Enzymatic Cleavage of a CD4 Immunoadhesin Generates Crystallizable, Biologically Active Fd-like Fragments†

Steven M. Chamow,*,‡ David H. Peers,‡ Randal A. Byrn,§ Michael G. Mulkerrin,∥ Reed J. Harris,⊥ Wen-Ching Wang,# Pamela J. Bjorkman,# Daniel J. Capon,° and Avi Ashkenazi°

Departments of Recovery Process Research and Development, Molecular Biology, Protein Engineering, and Medicinal and Analytical Chemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, Division of Hematology-Oncology, Harvard Medical School, New England Deaconess Hospital, Boston, Massachusetts 02215, and Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Pasadena, California 91125

Received June 19, 1990

ABSTRACT: CD4, the cell-surface receptor for the human immunodeficiency virus (HIV), is a member of the immunoglobulin (Ig) gene superfamily. It contains four extracellular sequences homologous to Ig V_L domains. The first of these (V_1) is sufficient for binding to HIV; however, the structural basis for this binding has yet to be elucidated. While several models for the structure of Ig-like domains in CD4 have been proposed on the basis of crystal structures of Ig V_L domains, direct evidence that CD4 and V_L domains fold similarly has not been obtained. To produce individual domains of CD4 for structural studies, we used molecular fusions of such domains with Ig heavy chain (CD4 immunoadhesins), which are very efficiently expressed and secreted in mammalian cells and can be easily isolated in single-step purification with protein A. Since these fusion molecules are antibody-like homodimeric proteins, we investigated the possibility that they might be cleaved enzymatically to produce Fd-like and Fc fragments. We found that cleavage with papain releases an Fd-like fragment containing the V_1 and V_2 CD4 domains; this fragment fully retains the ability to bind to the HIV-1 envelope glycoprotein gp120 and to block HIV infection in vitro. Moreover, folding of the CD4 domains in the Fd-like fragment and in the parent immunoadhesin is indistinguishable, as indicated by circular dichroism. Spectral analysis of the Fd-like fragment suggests that secondary structure content is identical with that predicted from the known structure of Ig V_L domains; this directly supports the hypothesis that the V_1 and V_2 domains of CD4 fold similarly to Ig V_L domains. Crystals of the Fd-like fragment diffract beyond 3-Å resolution and are suitable for detailed structural analysis. The approach described here may be useful, as an alternative to direct expression, in the study of receptors and other adhesion molecules which are members of the Ig gene superfamily.

CD4 is a 55-kDa cell-surface glycoprotein, found mainly on a subset of T lymphocytes and on macrophages. The normal function of CD4 is to associate with class II major histocompatibility molecules on the surface of antigen-presenting cells, facilitating antigen recognition by the T cell receptor. CD4 also serves as the major cellular receptor for the human immunodeficiency virus (HIV) (Dalgleish et al., 1984; Klatzmann et al., 1984). Infection by HIV is initiated by binding of the viral envelope glycoprotein, gp120, to CD4. It is thought that this binding event is followed by entry of the viral genome into the target cell through fusion of the viral envelope with the cell plasma membrane [for review, see Robey and Axel (1990)].

Molecular cloning of CD4 has demonstrated that it is a member of the immunoglobulin gene superfamily (Maddon et al., 1985). The extracellular portion of CD4 contains four immunoglobulin V like domains (V_1–V_4), the first of which is necessary and sufficient for gp120 binding (Maddon et al., 1987; Richardson et al., 1988; Landau et al., 1988). Mutational analyses of the V_1 domain indicate that the amino acids which interact with gp120 are located in a sequence that is similar to an antigen complementarity-determining region of immunoglobulins (Peterson & Seed, 1988; Mizukami et al., 1988; Clayton et al., 1988; Arthos et al., 1989), as well as in several other regions of the V_1 domains (Brodsky et al., 1990). Structural models for the V_1 domain of CD4 have been proposed on the basis of its homology to immunoglobulin variable (V_L) domains, for which crystal structures have been solved (Peterson & Seed, 1988; Clayton et al., 1988; Arthos et al., 1989; Bates et al., 1989). However, a clear picture of how the V-like domains of CD4 are folded, and how their interaction with gp120 is facilitated, has not yet been achieved, in part because the X-ray crystal structure of the molecule has not been solved. To this end, recombinant soluble CD4 (Smith et al., 1987; Harris et al., 1990) has not been useful (see Results and Discussion).

Previous studies (Chao et al., 1989; Berger et al., 1988) have shown that truncated forms of CD4 can be produced by direct expression. Here we describe an alternative approach for the production of isolated CD4 domains using enzymatic cleavage of CD4–immunoglobulin fusion molecules (CD4 immunoadhesins). Several properties of these fusion molecules suggested that they might be useful for producing individual CD4 domains. First, CD4 immunoadhesins can be expressed and

† This work was supported by Genentech, Inc., by Public Health Service Grants AI-28931 (W.-C.W. and P.J.B.), HL-43510, HL-42374, and HL-42112 (R.A.B.) from the National Institutes of Health, and by the Howard Hughes Medical Institute. P.J.B. is a Pew Scholar.

* To whom correspondence should be addressed at Genentech, Inc., 460 Point San Bruno Blvd., South San Francisco, CA 94080.

‡ Department of Recovery Process Research and Development, Genentech.

§ New England Deaconess Hospital.

∥ Department of Protein Engineering, Genentech.

⊥ Department of Medicinal and Analytical Chemistry, Genentech.

California Institute of Technology.

° Department of Molecular Biology, Genentech.

0006-2960/90/0429-9885$02.50/0

secreted efficiently in mammalian cells (Byrn et al., 1990). Second, purification of such fusion molecules can be achieved under nondenaturing conditions in a single step with *Staphylococcus aureus* protein A (Capon et al., 1989). Third, CD4 immunoadhesins have an antibody-like homodimeric structure which does not contain light chains, suggesting that it might be possible to release CD4 domains as Fd[1]-like fragments by enzymatic cleavage.

