table 1); this high number may explain in part their rarity.

The V_H-gene usage that we describe here for the CD4i antibodies not only is unusual in general but also is not observed with other gp120-reactive antibodies (Table 1). Selective V_H-gene usage has been observed previously, for example with salivary gland mucosa-associated lymphoma immunoglobulins (39), but the molecular details of selection have not been uncovered. Here we show that with the CD4i antibodies, such selective V_H-gene usage reflects the unusual properties of the CD4i epitope, being highly conserved but hidden by multiple layers of immune camouflage. While the small size of the epitope limits accessibility, it also limits the determinants of recognition. Once recognition occurs, it paradoxically allows repeated elicitation of similar antibodies. The findings suggest that, presented with the right immunogen, antibodies against the coreceptor binding site of gp120 may be readily elicited.

Our structural analysis reveals mechanisms of sulfation and V_H-gene selection. Whether either of these represents an Achilles heel on HIV-1 that can be exploited therapeutically or prophylactically remains to be seen. Perhaps we need only look to the immune system for further inspiration.

We thank R. Friedman for assistance with 17b genomic analysis; X. Ji for communication of X5 results before publication; J. Lidestri and the X4A and SER-CAT beamline staffs for assistance with data collection; D. Davies, J. Mascola, G. Ofek, and L. Shapiro for comments on the manuscript; and H. Alston for help with its preparation. These studies were supported by grants from the National Institutes of Health. P.D.K. was a recipient of a Burroughs Wellcome Career Development Award. J.S. was supported by a Bristol–Myers Squibb research grant and by the International AIDS Vaccine Initiative. Beamline X4A at the National Synchrotron Light Source, a Department of Energy facility, was supported by the Howard Hughes Medical Institute; use of the Southeast Regional Collaborative Access Team (SER-CAT) beamline at the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract W-31-109-Eng-38.

