

Adenovirus interaction with its cellular receptor CAR

Jason Howitt, Carl W. Anderson, and Paul Freimuth

Biology Department
Brookhaven National Laboratory
Upton, NY 11973

Contents

1. Introduction
2. Crystal Structures
 - 2.1. Ad5 knob
 - 2.2. Ad12 knob-CAR D1 complexes
 - 2.3. Ad2 and Ad3 knob
3. Knob-CAR D1 interface
 - 3.1. Binding studies and mutational analysis
 - 3.2. Modification of binding specificity--gene therapy
4. Structure of full-length CAR
5. Characterization of the human CAR gene, *CXADR*
 - 5.1. Chromosomal location of *CXADR*
 - 5.2. Sequence analysis of *CXADR*
 - 5.3. *CXADR* polymorphism analysis
6. Concluding remarks

1. Introduction

The mechanism of adenovirus attachment to the host cell plasma membrane has been revealed in detail by research over the past 10 years. It has long been known that receptor binding activity is associated with the viral fibers, trimeric spike proteins that protrude radially from the vertices of the icosahedral capsid (Philipson et al. 1968). In some adenovirus serotypes, fiber and other virus structural proteins are synthesized in excess and accumulate in the cell nucleus during late stages of infection. Fiber protein can be readily purified from lysates of cells infected with subgroup C viruses, for example Ad2 and Ad5 (Boulanger and Puvion 1973). Addition of purified fiber protein to virus suspensions during adsorption strongly inhibits infection, indicating that fiber and intact virus particles compete for binding sites on host cells (Philipson et al. 1968; Hautala et al. 1998). Cell binding studies using purified radiolabeled fiber demonstrated that fiber binds specifically and with high affinity to the cell plasma membrane, and that cell lines typically used for laboratory propagation of adenovirus have approximately 10^4 high-affinity receptor sites per cell (Persson et al. 1985; Freimuth 1996). Similar numbers of high-affinity binding sites for radiolabeled intact virus particles also were observed (Seth et al. 1994).

Adenovirus fibers have two distinct structural domains, a rod-like shaft which is attached to the capsid vertex by an amino-terminal anchor, and a distal globular domain referred to as the knob or head domain. The shaft domain consists of repetitive sequence motifs that vary in number between different serotypes and fold into a novel coiled-coil type structure (Stouten et al. 1992; van Raaij et al. 1999). The domain boundary between the shaft and knob is, therefore, easy to recognize in alignments of fiber protein sequences from different serotypes. Knob domains from several different serotypes have been produced as recombinant protein fragments in insect cell and bacterial expression systems, and this led to the key observation that receptor binding activity is associated with the knob domain (Henry et al. 1994; Louis et al. 1994). Recombinant knobs from some serotypes assemble into trimers in heterologous expression systems. The trimeric form of knob is necessary for its receptor binding activity. Recombinant full-length fiber also has been produced, using a vaccinia virus vector for expression in mammalian cells (Hong and Engler 1996). Mutant analysis demonstrated that the knob domain plays a

critical role in trimerization of the full-length fiber protein. Trimeric fiber protein is resistant to denaturation by sodium dodecylsulfate at room temperature (Hong and Engler 1996), indicating that the knob and shaft domains have highly stable protein folds.

CAR, the cellular receptor recognized by fibers of the adenoviruses that are commonly used in the laboratory (Ad2, Ad5), was recently identified and molecularly cloned (Bergelson et al. 1997; Tomko et al. 1997). As its name indicates, CAR serves as a receptor for both subgroup B coxsackieviruses and adenovirus. Many adenovirus serotypes recognize CAR with the notable exception of subgroup B viruses (Ads 3 and 7) (Defer et al. 1990; Roelvink et al. 1998). CAR has a mass of 46 kD and is a member of the immunoglobulin superfamily. The amino-terminal extracellular region of CAR consists of 2 Ig-like domains, a distal Ig variable-type domain (CAR D1) and a proximal C2-type Ig domain (CAR D2). CAR has a single hydrophobic membrane-spanning region and a ~110 residue carboxy-terminal cytoplasmic domain. CAR D1 and the entire CAR ectodomain have been expressed as recombinant protein fragments in both bacterial and mammalian cell expression systems. Ad12 knob and CAR D1 that were produced in *E. coli* formed specific complexes *in vitro*, and both bacterially-expressed proteins inhibited infection of HeLa cells by Ad2 and Ad12 virus (Freimuth et al. 1999). Thus the CAR D1 domain alone is sufficient for interaction with fiber knob. A GPI-linked form of CAR also supported adenovirus infection (Wang and Bergelson 1999), indicating that the membrane-spanning and cytoplasmic domains of CAR are not required for infection of cultured cells. From these results it can be surmised that the knob-CAR interaction serves mainly to attach virus particles to the host cell plasma membrane. Subsequent endocytosis of bound particles can be accelerated in certain cell types by interaction of virus-CAR complexes with integrin coreceptors (Bai et al. 1993; Wickham et al. 1993). Integrin-independent pathways for infection do exist, however, since mutant viruses that have lost the ability to bind to integrins do infect cells with wild-type efficiency, albeit at a slower rate on tightly adherent cells (Bai et al. 1993). The role of integrin coreceptors in adenovirus infection is beyond the scope of this discussion but has been reviewed in detail elsewhere (Nemerow 2000).