The goal of the present study was to investigate whether CD4 immunoadhesins can be cleaved enzymatically, in a manner similar to that of immunoglobulins. We show that Fd-like fragments containing specific CD4 domains can be generated by cleavage of CD4 immunoadhesins with papain. Fragments so generated, containing the V_1 and V_2 domains of CD4, are biologically active, are properly folded, and form well-ordered crystals. This approach of using immunoglobulin fusion molecules, combined with enzymatic cleavage, may be generally useful in studying the structure and function of specific domains of proteins encoded by genes of the immunoglobulin gene superfamily.

Experimental Procedures

Reagents. Immunoadhesins containing the four V-like domains of CD4 ($CD4_4$-IgG) or the first two such domains ($CD4_2$-IgG) (Capon et al., 1989) were purified by affinity chromatography on immobilized protein A from supernatants of transfected Chinese hamster ovary cells. Recombinant soluble CD4 containing the extracellular portion of the molecule (rCD4) (Smith et al., 1987) was purified from supernatants of similarly transfected cells. Human IgG (hIgG) was purchased from Jackson Immunoresearch (West Grove, PA). Mercuripapain was obtained from Worthington Biochemicals (Freehold, NJ). Phenyl- and Q-Sepharose were obtained from Pharmacia (Piscataway, NJ). All other chemicals were reagent grade.

Papain Cleavage. Mercuripapain was preactivated as described (Parham, 1986). rCD4, $CD4_4$-IgG, $CD4_2$-IgG, and hIgG at concentrations of 1 mg/mL in a buffer containing 20 mM 2-(*N*-morpholino)ethanesulfonic acid, pH 6.0, 23 mM glycine, and 230 mM mannitol were digested with activated mercuripapain (1/50 w/w) for 15–60 min at 37 °C. Reactions were stopped by addition of iodoacetamide to 25 mM final concentration, and mixtures were analyzed by nonreduced sodium dodecyl sulfate–polyacrylamide gel electrophoresis in 12% gels and immunoblotting (Burnette, 1981). Blots were probed with a purified mouse monoclonal antibody (MAb 460) directed against the first domain of CD4 (4 μg/mL in Tris-buffered saline/3% bovine serum albumin). Blots were developed with horseradish peroxidase conjugated goat anti-mouse antibody (Jackson Immunoresearch, West Grove, PA) diluted 1:2000 in Tris-buffered saline/3% bovine serum albumin, followed by 4-chloro-1-naphthol/H_2O_2 as substrate. Molecular masses were estimated by comparing mobilities to migration of standard proteins (Bio-Rad, Richmond, CA). Because our interest was in a CD4 fragment of two domains, only digestion products of $CD4_2$-IgG (Figure 1A) were purified, characterized, and evaluated further.

Purification of Fd-like and Fc Fragments. Digested $CD4_2$-IgG (50 mg) was concentrated and conditioned by addition of ammonium sulfate and glycerol to final concentrations of 0.7 M and 20% (w/v), respectively. The conditioned digest was applied to a column of phenyl-Sepharose equilibrated in

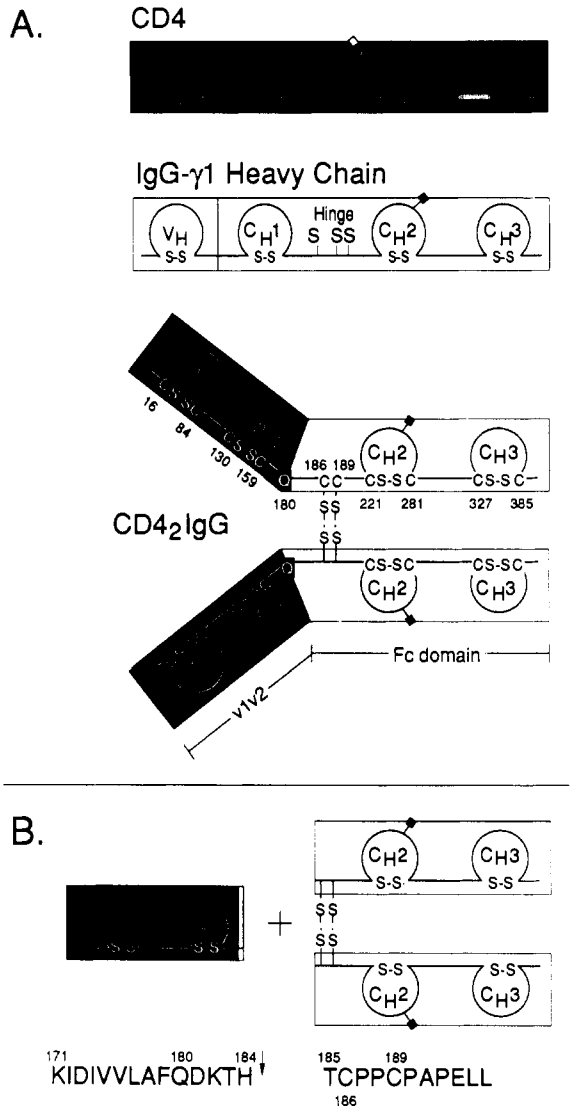

FIGURE 1: Molecular constructions and identification of papain cleavage site. (A) Structure of CD4 immunoadhesin and the parent human CD4 and IgG1 heavy chain molecules. CD4- and IgG1-derived sequences are indicated by shaded and unshaded regions, respectively. The immunoglobulin-like domains of CD4 are numbered 1–4; TM and CYT refer to the transmembrane and cytoplasmic domains. Locations of N-linked carbohydrate chains (square lollipops) are shown. The variable (V_H) and constant (C_H1, hinge, C_H2, and C_H3) regions of IgG-γ1 heavy chain are shown. $CD4_2$-IgG consists of residues 1–180 (Q) of the mature CD4 protein fused to γ1 sequence beginning at Asp-216, which is the first residue in the IgG1 hinge after the cysteine residue involved in heavy-to-light chain bonding. Cysteines are numbered and their corresponding disulfides indicated (R. J. Harris, unpublished data). $CD4_2$-IgG was purified by protein A affinity chromatography from supernatants of transfected Chinese hamster ovary cells. (B) Papain cleaves $CD4_2$-IgG in the hinge region, generating a two-chain, disulfide-linked Fc fragment and two V_1V_2 fragments per molecule. The primary site of cleavage is between His-184 and Thr-185, with secondary cleavage between Asp-181 and Lys-182.