1. Profy, A. T., Salinas, P. A., Eckler, L. I., Dunlop, N. M., Nara, P. L. & Putney, S. D. (1990) *J. Immunol.* **144,** 4641–4647.
2. Wyatt, R., Kwong, P. D., Desjardins, E., Sweet, R. W., Robinson, J., Hendrickson, W. A. & Sodroski, J. (1998) *Nature* **393,** 705–711.
3. Kwong, P. D., Doyle, M. L., Casper, D. J., Cicala, C., Leavitt, S. A., Majeed, S., Steenbeke, T. D., Venturi, M., Chaikin, I., Fung, M., *et al*. (2002) *Nature* **420,** 678–682.
4. Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, G. M., Kilby, M. J., Saag, M. S., *et al*. (2003) *Nature* **422,** 307–312.
5. Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., *et al*. (1983) *Science* **220,** 868–871.
6. Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., *et al*. (1984) *Science* **224,** 500–503.
7. Dalgleish, A. G., Beverley, P. C., Clapham, P. R., Crawford, D. H., Greaves, M. F. & Weiss, R. A. (1984) *Nature* **312,** 763–767.
8. Klatzmann, D., Champagne, E., Charmaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C. & Montagnier, L. (1984) *Nature* **312,** 767–768.
9. Berger, E. A., Murphy, P. M. & Farber, J. M. (1999) *Annu. Rev. Immunol.* **17,** 657–700.
10. Atchinson, R. E., Gosling, J., Monteclaro, F. S., Franci, C., Digilio, L., Charo, I. F. & Goldsmith, M. A. (1996) *Science* **274,** 1924–1926.
11. Doranz, B. J., Lu, Z. H., Rucker, J., Zhang, T. Y., Sharron, M., Cen, Y. H., Wang, Z. X., Guo, H. H., Du, J. G., Accavitti, M. A., *et al*. (1997) *J. Virol.* **71,** 6305–6314.
12. Farzan, M., Choe, H., Martin, K. A., Sun, Y., Sideldo, M., Mackay, C. R., Gerard, N. P., Sodroski, J. & Gerard, C. (1997) *J. Biol. Chem.* **272,** 6854–6859.
13. Rucker, J., Samson, M., Doranz, B. J., Libert, F., Berson, J. F., Yi, Y., Smyth, R. J., Collman, R. G., Broder, C. C., Vassart, G., *et al*. (1996) *Cell* **87,** 437–446.
14. Farzan, M., Mirzabekov, T., Kolchinsky, P., Wyatt, R., Cayabyab, M., Gerard, N. P., Gerard, C., Sodroski, J. & Choe, H. (1999) *Cell* **96,** 667–676.
15. Thali, M., Moore, J. P., Furman, C., Charles, M., Ho, D. D., Robinson, J. & Sodroski, J. (1993) *J. Virol.* **67,** 3978–3988.
16. Labrijn, A. F., Poignard, P., Raja, A., Zwick, M. B., Delgado, K., Franti, M., Binley, J., Vivona, V., Grundner, C., Huang, C., *et al*. (2003) *J. Virol.* **77,** 10557–10565.
17. Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998) *Nature* **393,** 648–659.
18. Kwong, P. D., Wyatt, R., Majeed, S., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (2000) *Structure* **8,** 1329–1339.
19. Choe, H., Li, W., Wright, P. L., Vasilieva, N., Venturi, M., Huang, C., Grundner, C., Zwick, M. B., Wang, L., Rosenberg, E. S., *et al*. (2003) *Cell* **114,** 161–170.
20. Montefiori, D. C., Hill, T. S., Vo, H. T., Walker, B. D. & Rosenberg, E. (2001) *J. Virol.* **75,** 10200–10207.
21. Xiang, S. H., Doka, N., Choudhary, R. K., Sodroski, J. & Robinson, J. E. (2002) *AIDS Res. Hum. Retroviruses* **18,** 1207–1217.
22. Vujcic, L. K. & Quinnan, G. V., Jr. (1995) *AIDS Res. Hum. Retroviruses* **11,** 783–787.
23. Zhang, M. Y., Shu, Y., Phogat, S., Xiao, X., Cham, F., Bouma, P., Choudhary, A., Feng, Y.-R., Sanz, I., Rybak, S., *et al*. (2003) *J. Immunol. Methods* **283,** 17–25.
24. Altschul, S., Gish, W., Miller, W., Myer, E. & Lipman, D. (1990) *J. Mol. Biol.* **215,** 403–410.
25. LeFranc, M. P. (2003) *Nucleic Acids Res.* **31,** 307–310.
26. Kwong, P. D., Wyatt, R., Desjardins, E., Robinson, J., Culp, J. S., Hellmig, B. D., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1999) *J. Biol. Chem.* **274,** 4115–4123.
27. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D* **50,** 760–763.
28. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins Struct. Funct. Genet.* **11,** 281–296.
29. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165.
30. Merritt, E. A. & Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524.
31. Moore, K. L. (2003) *J. Biol. Chem.* **278,** 24243–24246.
32. Monigatti, F., Gasteiger, E., Bairoch, A. & Jung, E. (2002) *Bioinformatics* **18,** 769–770.
33. Nguyen, H. P., Seto, N. O. L., MacKenzie, C. R., Brade, L., Kosma, P., Brade, H. & Evans, S. V. (2003) *Nat. Struct. Biol.* **10,** 1019–1025.
34. Cormier, E., Persuh, M., Thompson, D. A. D., Lin, S. W., Sakmar, T. P., Olson, W. C. & Dragic, T. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5762–5767.
35. Quiocho, F. A. (1996) *Kidney Int.* **49,** 943–946.
36. Sharp, P. M., Bailes, E., Gao, F., Beer, B. E., Hirsch, V. M. & Hahn, B. H. (2000) *Biochem. Soc. Trans.* **28,** 275–282.
37. Saphire, E. O., Parren, P. W., Pantophlet, R., Zwick, M. B., Morris, G. M., Rudd, P. M., Dwek, R. A., Stanfield, R. L., Burton, D. R. & Wilson, I. A. (2001) *Science* **293,** 1155–1159.
38. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., *et al*. (2003) *Science* **300,** 2065–2071.
39. Miklos, J. A., Swerdlow, S. H. & Bahler, D. W. (2000) *Neoplasia* **95,** 3878–3884.