The objective of this chapter is to review the structural biology and biochemistry of the knob-CAR interaction. To date, knob-CAR and HIV gp120-CD4 are the only

examples where the interface between virus particles and their cellular protein receptors have been examined by X-ray crystallography at high resolution (Kwong et al. 1998; Bewley et al. 1999). These 2 systems have certain striking similarities and unusual characteristics which may reflect the struggle that all animal viruses have in conserving specificity for a particular receptor while simultaneously varying the antigenic structure of the receptor-binding protein itself. The relatively large number of adenovirus serotypes that bind CAR and the ease with which recombinant knob-CAR systems can be derived from these serotypes, therefore provide excellent model systems to investigate how receptor specificity is maintained during virus evolution, and conversely, how viruses manage to change receptor specificity.

2. Crystal Structures

2.1 Ad5 knob

The adenovirus-5 (Ad5) fiber protein knob domain was the first knob to have its atomic resolution structure solved by X-ray crystallography (Xia et al. 1994). For this analysis, the Ad5 knob was expressed as a soluble protein fragment in *E. coli*. Knob was found to have an eight-stranded β -sandwich structure containing two distinct antiparallel β -sheets that accommodate a high number of reverse turns with nearly all hydrogen-bond donors and acceptors found to be properly paired. The majority of the protein (65%) is made up of surface loops with approximately 35% of the structure containing β -strands. The knob fold was found to be similar to other structures solved at that time, the closest being superoxide dismutase (Tainer et al. 1982); however, the folding topologies of knob differed from these other structures. Knob, therefore, represents a new class of antiparallel β -structure.

The nomenclature adopted by Xia *et al* labeled the two β -sheets R and V, for receptor and virion, respectively. The R sheet is made up of β -strands D, G, H and I, which mainly are exposed to solvent. The V sheet is made up of the β -strands A, B, C and J and is partially buried in the trimer structure. Two very small strands, labeled E and F, also occur within the V sheet (see Figure 1A for details). Loop regions are labeled according to the strands that they connect. An angle of ~ 30 degrees separates the two β -sheets creating a hydrophobic core within the molecule. Previously, it had been

postulated that the fiber protein might be a dimer; however, stoichiometric analysis of radiolabeled protein indicated that the fiber knob was trimeric (van Oostrum and Burnett 1985). In agreement with this geometry, Ad5 knob monomers were found to form homotrimers in the crystal lattice (Figure 1B) and in solution (Xia et al. 1994). The trimer is approximately 62 Å in diameter by 40 Å high and contains extensive contacts between monomers which involve hydrogen bonding, salt bridges, van der Waals contacts and water-mediated charge interactions. Strands G and J of each monomer play a critical role in the formation of the trimeric protein with a loss of accessible surface area upon trimerization of $\sim 1950 \text{ \AA}^2$.

The cellular receptor for adenovirus had not been identified at the time the Ad5 knob structure was solved, and it was difficult to predict the location of the receptor-binding site(s) based on sequence conservation given the extensive sequence variation between knobs from different adenovirus serotypes. It was postulated, however, that a large cavity located along the central threefold axis of symmetry might accommodate binding of a single receptor molecule. Also suggested as a possible receptor binding site was the floor of the valley formed by the R-sheet and the HI loop, features that are present in 3 copies per knob trimer. According to this second model, knob would have a capacity to bind three receptor molecules.

2.2. Crystal structure of the Ad12 knob-CAR D1 complex.

The mode of knob-CAR interaction was clarified in 1999 when atomic resolution structures of the adenovirus-12 (Ad12) knob alone and in complex with CAR D1 were determined by X-ray crystallography (Bewley et al. 1999). Ad12 is a subgroup A adenovirus, and by this time it had been shown that representatives of all subgroups except subgroup B bound to CAR (Defer et al. 1990; Bergelson et al. 1997; Tomko et al. 1997; Roelvink et al. 1998). Both Ad12 knob and CAR D1 were expressed as a soluble protein fragments in *E. coli*. The structures of Ad12 and Ad5 knob trimers are very similar, with a root mean deviation of 1.2 Å when equivalent C α atoms are superimposed. The only major difference between the structures centers around the HI loop, which is well-ordered in Ad12 knob but extended and disordered in Ad5 knob. CAR D1 has a β -sandwich fold that is characteristic of immunoglobulin variable (IgV) domains, with an