50 mM sodium phosphate, pH 7.2, 0.7 M ammonium sulfate, and 20% (w/v) glycerol. The column flow-through fraction containing 10 mg of the Fd-like fragment (termed V_1V_2) was concentrated to 7.5 mg/mL and buffer-exchanged by gel filtration on Sephadex G-25 into 0.1 M Tris-HCl, pH 8.0, for use.

The bound fraction from the phenyl-Sepharose column containing the Fc fragment and residual $CD4_2$-IgG was eluted with 6 M urea, dialyzed against 10 mM Tris, pH 8.0/150 mM NaCl, and applied to a protein A affinity column (IPA-300, Repligen Corp., Cambridge, MA). Bound protein was eluted,

[1] Fd is the portion of an immunoglobulin Fab fragment contributed by the heavy chain.

dialyzed against 10 mM Tris, pH 8.0/25 mM NaCl, and applied to a Q-Sepharose column equilibrated in the same buffer. The flow-through fraction containing Fc fragment (24 mg) was concentrated to 1.0 mg/mL and buffer exchanged into 0.1 M Tris-HCl, pH 8.0, for use.

V_1V_2, $CD4_2$-IgG, and Fc were quantitated by absorbance at 280 nm with $\epsilon^{0.1\%}$ of 1.02, 1.34, and 1.35, respectively (R. J. Harris, unpublished data), or by bicinchoninic acid protein assay (Smith et al., 1985).

gp120 Binding. Microtiter plates were coated with goat anti-human IgG antibody (1 μg/mL) for 12 h at 4 °C. After washing and blocking with phosphate-buffered saline containing 1% bovine serum albumin, plates were incubated with $CD4_2$-IgG (1 μg/mL) for 1 h at 24 °C. The plates were washed again, then incubated (1 h, 24 °C) with ^{125}I-gp120 (5 nM), and added together with various concentrations of purified $CD4_2$-IgG, V_1V_2, or Fc to compete for the binding of ^{125}I-gp120 to the plate-bound $CD4_2$-IgG.

HIV Infectivity. Infection of H9 cells by HIV-1_{IIIB} strain (Gallo et al., 1984) was performed as previously described (Byrn et al., 1989). Reverse transcriptase (RT) activity of infected cultures was determined after incubation for 7 days (Poiesz et al., 1980).

N- and C-Terminal Sequencing. Gel-separated fragments were transferred by electroblotting onto poly(vinylidene difluoride) membrane (Matsudaira, 1987). Membrane sections containing electroblotted fragments were inserted into an Applied Biosystems Model 477A sequencer reaction cartridge atop a preconditioned polybrene-coated filter. For amino acid analysis, the two V_1V_2 fragments were resolved by reverse-phase HPLC. The peaks were hydrolyzed for 24 h in 6 N HCl at 110 °C in vacuo and then analyzed on a Beckman 6300 system.

Circular Dichroism. V_1V_2, $CD4_2$-IgG, and Fc were prepared in 0.1 M Tris-HCl, pH 8.0, as described above at concentrations of approximately 1.0 mg/mL. Circular dichroism spectra were obtained on an Aviv Cary 60 spectropolarimeter. All spectra were determined as the average of five scans with an integration time of 2 s for each datum. The far-UV CD spectra (190–250 nm) were obtained at 20 °C with an interval of 0.2 nm in a 0.01-cm cell. The near-UV CD spectra (250–350 nm) were obtained, also at 20 °C, with an interval of 0.5 nm in a 1.0-cm cell. Secondary structure content of V_1V_2 was calculated with the modified program (CONTIN V2DP, March 1984) of Provencher and Glockner (1981).

Crystallization. Crystals were grown from protein solutions (7.5 mg/mL) in 0.1 M Tris, pH 8.0, in 2-μL droplets with 20% (w/v) poly(ethylene glycol) 3350, by vapor diffusion.

Results and Discussion

The cleavage of human IgG with papain (Porter, 1959) results in the generation of two fragments with distinct properties: (i) an Fc fragment, comprised of two heavy chains held together by disulfide bonds, each consisting of a hinge region and constant domains C_H2 and C_H3; and (ii) two Fab fragments, each comprised of one heavy and one light chain, consisting respectively of a constant domain C_H1 or C_L and a variable domain V_H or V_L. Each Fab fragment contains a monovalent antigen combining site. The polypeptide chain structure of the immunoadhesin $CD4_2$-IgG used in this study is illustrated in Figure 1A. This molecule folds as a disulfide-bonded homodimer, in which each monomer consists of the V_1 and V_2 domains of human CD4 linked at its carboxyl terminus to the hinge region and constant domains C_H2 and C_H3 of a human IgG-γ1 heavy chain (Byrn et al., 1990). We

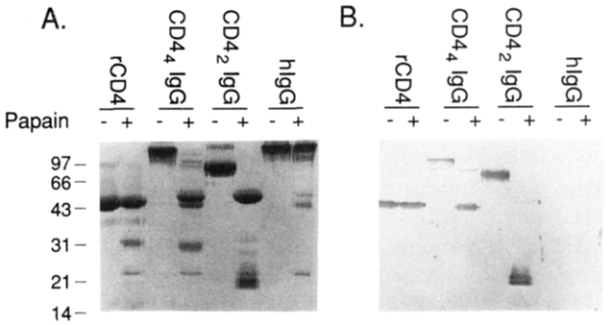

FIGURE 2: Cleavage of immunoadhesins with papain. rCD4, $CD4_4$-IgG, $CD4_2$-IgG, and hIgG were incubated without (−) or with (+) preactivated papain in the absence of cysteine for 30 min at 37 °C; iodoacetamide was added, and reaction mixtures were analyzed. (A) For SDS–PAGE, 10 μg was analyzed per lane, and the gel was stained with Coomassie blue; (B) for immunoblotting, 1 μg was used, and the immunoblot was processed with the anti-CD4 monoclonal antibody MAb 460 as described under Experimental Procedures. Papain appears in the (+) lanes of (A) as an approximately 23-kDa band.

also investigated the immunoadhesin $CD4_4$-IgG, which is similar to $CD4_2$-IgG but contains all four V-like domains of CD4.

To generate intact Fd-like fragments from these CD4 immunoadhesins, we sought conditions under which papain would efficiently cleave these hybrid molecules but not recombinant soluble CD4 (rCD4). We found that, at pH 6.0 and in the absence of cysteine, rCD4 is largely resistant to papain cleavage for at least 60 min, whereas $CD4_2$-IgG and $CD4_4$-IgG are digested virtually completely within 30 min (Figure 2A). Human IgG is also resistant under these conditions (Figure 2A). Both $CD4_2$-IgG and $CD4_4$-IgG produced a major cleavage product of 55 kDa. In addition, $CD4_2$-IgG produced a doublet of approximately 20 kDa, and $CD4_4$-IgG produced a 50-kDa fragment. Immunoblotting with a mouse monoclonal antibody (MAb 460) directed against the V_1 domain of CD4 (K. Rosenthal and B. Fendly, personal communication) indicated that this domain was present in the 20-kDa doublet generated from $CD4_2$-IgG and in the 50-kDa fragment from $CD4_4$-IgG (Figure 2B). These results indicate that the 20- and 50-kDa fragments represent the sequences contributed by the V_1V_2 or the V_1–V_4 CD4 domains, respectively, while the 55-kDa fragment seen in digests of both immunoadhesins represents the Fc fragment of the IgG1 heavy chain. Reduction of the 55-kDa fragment led to the appearance of a single 28-kDa species (not shown), further supporting the conclusion that the 55-kDa papain cleavage product represents an Fc fragment. No lower molecular mass bands appeared after reduction of the 20- and 50-kDa fragments (not shown), indicating that nicks in the polypeptide chains had not occurred. In addition to the major fragments, a series of minor fragments ranging in size from 20 to 35 kDa were produced by papain digestion, and their nature and abundance varied with the pH of the reaction in the range 5.0–7.0. Digestion at pH 6.0 gave the best yield of the V_1V_2 doublet.

Papain is a thiol protease and, therefore, can be activated by free cysteine (Parham, 1986); generation of Fab fragments from immunoglobulins requires preactivation of the enzyme with cysteine, as well as the presence of cysteine in the reaction mixture (Poulik, 1966). Indeed, omission of cysteine from the reaction mixture resulted in inefficient cleavage of human IgG by papain (Figure 2A). Production of Fd-like fragments from $Cd4_2$-IgG and $CD4_4$-IgG by papain also required preactivation of the enzyme with cysteine. However, in contrast to human IgG, cysteine was not required in the reaction mixture for

efficient production of Fd-like fragments from these immunoadhesins (Figure 2A). In fact, inclusion of cysteine in the reaction mixture destroyed the Fd-like fragments as a result of multiple-site cleavage in portions of the molecule not restricted to the hinge region (not shown). Because papain does not require the continuous presence of cysteine in order to cleave the immunoadhesins, it may be more active in cleaving these molecules than in cleaving immunoglobulins. One possible explanation is that the hinge region, in which the site of cleavage is located (see below), is more accessible to the enzyme in the former molecule than in the latter. This may result from the fact that these immunoadhesins do not contain covalently attached light chains. In this context, it is interesting to note that a different immunoadhesin construct ($CD4_2$-γ1) (Capon et al., 1989), which is also homodimeric but includes all three C_H domains of human IgG, cleaved only poorly with papain under similar conditions (not shown). This suggests that removal of the C_H1 domain also increases the accessibility of the hinge region to cleavage by papain. A second enzyme, pepsin, was not useful in generating isolated domains from CD4 immunoadhesins, because the CD4 portion itself was not stable to treatment with the enzyme (not shown).

To identify the site of cleavage by papain, gel-separated fragments were transferred to poly(vinylidene difluoride) membrane and subjected to N-terminal sequence analysis (Figure 1B). The V_1V_2 fragments gave the common sequence KKVVLG, corresponding to the N-terminus of the intact molecule (Littman et al., 1988; Maddon et al., 1985). On the basis of the N-terminal sequence of the Fc fragment (TCPPCP), we deduced that cleavage had occurred between His-184 and Thr-185, immediately upstream of the interchain disulfide bonds in the IgG hinge region of the molecule. This site corresponds exactly to the cleavage site identified for human IgG (Steiner & Porter, 1967; Piggot & Press, 1967).

Two forms of V_1V_2 that differ by approximately 1 kDa in apparent molecular mass (Figure 2) were consistently observed. Reverse-phase HPLC was employed to resolve these two forms (not shown). Amino acid and mass spectrometric analysis of the collected fragments demonstrated that the higher molecular mass form (representing 70% of total V_1V_2) is comprised of residues 1–184, while the lower molecular mass form is three residues shorter (1–181). Since 1–180 is the CD4 coding sequence, the two forms of V_1V_2 contain four- and one-residue C-terminal γ1 extensions, respectively. However, Fc fragments that include residues 182–184 were not found. These results suggest the following sequence of cleavage events: the 1–184 form of V_1V_2 is generated first by cleavage between His-184 and Thr-185; subsequently, the 1–181 form is produced by secondary papain cleavage between Asp-181 and Lys-182. Secondary cleavages are also known to occur with human IgG (Gorevic et al., 1985). These two forms of V_1V_2 cocrystallized and were otherwise indistinguishable (see below).