insertion of 2 short β strands, C' and C'', between strands C and D (Figure 2A). Surprisingly, the binding site for CAR D1 was not located in the floor of the R sheets as predicted from structural analysis of the Ad5 knob, but rather were located at the interface between adjacent knob monomers (Figure 2B). All 3 receptor binding sites were fully saturated with CAR D1 molecules in the Ad12 knob-CAR D1 crystal structure, indicating that individual fibers on the virus capsid potentially can make multivalent contact with 2 or 3 CAR molecules on the host cell surface. The receptor binding site is made up of four loop regions from the knob trimer interacting with a single surface of CAR D1. The AB loop in particular contributes over 50 % of the interfacial protein-protein interactions (discussed in section 3). Mutational analysis of fiber knob and CAR D1 confirmed the structure of the bound complex (Bewley et al. 1999; Roelvink et al. 1999). No large conformational rearrangements of knob are required for receptor binding, but small backbone movements are observed within contact regions of CAR D1.

CAR D1 also was crystallized independently of fiber knob, and its structure was solved at high resolution (van Raaij et al. 2000). In this study CAR D1 was expressed as a soluble protein within the periplasm of *E. coli*. The structure shows CAR D1 to be a homodimer in the crystal lattice; solution based ultracentrifugation experiments also indicated a dimeric form of the protein with a dissociation constant of 16 μ M. These findings are consistent with results suggesting that CAR may be a homophilic adhesion molecule likely to mediate tissue organization as its native function (Honda et al. 2000). The high resolution structure of the dimeric CAR D1 is almost identical to the structure of CAR D1 in complex with knob; the BC and FG loops, which are involved in homodimerization and knob binding, show minor rearrangements in conformation. However, a larger difference is observed in the C'D loop, which adopts a different backbone conformation resulting in up to 4 Å movement of this loop relative to its position in the knob-CAR complex. Interestingly, the CAR D1 surface involved in the formation of the homodimer is very similar to that used for fiber knob binding, suggesting this surface may need to be conserved for proper CAR function in normal, uninfected cells. This perhaps accounts for the low frequency of single nucleotide polymorphisms that map to this region of the CAR gene (discussed below).

2.3. Crystal structures of Ad2 and Ad3 knob

The crystal structure of knob from Ad2, another subgroup C virus, was solved in 1999 (van Raaij et al. 1999). Ad2 knob was expressed as a soluble product from a baculovirus expression system. The crystal structure again displayed the same 8 stranded β -sheet fold and has a root mean deviation of 1.95 Å based on C α when compared with the Ad5 knob structure. Recently, a crystal structure of Ad3 knob from subgenus B was solved (Durmort et al. 2001), which represents the first structure of a knob that does not use CAR as its cellular receptor. The protein was expressed in *E. coli* as an insoluble inclusion body, and through subsequent denaturation and dilution renaturation, a stable, trimeric protein was obtained. The overall topology of the Ad3 structure is again very similar to that of the other solved adenovirus knobs with each monomer consisting of two antiparallel β -sheets. Observed differences are a shorter β -strand F and an insertion of two short helical folds in the CD and EG loop regions. A comparison of C α atoms between the Ad3 and Ad5 knobs revealed a root mean deviation of 1.62 Å, indicating the close relationship between the two subgroups. In keeping with the notion that major antigenic determinants of knob map to variable loop regions of the subgenera, larger differences in structure were observed between the surface loops of Ad3 knob compared to Ad5 knob. In particular, the CD and EG loops, and part of the AB loop of Ad3 knob were shown to vary considerably from that of the other solved structures. A number of key residues in these loops are substituted in Ad3 knob, accounting for its inability to bind CAR D1.

3. Knob-CAR Interface

The interface between adenovirus type 12 knob and CAR D1 is made up of a number of contacts that contribute to the formation of a high affinity complex. Four loop regions of knob have been found to interact with a single face of CAR D1; there is no structural or mutational evidence for an interaction with the second domain of CAR (D2). The complex forms between two adjacent knob monomers with residues from each subunit required for binding (Figure 3A, B, C). Specifically, this interaction involves the AB, CD and DG loops of one monomer, and the FG loop of the second monomer. The

majority of contacts occur within the AB loop region of knob where hydrogen bonds, salt bridges and van der Waals contacts all take place and appear to form a key anchor for the complex. Residue Asp415 forms a hydrogen bond and salt bridge with Lys125 of CAR D1 and also a second salt bridge with Lys123; prolines 417 and 418 make a number of van der Waals contacts with CAR D1 residues Glu58, Val72, Leu75, Tyr85 and Lys 125. The importance of the AB loop is highlighted by mutational analyses which showed that amino acid substitutions inhibit formation of the complex (Bewley et al. 1999; Roelvink et al. 1999; Kirby et al. 2000). Two other loops within the same monomer also show important contacts; residues Val450 and Lys451 of the CD loop and residues Gln487, Gln494, Ser497 and Val498 of the DG loop (which contains the E and F strands) all interact with CAR D1. The DE loop residue Gln487 forms a hydrogen bond with Ser77 of CAR D1. Residues from the adjacent monomer of the FG loop, Pro517, Pro 519, Asn520 and Glu523 also have direct contact with CAR D1, with Asn520 forming a hydrogen bond with Asp70 of CAR D1. An important residue conserved throughout all human adenovirus serotypes, apart from the short fibers of subgenus F (Ad40 and Ad41), is Lys429. This residue lies just below the AB loop on β -strand B down the central y-axis of the interface and forms a hydrogen bond and salt bridge with Glu58 of CAR D1. The strict conservation of this residue throughout the subgenera and the important contacts made with CAR makes this one of the few residues that are crucial for cellular attachment.