To investigate whether the V_1V_2 fragments retain the biological and structural characteristics exhibited by these same domains in the parent immunoadhesin, we purified the V_1V_2 and Fc fragments (see Experimental Procedures) and tested the ability of the V_1V_2 fragment to bind to gp120. The ability of the fragments to compete with immobilized $CD4_2$-IgG for binding to ^{125}I-labeled gp120 was measured and compared to that of free $CD4_2$-IgG (Figure 3a). Whereas the purified Fc fragment was unable to bind to gp120, competition by purified V_1V_2 was virtually identical with that of $CD4_2$-IgG (Figure 3a).

To characterize further the biological activity of the V_1V_2 fragment, we compared the potency of this polypeptide with

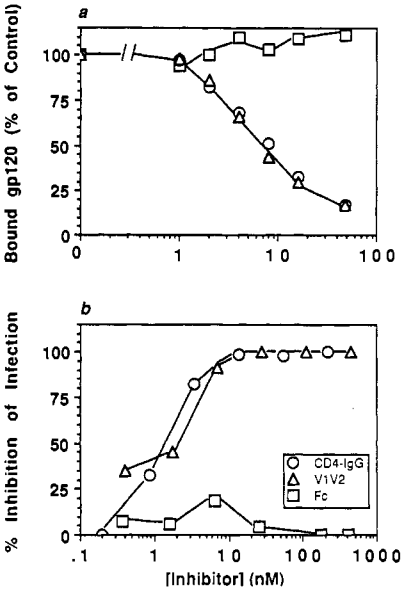

FIGURE 3: gp120 binding and inhibition of HIV infectivity by V_1V_2. (a) Binding of $CD4_2$-IgG, V_1V_2, and Fc to gp120. $CD4_2$-IgG was bound to plates coated with anti-human IgG antibody. Subsequently, ^{125}I-gp120 (5 nM) was added, together with purified $CD4_2$-IgG (circles), V_1V_2 (triangles), or Fc (boxes), as the displacer. The K_d for ^{125}I-gp120 binding to plate-bound $CD4_2$-IgG was 3.6 nM, as measured independently by saturation binding analysis (not shown). The K_d for gp120 binding to soluble $CD4_2$-IgG or to V_1V_2 (K_d displacer), as derived from the competition data and with the equation K_d displacer = $IC_{50}/(1 + [gp120]/K_d)$, is 4.1 nM. (b) Inhibition of HIV infectivity. H9 cell cultures were infected by HIV-1_{IIIB} in the presence of $CD4_2$-IgG (circles), V_1V_2 (triangles), or Fc (boxes). Reverse transcriptase activity was determined after incubation of cultures for 7 days at 37 °C. Percent inhibition is defined as [1 − (average of duplicate experimental culture RT values)/(average of replicate media control culture RT values)] × 100.

that of $CD4_2$-IgG in blocking the infection of H9 lymphoblastoid cells by HIV-1_{IIIB} in vitro (Figure 3b). Both $CD4_2$-IgG and V_1V_2 blocked infection, essentially completely, at concentrations greater than 10 nM. Half-maximal inhibition occurred at approximately 1.5 and 2.5 nM, respectively. Purified Fc, prepared from $CD4_2$-IgG, had no significant inhibitory activity in this assay. None of the three agents had any adverse effect on cell viability over the concentration range used. The observation that V_1V_2 and $CD4_2$-IgG are equivalent both in binding to gp120 and in blocking HIV infection of H9 cells suggested that at least the CD4 V_1 domain, which is known to be sufficient for these functions, retains its native conformation in this papain-generated fragment.

To compare more directly the structure of the isolated V_1V_2 domains to that in the parent $CD4_2$-IgG molecule, we characterized these molecules by circular dichroism (Figure 4). The contributions of the individual spectra of V_1V_2 and Fc were summed and compared to the spectrum of the parent molecule. Differences in the far-UV CD spectra (190–250 nm) would suggest changes in secondary structural elements; differences in the near-UV CD spectra (250–350 nm), changes in chromophore environments. A comparison of the far-UV CD spectra (Figure 4A) indicates that the two spectra overlap in the region above 205 nm. Thus, the total number of residues in each of the secondary structural elements in the fragments is unchanged relative to that in the parent molecule. At shorter wavelengths, there is some diminution in the intensity of the immunoadhesin spectrum compared to that of the sum of the spectra of the fragments. This difference may reflect the juxtaposition of stacked β-sheets from different domains of the parent molecule that is absent in separated fragments;

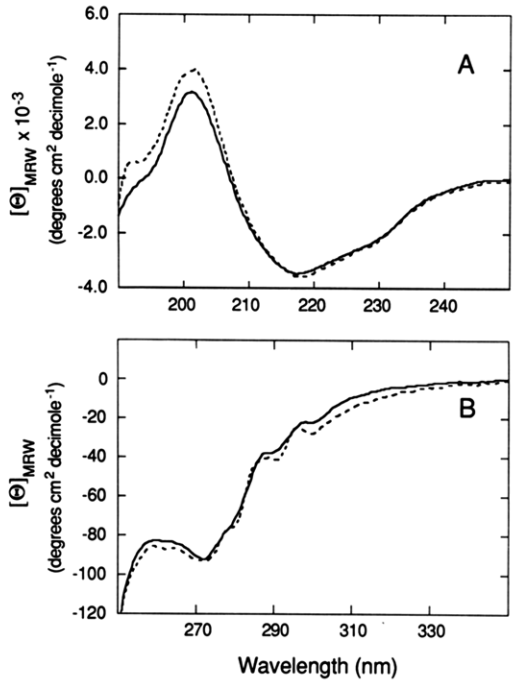

FIGURE 4: No structural changes in V_1V_2 or Fc domains are indicated by circular dichroism as a result of cleavage. The individual spectra of V_1V_2 + Fc were summed (—) and compared to that of intact $CD4_2$-IgG (– – –). (A) Far-UV CD (190–250 nm) and (B) near-UV CD (250–350 nm).