The protein-protein interface formed by knob and CAR D1 is atypical due to the low topological complementarity between the two molecules. Although cavities within protein interfaces are common (Hubbard and Argos 1994), the size of the pocket formed within the complex is larger than usually observed. This cavity is most likely filled with solvent and allows for water-mediated bonding between the two molecules. The cavity sits between the AB and DE loops of knob and is approximately 240 \AA^3 in size (Figure 3B). The cavity increases the footprint at the interface by approximately 15% to a value of $\sim 1880 \text{ \AA}^2$ (820 \AA^2 from one knob monomer, 140 \AA^2 from the adjacent monomer and 920 \AA^2 by CAR D1) compared to what would be predicted from the values of the buried surface area. A mixture of hydrophobic and polar groups line the walls of the cavity, with over 60% being either backbone atoms or side chains that are conserved in other

adenoviruses. This cavity theoretically can accommodate four water molecules, although in the structural model, only one water molecule was observed (Bewley et al. 1999). The use of water as a bonding agent to stabilize the knob-CAR interface lowers the constraints on the virus to maintain specific amino acid side chain contacts with CAR. As a result, there is not strict conservation of residues in the AB and DE loops which form the walls of the water-trapping cavity (Figure 3D). This indirect, water-mediated mode of binding allows a greater degree of antigenic drift of residues which form the receptor binding site without loss of binding specificity.

Interestingly, the structurally unrelated HIV gp120-CD4 complex shows a similar unusual mismatch in surface topology which results in the formation of large cavities (Kwong et al. 1998). The largest of these is 273 \AA^3 and is formed by an equal number of gp120 and CD4 residues. The authors note that the cavity acts as a water buffer which allows for variation on the gp120 surface and the formation of an 'anti-hotspot', which may help the virus escape from antibodies directed against the CD4 binding site. Thus, the gp120-CD4 and knob-CAR complexes, although structurally different, may both employ a similar mechanism for evasion of the host immune response. More recent results of the gp120-CD4 complex suggest that conformational rearrangements of the gp120 protein also provide a mechanism for evasion of the immune system (Myszka et al. 2000). Such rearrangements are not observed with the adenovirus knob protein, where little conformational change accompanies formation of the knob-CAR complex.

The subgroup B adenoviruses are suggested to use a still unknown cellular receptor for viral attachment (Defer et al. 1990; Roelvink et al. 1998). Sequence alignments of this subgroup with other CAR-binding serotypes suggests that differences within the AB loop of the proteins may affect receptor binding (Figure 4). Although it is clear that the AB loop is a major determinant of binding specificity, the contribution of individual residues within the AB loop or other loops is difficult to assess from sequence alignments alone, due to underlying local differences in the structures or surface properties of knobs from different serotypes. Thus, mutational analysis, binding studies and structural characterization together are required for a complete understanding of the contribution that individual residues make to the interaction of knob with the cellular receptor, and for the precise localization of the receptor binding sites on these knobs.

Models based on structural data of Ad3 knob predict that AB loop residue Glu140 would form charge clashes with two acidic residues of CAR, Asp56 and Glu58, in a hypothetical Ad3-CAR D1 complex (Durmort et al. 2001). Other differences between the Ad3 and Ad12 knobs that likely account for the inability of Ad3 knob to interact with CAR D1 are substitution of critical Ad12 knob AB loop residue Asp415 for lysine at the equivalent position in Ad3 knob, and also differences in conformation of the DE and EG loops between these 2 serotypes. This change of conformation within the DE and EG loops of Ad3 knob may allow for interaction with receptors larger than CAR D1.

3.1. Binding studies and mutational analysis of knob CAR complexes

The adenovirus knob domain has been extensively studied through site-directed mutational analysis with a view to understanding the interactions important for receptor binding. A number of methods have been used to quantify the binding of mutant knob proteins to CAR D1. Techniques, such as cell binding assays, have been employed by a number of investigators, who used different indirect detection mechanisms (Roelvink et al. 1999; Santis et al. 1999; Kirby et al. 2000). These assays present the native receptor in an orientation that mimics that of a normal cellular infection and gives rise to affinity values in the sub-nanomolar to nanomolar range for knob-CAR binding. More recently, surface plasmon resonance (SPR) has been used to investigate binding of recombinant knob and CAR molecules (Kirby et al. 2000; Kirby et al. 2001; Lortat-Jacob et al. 2001). This technique more readily lends itself to high throughput analysis of mutant proteins. Results of Lortat-Jacob *et al* show the importance of the orientation of knob and its receptor in such assays. Ad2 knob binding to immobilized CAR, as found in the native state on cells, has an affinity of ~1 nM; however, if knob is immobilized and exposed to soluble CAR D1 then the affinity decreases to ~24 nM. Analysis of these data indicated that in the native orientation, the first binding event is 1:1 for Ad2 knob and CAR D1; subsequently, the complex is further stabilized by the association of two further CAR D1 molecules with the trivalent knob. This avidity mechanism of trivalent binding results in the high overall affinity that is observed between fiber knob and CAR and is analogous to the common bivalent antibody-antigen binding.

Other SPR studies using CAR D1, CAR D2 and CAR D1D2 indicate that the knob-CAR interaction requires only the first domain of CAR for efficient binding (Kirby et al. 2000) as reported previously through direct and indirect methods (Freimuth et al. 1999). Different serotypes of knob have been shown, through SPR, to have varying affinities for CAR D1. Ad5 and Ad12 knobs and the long fiber knob from Ad41 have affinities between 8-15 nM for CAR D1, whereas Ad9 knob has an affinity of 6400 nM (Kirby et al. 2001). Thus, the subgroup D Ad9 and the subgroup C Ad5 knobs differ by about 800-fold in binding affinity for CAR D1. Substituting Ad9 knob residue Asp222 with lysine, the structurally-homologous residue in Ad5 knob, increased the affinity 20-fold to 310 nM. The converse mutation in Ad5 knob (Lys442Asp) decreased its affinity for CAR D1 by 350-fold to 2800 nM. Other mutations substituting Ad5 and Ad12 knob residues into Ad9 knob (e.g. Lys260Pro and Lys260Gln, respectively) did not have any measurable effect on affinity. Earlier studies indicated that Ad2 and Ad9 knob utilize the same primary receptor (later found to be CAR); however, the shorter fiber length of adenovirus type 9 was suggested to allow fiber-independent binding of the penton base to integrins on the cell surface (Roelvink et al. 1996). This situation suggests a limited role for knob in the attachment of Ad9 to cells, which, therefore, would not require such high affinity interactions as those seen for the longer fibers. Importantly, these results indicate that the binding sites on adenovirus knob, although similar between serotypes, often do not involve identical residues or contacts.

In an effort to further define the differences between serotypes within the knob domain, we have developed a fluorescence polarization method of measuring knob binding to CAR D1. This technique is entirely solution-based and thus removes the effects of immobilization of the cellular receptor. Therefore, results pertain solely to the affinity of individual binding sites on knob for CAR D1 without the avidity effects observed in other techniques. These studies show that Ad2 has a 5-fold greater affinity for CAR D1 than Ad12 knob (Howitt and Freimuth, unpublished data).

It was inferred from modeling studies overlaying the Ad2 knob structure with the Ad12 knob-CAR D1 structure that two residues within Ad2 knob could form hydrogen bonds that were not observed in the Ad12 knob-CAR D1 complex. Using this structural model as a guide, it has been possible to change the affinity of both Ad2 and Ad12 knobs

for CAR D1. Substitution of Ad12 knob residue Pro417 with serine, the structurally-equivalent residue in Ad2 knob, increased the binding affinity of the resulting Ad12 knob mutant P417S for CAR D1 by 3-fold. A crystal structure of the Ad12 P417S mutant-CAR D1 complex confirmed the mutation and the formation of the predicted extra hydrogen bond (Howitt and Freimuth, unpublished data). Interestingly, Glu58 of CAR D1, the residue which forms the new hydrogen bond with Ser417 of the Ad12 knob mutant P417S, also is hydrogen bonded to Lys429 of Ad12 knob, a conserved residue in the majority of fiber knobs.

The second key residue is Tyr477 of Ad2 knob, which corresponds to serine 489 in Ad12 knob. Mutation of Ad12 knob Ser489 for tyrosine increased the affinity of Ad12 knob for CAR D1 by 10-fold, to a value greater than that observed for Ad2 knob (Howitt and Freimuth, unpublished data). This mutation also is thought to add a hydrogen bond at the interface; however, the bonding pair in CAR D1 is not currently known as no crystal structure has been solved. Tyrosine residues are commonly found in the interfaces of complexes (Hubbard and Argos 1994) and have considerably lower hydrophobic transfer free energies compared with aliphatic side chains (Stites 1997). This configuration may provide a way to cover large, normally solvent-filled surfaces without destabilizing the subunit but while allowing for the contribution of an important hydrophobic functional group within the receptor-ligand interface. The addition of this residue to the binding interface may, therefore, have a 2-fold effect and lead to the higher affinity binding that we observed.