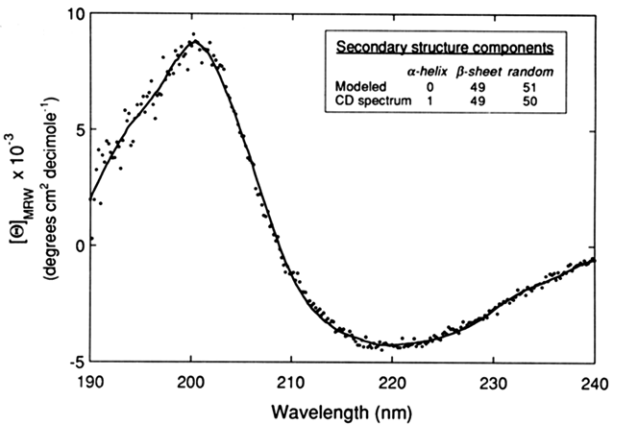

FIGURE 5: Secondary structure calculation of V_1 and V_2 domains. The far-UV CD spectrum of V_1V_2 (●) and the fit of the data (—) from the secondary structure calculation. The table insert shows the secondary structure content (%) predicted from modeling (Bates et al., 1989) and the result of the analysis from the CD spectrum for V_1V_2. The differences in α-helix and random coil components are not statistically significant. The spectrum consisting of 251 data points was used.

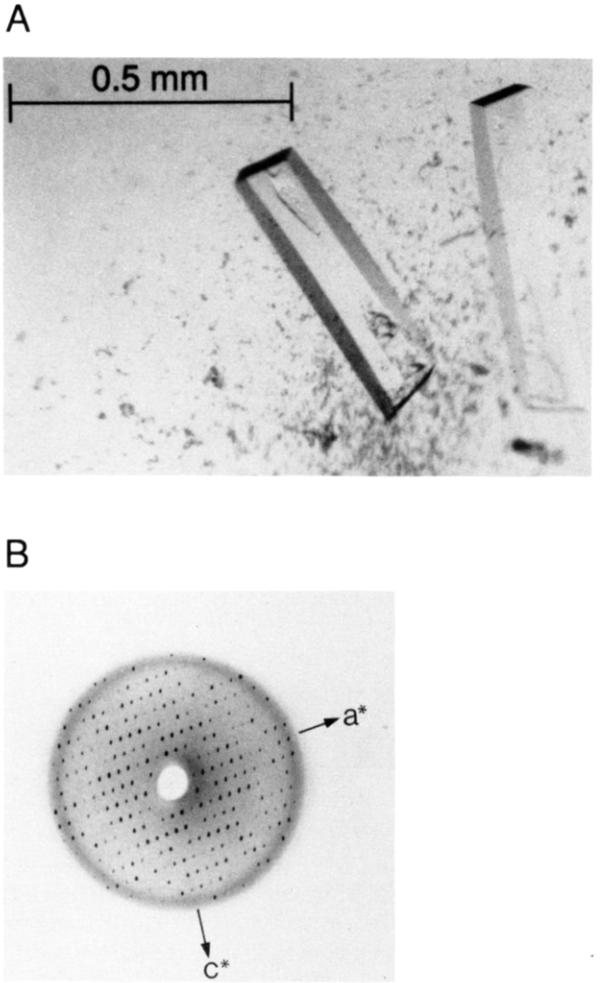

FIGURE 6: Crystals of purified V_1V_2 fragment. (A) Crystals were generated from a 7.5 mg/mL protein solution by the hanging-drop method. Length of crystals is indicated by the bar. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis of washed crystals demonstrated that both V_1V_2 fragments were present (not shown). (B) An 8° screened precession photograph (*h0l* zone) taken with nickel-filtered, CuKα radiation (50 kV, 150 mA) at a crystal-to-film distance of 10 cm.

interaction between β-sheets, or a change in their relative orientation, is known to perturb the circular dichroism spectrum (Manning & Woody, 1987; Manning et al., 1988). That the V_1V_2 CD4 domains are structurally unchanged is further indicated by near-UV CD data (Figure 4B). Sensitive to the environments of individual chromophores in the molecule, these spectra reflect differences in the packing of secondary structural elements. As before, the sum of the spectra of the individual fragments overlaps with the spectrum of the intact immunoadhesin. Taken together, these results demonstrate structural equivalence of the V_1V_2 and Fc domains in the fragments and in the parent $CD4_2$-IgG molecule, suggesting that papain cleavage in the hinge has no effect on the three-dimensional structure of the individual domains.

For a more detailed structural analysis of the V_1V_2 domains of CD4, we calculated the amount of each of the protein secondary structural elements represented in the far-UV CD spectrum of V_1V_2 (Figure 5). Our analysis suggests a distribution of 1% α-helix, 49% β-sheet, and 50% random. A recent model for the structure of the V_1V_2 domains based on homology to solved crystal structures of immunoglobulin variable domains (Bates et al., 1989) predicts that 87 of a total of 178 residues (49%) are involved in β-sheet, with the remaining 91 residues (51%) in loop structures. The result of our analysis is consistent with prediction (see Figure 5 inset). These data therefore support the hypothesis that the V_1V_2 domains of CD4 possess the secondary structural elements of immunoglobulin variable domains.

Many previous attempts have been made to crystallize CD4. Purified rCD4, one form of the molecule that is available in sufficient quantity, has been crystallized; however, these crystals are disordered and diffract to only 10-Å resolution, possibly due to glycosylation (M. Ultsch and S. M. Chamow, unpublished data). In contrast, the purified V_1V_2 fragments described here were found to produce highly ordered crystals which diffract beyond 3-Å resolution. The V_1V_2 crystals are needle-like, with approximate dimensions of 0.5 mm × 0.1 mm × 0.1 mm (Figure 6). Screened precession photographs show reciprocal lattice symmetry *C*2, such that only reflections for which $h + k = 2n$ are present. The space group is therefore

*C*2 with unit-cell dimensions $a = 134.5$, $b = 33.1$, and $c = 45.6$ Å and $\beta = 96°$. The asymmetric unit of the crystal is estimated to contain one molecule, on the basis of average volume-to-mass ratios of protein crystals ($Vm = 2.5$) (Mathews, 1968).