Why do the knob domains from different adenovirus serotypes vary in binding affinity for CAR D1? It is clear that viral receptor-binding proteins are under strong selective pressure both to evade neutralization by varying their antigenic structure and to simultaneously maintain the ability to interact with the cellular receptor. Interestingly, although residues which form the receptor-binding sites of CAR-binding knobs are relatively more highly conserved than other regions of knob, these residues in fact are not strictly conserved (Bewley et al. 1999). The receptor-binding sites on knob, therefore, may be designed to tolerate a certain degree of sequence variation without complete loss of receptor binding affinity and specificity. One feature that may serve to buffer the binding sites against sequence variations is that they are constructed primarily from

loops, which have a greater degree of flexibility than other regions of regular secondary structure. Another aspect that may accommodate sequence variations is the avidity effect of having a trivalent structure; although certain residue changes may weaken the affinity of individual binding sites for CAR D1, the overall binding affinity increases when one or two additional receptors are recruited into the initial knob-CAR complex.

Clearly not all adenovirus serotypes have evolved to bind CAR at maximal affinity. This finding may imply either that low affinity knobs are younger in evolutionary terms and have not matured to form high affinity contacts, or that the high affinity binding knobs represent a dead-end in evolutionary terms as they are irreversibly constrained to a single receptor. Lowering the affinity of interaction between the receptor and ligand would likely have the same effect on efficiency of virus infection as decreasing the concentration of available receptors. Indeed, the efficiency of Ad2 infection was found to vary in direct proportion to the concentration of receptors in the cell membrane (Freimuth 1996); therefore, it appears that the transmission efficiency of viruses with low binding affinities would be poor. In viral evolution, this arrangement may allow for emergence of viruses that have novel receptor-binding specificities. Supporting this view is the notion that viruses cannot survive a complete loss of receptor binding at anytime during their evolution. Thus, it is likely that changes in receptor specificity would occur through a stepwise reduction in affinity for an initial receptor, with an accompanying decrease in transmission efficiency, followed by an increase in affinity for a new receptor. Alternatively, a second site for receptor recognition may be employed by the virus at certain times, which would allow for large mutations within the former receptor binding site in a single or multi-step process. Although actual pathways for evolution of new specificities are unknown, the current structural and functional data determined for knob suggests that viral tropism is mediated through a stepwise process and results in similar structures which are able to recognize different receptors.

3.2. Modification of binding specificity--Implications for adenovirus-based vectors for gene therapy

Adenoviruses have a number of advantages as vehicles for gene therapy, including broad host range, high efficiency of gene transfer *in vivo* and failure to

integrate the viral genome into host cell chromosomal DNA (Wilson 1996; Robbins et al. 1998; Benihoud et al. 1999). However, if adenoviruses are to be used in more sophisticated applications, retargeting of the viral vector towards specific cell types will be required. Such retargeting may be achieved through two major methods of altering the cellular tropism. The first is the removal of the innate receptor-binding specificity through the *in vitro* addition of bispecific molecular bridges that contain other ligands known to bind to other cellular markers (Douglas et al. 1996; Dmitriev et al. 1998; Dmitriev et al. 2000; Haisma et al. 2000; Trepel et al. 2000). The second technique is genetic engineering of the knob domain to recognize other receptors and inhibit the natural tropism for CAR. The first method of retargeting through the fusion of separate ligands is outside the scope of this review. The genetic method of vector targeting requires knowledge of the residues in functional contact with the receptor and an understanding of the binding process. The use of computational techniques combined with known structural data may facilitate redirecting the ligand towards specific cells. Also, the use of *in vitro* evolution methods may accelerate the selection of variant knob domains that recognize novel receptors. The recent development of a phage display system for knob (Pereboev et al. 2001) may represent the first step towards high-throughput techniques for altering knob binding specificity. The architecture of the knob protein appears to be well suited for the development of novel binding specificities, since the exposed loops used for binding receptors have a high tolerance for changes without impacting the structural integrity of the protein. Also, the trivalent nature of the domain may allow for low receptor affinity to be enhanced through avidity effects to levels where cellular recognition may occur. Therefore, future studies aimed at redirecting knob to bind to novel receptors may prove highly beneficial for the development of gene delivery vectors for treatment of specific cell types.

4. Structure of full-length CAR protein

CAR serves as a cellular receptor for both adenovirus and group B coxsackieviruses. In collaboration with Michael Rossmann's group at Purdue University, we investigated the interaction of coxsackievirus B3 (CVB3) with human CAR (He *et al*, *in press*). Initial attempts to visualize complexes of CVB3 bound to the recombinant

CAR D1 domain by cryoelectron microscopy were disappointing in that electron density corresponding to CAR D1 was very weak, an indication that binding sites on the virus capsid were not completely saturated with CAR D1. We, therefore, focused on production of the complete CAR ectodomain (CAR D1D2) and the full-length intact CAR protein, in hopes that these forms of CAR might interact more strongly with the CVB3 capsid.

To produce the complete CAR ectodomain, we turned to mammalian cell expression systems since the CAR D1D2 fragment was not soluble when expressed in bacteria, and its solubility was not increased by carboxy-terminal peptide extensions that enhanced the solubility and proper folding of CAR D1. Just prior to embarking on this project, Rose and colleagues had described a system in which cDNAs were expressed in mammalian cells using vectors based on vesicular stomatitis virus (VSV) (Schnell et al. 1996). Due to the aggressive nature of VSV infection, this system was attractive for rapid, high-level protein expression. Furthermore, because VSV overtakes the host cell biosynthetic pathways very rapidly, cell cultures can be maintained in serum-free media during the course of infection. These features, in the case of CAR D1D2 and also for other secreted proteins, greatly facilitate protein purification since the recombinant protein accounts for a large fraction of the total protein contained in the spent culture medium.

In collaboration with Michael Whitt and colleagues at UT-Memphis, we constructed a VSV-based expression vector for production of the CAR D1D2 ectodomain fragment. Earlier work from Whitt and others had shown that the VSV glycoprotein (G) gene could be deleted and that the resulting Δ G virus could be propagated on host cells that were transfected with plasmid DNA encoding the G gene (Robison and Whitt 2000). In our initial construct, the VSV G gene was substituted by a cDNA fragment encoding the CAR signal peptide, domains D1 and D2, and a stop codon immediately after D2. Although detectable CAR D1D2 was present in the culture fluid of BHK cells infected with this initial VSV-CAR virus, the levels of secreted protein were substantially lower than seen for other secreted proteins produced from comparable vectors. We noticed that the CAR signal peptide is shorter and less hydrophobic than typical mammalian signal peptides; therefore, the cDNA sequence encoding the CAR signal peptide was deleted

from the initial construct and replaced by a cDNA fragment encoding the VSV-G protein signal peptide. This substitution improved secretion levels, and the chimeric CAR D1D2 protein accumulated in the culture fluid to approximately 0.01 mg/ml. The secreted CAR D1D2 protein was further purified and concentrated by affinity chromatography on a matrix of Ad12 knob protein covalently coupled to Sepharose beads.

A mammalian cell expression system also was developed for production of full-length CAR protein. As described below, we isolated a bacterial artificial chromosome (BAC) containing the intact, functional human CAR gene and the 3' end of an adjacent gene, ANA. When transfected into mouse A9 cells along with a selectable neomycin phosphotransferase gene, this BAC conferred a receptor-positive phenotype to a large percentage of neomycin-resistant colonies. A number of A9 subclones that were stably transfected with this BAC were characterized for cell surface expression of CAR protein. Cell monolayers grown on coverslips were incubated with biotinylated Ad12 knob and then stained with fluorescein-labeled anti-biotin antibody. As shown in Figure 5, CAR protein levels varied over a wide range in individual subclones. Furthermore, Southern and Northern blot analyses showed that, in most cases, CAR protein and mRNA levels were proportional to the number of CAR genes integrated in these cell subclones. Such dosage-dependent and position-independent transgene expression has been observed in other systems and often results from the activity of DNA elements that maintain the integrated transgene in an open chromatin conformation (Bell et al. 2001). The close proximity of the CAR and ANA genes at this chromosomal locus suggests a possible need for regulatory elements to insulate these 2 genes from each other's regulatory signals, possibly accounting for the dosage-dependent expression we observed in transfected cell lines.

One A9 subclone, A9.5, expressed very high levels of CAR protein on the cell surface and was used as a source of full-length CAR protein. CAR protein was purified from Triton X100 extracts of A9.5 cells in a single step, by chromatography on a knob-Sepharose column as described above. The eluate from this column contained a major protein species of ~46 kD, which subsequently was identified as CAR by Western blotting and amino-terminal sequencing.

Gradient-purified CVB3 virions were incubated with full-length CAR or the VSV-produced CAR D1D2 fragment and then examined by cryoelectron microscopy (cryoEM). This analysis generated $\sim 20\text{\AA}$ resolution electron density maps for both complexes, which revealed the relative positions of the CAR D1 and D2 domains (Figure 6). In addition, the cryoEM structure provided sufficient detail to allow unambiguous fitting of the high resolution structure of the CAR D1 domain, which had previously been determined by X-ray crystallography. The resulting model demonstrated that CVB3 also interacts solely with the CAR D1 domain, and it revealed that the surface of CAR D1 in contact with the CVB3 capsid is different from the surface which contacts the adenovirus fiber knob domain.