Truncated forms of CD4 or CD4 domains have been previously produced in bacteria (Chao et al., 1989) and in mammalian cells (Berger et al., 1988) by direct expression of truncated CD4 cDNA. However, the approach of using immunoglobulin fusion molecules to generate isolated domains of CD4 might have several advantages over direct-expression strategies. First, CD4 immunoadhesins are generally secreted very efficiently (levels range from 0.5 to 2 μg/mL) and are stable in cell supernatants (D. J. Capon and A. Ashkenazi, unpublished data). Second, purification in nondenaturing solvents can be achieved in a single step with *S. aureus* protein A, with quantitative recovery. Since protein A is a group-specific adsorbent (Lindmark et al., 1983), it can be used for many types of IgG-derived constructs; generation of Fd-like fragments from immunoadhesins purified in this way is straightforward, provided that the constructs are not susceptible to intradomain proteolysis. This situation is favored by a cleavage condition which does not include cysteine. Third, this approach is easily scaled to enable preparation of milligram quantities of purified fragments, a prerequisite of crystallization.

In conclusion, we have described a novel method for the preparation of Fd-like fragments of CD4 from a CD4–immunoglobulin fusion molecule by enzymatic cleavage with papain. The CD4 domains in these fragments retain their biological functions of binding to gp120 and of blocking HIV infection and appear to be folded in their native conformation. These fragments are crystallizable and give rise to well-ordered crystals which can be structurally analyzed at atomic resolution. The approach described here may be applicable to the study of other receptors and adhesion molecules. Indeed, a similar approach has recently been used successfully to generate isolated domains of lymphocyte homing receptor for functional and structural analysis (S. Watson, personal communication).

ACKNOWLEDGMENTS

We thank Drs. P. Griffin and J. Stults for mass spectrometric analysis, Dr. T. Gregory for advice and helpful discussions, Dr. R. Ward for critical reading of the manuscript, and K. Andow for graphics assistance. The anti-CD4 monoclonal antibody (MAb 460) was a generous gift from J. Porter and Dr. B. Fendly.

REFERENCES

Arthos, J., Deen, K. C., Chaikin, M. A., Fornwald, J. A., Sathe, G., Sattentau, Q. J., Clapham, P. R., Weiss, R. A., McDougal, J. S., Pietropaolo, C., Axel, R., Truneh, A., Maddon, P. J., & Sweet, R. W. (1989) *Cell 57*, 469–481.

Bates, P. A., McGregor, M. J., Islam, S. A., Sattentau, Q. J., & Sternberg, M. J. E. (1989) *Protein Eng. 3*, 13–21.

Berger, E. A., Fuerst, T. R., & Moss, B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2357–2361.

Brodsky, M. H., Warton, M., Myers, R. M., & Littman, D. (1990) *J. Immunol. 144*, 3078–3086.

Burnette, W. N. (1981) *Anal. Biochem. 112*, 195–203.

Byrn, R. A., Sekigawa, I., Chamow, S. M., Johnson, J. S., Gregory, T. J., Capon, D. J., & Groopman, J. E. (1989) *J. Virol. 63*, 4370–4375.

Bryn, R. A., Mordenti, J., Lucas, C., Smith, D., Marsters, S. A., Johnson, J. S., Cossum, P., Chamow, S. M., Wurm, F. M., Gregory, T., Groopman, J. E., & Capon, D. J. (1990) *Nature 344*, 667–670.

Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E., Broder, S., & Smith, D. H. (1989) *Nature 337*, 525–531.

Chao, B. H., Costopoulos, D. S., Curiel, T., Bertonis, J. M., Chisholm, P., Williams, C., Schooley, R. T., Rosa, J. J., Fisher, R. A., & Maraganore, J. M. (1989) *J. Biol. Chem. 264*, 5812–5817.

Clayton, L. K., Hussey, R. E., Steinbrich, R., Ramachandran, H., Husain, Y., & Reinherz, E. L. (1988) *Nature 335*, 363–366.

Dalgleish, A. G., Beverly, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., & Weiss, R. A. (1984) *Nature 312*, 763–766.

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., White, G., Foster, P., & Markham, P. D. (1984) *Science 224*, 500–503.

Gorevic, P. D., Prelli, F. C., & Frangione, B. (1985) *Methods Enzymol. 116*, 3–25.

Harris, R. J., Chamow, S. M., Gregory, T. J., & Spellman, M. W. (1990) *Eur. J. Biochem. 188*, 291–300.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J. C., & Montagnier, L. (1984) *Nature 312*, 767–768.

Lammare, D., Ashkenazi, A., Fleury, S., Smith, D. H., Sekaly, R. P., & Capon, D. J. (1989) *Science 245*, 743–746.

Landau, N., Warton, M., & Littman, D. (1988) *Nature 334*, 159–162.

Lindmark, R., Thoren-Tolling, K., & Sjoquist, J. (1983) *J. Immunol. Methods 62*, 1–13.

Littman, D. R., Maddon, P. J., & Axel, R. (1988) *Cell 55*, 541.

Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L., & Axel, R. (1985) *Cell 42*, 93–104.

Maddon, P. J., Molineaux, S. M., Maddon, D. E., Zimmerman, K. A., Godfrey, M., Alt, F. W., Chess, L., & Axel, R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 9155–9159.

Manning, M. C., & Woody, R. W. (1987) *Biopolymers 26*, 1731–1752.

Manning, M. C., Illangasekare, M., & Woody, R. W. (1988) *Biophys. Chem. 31*, 77–86.

Mathews, B. W. (1968) *J. Mol. Biol. 33*, 491–497.

Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035–10038.

Mizukami, T., Fuerst, T., Berger, E., & Moss, B. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 9273–9277.

Parham, P. (1986) in *Handbook of Experimental Immunology*, 4th ed. (Weir, D. M., Herzenberg, L. A., Blackwell, C., & Herzenberg, L. A., Eds.) Vol. 1, p 14.8, Blackwell Scientific Publications, Oxford, U.K.

Peterson, A., & Seed, B. (1988) *Cell 54*, 65–72.

Piggot, P. J., & Press, E. M. (1967) *Biochem. J. 104*, 616–626.

Poiesz, B. J., Ruscetti, F. W., Gadzer, A. F., Bunn, P. A., Minna, J. D., & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 7415–7419.

Porter, R. R. (1959) *Biochem. J. 73*, 119–126.

Poulik, M. D. (1966) *Nature 210*, 133–134.

Provencher, S. W., & Glockner, J. (1981) *Biochemistry 20*, 33–37.

Richardson, N. E., Brown, N. R., Hussey, R. E., Vaid, A., Matthews, T. J., Bolognesi, D. P., & Reinherz, E. L. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6102–6106.

Robey, E., & Axel, R. (1990) *Cell 60*, 697-700.
Smith, D. H., Byrn, R. A., Marsters, S. A., Gregory, T., Groopman, J. E., & Capon, D. J. (1987) *Science 238*, 1704-1707.
Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., & Klenk, D. C. (1985) *Anal. Biochem. 150*, 76-85.
Steiner, L. A., & Porter, R. R. (1967) *Biochemistry 6*, 3957-3970.

Incorporation of a Complete Set of Deoxyadenosine and Thymidine Analogues Suitable for the Study of Protein Nucleic Acid Interactions into Oligodeoxynucleotides. Application to the *Eco*RV Restriction Endonuclease and Modification Methylase†

Patrick C. Newman,‡ Victor U. Nwosu,‡,§ David M. Williams,‖,⊥ Richard Cosstick,‖ Frank Seela,# and Bernard A. Connolly*,‡

Department of Biochemistry (SERC Molecular Recognition Centre), University of Southampton, Southampton SO9 3TU, U.K., Department of Chemistry, University of Liverpool, Liverpool L69 3BX, U.K., and Laboratorium fur Organische and Bioorganische Chemie, Universitat Osnabruck, D-4500 Osnabruck, West Germany

Received April 11, 1990; Revised Manuscript Received July 20, 1990

ABSTRACT: A complete set of dA and T analogues designed for the study of protein DNA interactions has been prepared. These modified bases have been designed by considering the groups on the dA and T bases that are accessible to proteins when these bases are incorporated into double-helical B-DNA [Seeman, N. C., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804-808]. Each of the positions on the two bases, having the potential to interact with proteins, have been subject to nondisruptive, conservative change. Typically a particular group (e.g., the 6-NH_2 of dA or the 5-CH_3 of T) has been replaced with a hydrogen atom. Occasionally keto groups (the 2- and 4-keto oxygen atoms of T) have been replaced with sulfur. The base set has been incorporated into the self-complementary dodecamer d(GACGATATCGTC) at the central d(ATAT) sequence. Melting temperature determination shows that the modified bases do not destabilize the double helix. Additionally, circular dichroism spectroscopy shows that almost all the altered bases have very little effect on overall oligodeoxynucleotide conformation and that most of the modified oligomers have a B-DNA type structure. d(GATATC) is the recognition sequence for the *Eco*RV restriction modification system. Initial rate measurements (at a single oligodeoxynucleotide concentration of 20 μM) have been carried out with both the *Eco*RV restriction endonuclease and modification methylase. This has enabled a preliminary identification of the groups of the dA and T bases within the d(GATATC) sequence that make important contacts to both proteins.

Many proteins show a very high specificity of interaction with particular DNA sequences. Two classes of proteins have been extensively studied and are responsible for many of the ideas concerning the mechanisms by which this high specificity arises. These are the repressor proteins that bind to DNA sequences between 14 and 17 base pairs in length (Ohlendorf & Matthews, 1983; Pabo & Sauer, 1984; Brennan & Matthews, 1989) and the type II restriction enzymes that cleave DNA at recognition sequences usually 4 or 6 base pairs long (Modrich, 1982; Chirikjian, 1987; Bennett & Halford, 1989). Results obtained with both groups have led to two general models to explain how specific protein DNA interactions occur. The first has been termed direct readout, and here proteins interact directly with the bases in DNA (Otwinowski et al., 1988; Matthews, 1988; Brennan & Matthews, 1989). At its simplest this model assumes that the overall B-DNA geometry does not change with base sequence and that specificity arises from a complementary binding of a protein to a unique set of functional groups that may only be presented by its cognate sequence. The direct interaction of proteins with bases has been observed for several repressor proteins (Anderson et al., 1987; Jordan & Pabo, 1988; Aggarwal et al., 1988; Wohlberger et al., 1988) and the *Eco*RI restriction endonuclease (McClarin et al., 1986) by using X-ray crystallography. The second model is called indirect readout. Here a particular sequence of bases changes the conformation of the DNA, for example, by altering the relative dispositions of the phosphate

† B.A.C. was supported by grants from the U.K. SERC and particularly by the setting up of a molecular recognition initiative centre by this body. Additional funds were contributed by the EEC (Science Plan). B.A.C. is a Lister Institute research fellow. R.C. was supported by the U.K. SERC. P.C.N. and D.M.W. were SERC-supported research students.

* Address correspondence to this author.
‡ University of Southampton.
§ Present address: European Molecular Biology Laboratory, D-6900 Heidelberg, West Germany.
‖ University of Liverpool.
⊥ Present address: Max-Planck-Institut fur experimentelle Medizin, D-3400 Göttingen, West Germany.
Universitat Osnabruck.

0006-2960/90/0429-9891$02.50/0