More important for this discussion, however, was the striking observation that full-length CAR forms dimers when bound to the CVB3 capsid. The cryoEM structure shows that the cytoplasmic domains of adjacent CAR subunits form a region of ordered electron density, suggesting that the cytoplasmic domain mediates CAR dimerization. Although no supporting biochemical evidence for this CAR dimer has yet been reported, comparison of the cryoEM structures of the CVB3-CAR D1D2 and CVB3-full length CAR structures revealed that the orientation of the CAR ectodomain relative to the CVB3 capsid is essentially superimposable for both structures. Therefore, it seems unlikely that dimerization of CAR could result from non-specific binding between cytoplasmic domains, since this would most likely generate a disordered electron density and also might force the CAR ectodomain to dissociate from the CVB3 capsid or alter the bending angle between the CAR D1 and D2 domains.

The formation of CAR dimers would have potentially important implications for both CVB and adenovirus infection. Receptor dimerization would likely increase the avidity of virus attachment to the host cell membrane, thereby increasing the efficiency of infection. At present, it is unknown whether CAR dimers pre-exist on the cell membrane or whether dimerization is induced by virus binding. If the latter scenario proves to be correct, then it is possible that dimerization might generate signals that contribute to post-binding events in infection. It should be noted again here that GPI-linked forms of CAR can serve as functional receptors for both CVB and adenovirus, indicating that dimerization through the cytoplasmic domain is not required for infection (Wang and

Bergelson 1999). However, this does not exclude the possibility that dimeric receptors play a more important role in infections *in vivo* or in the normal cellular function of CAR.

5. Characterization of the human CAR gene, *CXADR*

5.1. Chromosomal Location of *CXADR*

The gene for human CAR, *CXADR*, is located on the long arm of chromosome 21 at position q21.1, about 15.96 Mbp from the telomere and 4.55 Mbp from the centromere (Hattori et al. 2000; Antonarakis 2001). Transcription is in the direction of the telomere, and the transcribed region extends over about 57 kbp. Upstream of *CXADR* (toward the centromere) are several predicted genes, corresponding to ESTs of unknown function, and several pseudogenes. The first gene of known function, *USB25*, a ubiquitin-specific proteinase, is more than 16 Mbp away. On the downstream side, *ANA (BTG3/TOFA)* is transcribed toward the *CXADR* gene from the opposite strand, and the 3' ends of the *CXADR* and *ANA* transcripts are separated by about 23.5 kbp (Figure 7). Chromosome 21, which is perhaps best known for its participation in Down's syndrome, is relatively sparsely populated with genes, especially the proximal half of the long arm on which *CXADR* is located. The entire chromosome is thought to contain only about 230 known or predicted genes, whose coding sequences collectively account for about 3 percent of the total DNA sequence of chromosome 21 (Antonarakis 2001).

The genomic location of human *CXADR* was first inferred from an analysis of a panel of mouse-human somatic cell hybrids, each of which contained only a single human chromosome in the background of mouse A9 cells (Mayr and Freimuth 1997). Mouse A9 cells, a subline derived from L929 cells, like most cultured rodent cell lines, express no or few adenovirus receptors. Among the cell panel, only cells containing human chromosome 21 were capable of binding human group A or C adenoviruses, suggesting that chromosome 21 encoded at least one essential protein component of the adenovirus receptor. This conclusion was further supported by analysis of several human-hamster hybrid cell lines containing human chromosome 21 fragments. Identification of the CAR cDNA (Bergelson et al. 1997; Tomko et al. 1997), followed shortly by reports

of the complete sequence of human chromosome 21 (Hattori et al. 2000), allowed a precise determination of the location of the CAR gene, as well as several CAR pseudogenes, two of which also are located on chromosome 21 (Bowles et al. 1999). It should be noted that most if not all vertebrates, including mice, have a gene homologous to *CXADR*; however, in many established mouse and hamster cell lines this gene is not expressed.

5.2. Sequence Analysis of *CXADR*

As noted above, we had observed that mouse A9 and Chinese hamster ovary (CHO) cell lines failed to bind radiolabeled adenovirus particles but that Ad2 virus did bind to derivatives of these cells which contained human chromosome 21. To isolate the human gene(s) which encoded receptor activity, high molecular weight genomic DNA from the chromosome 21-containing hybrid cells (A9-21 cells) was transfected into normal A9 cells along with a selectable neomycin phosphotransferase gene, which confers resistance to the cytotoxic drug G418. G418-resistant colonies that expressed adenovirus receptors were identified by their ability to bind modified sheep erythrocytes whose membranes were coated with purified viral fiber protein. Several independent, receptor-bearing A9 cell clones were isolated and shown to contain a similar pattern of DNA restriction fragments that hybridized to a human Alu repeat probe on Southern blots (Mayr and Freimuth 1997). One Alu repeat-containing fragment diagnostic of this locus was cloned and then used to probe a phage library of human (HeLa cell) genomic DNA. This approach ultimately led to the isolation of several phage clones whose HeLa DNA inserts overlapped to form a contig spanning about 30 kb. Hybridization probes derived from the ends of this contig were then used to screen a BAC library of large human genomic DNA fragments. One clone was isolated that contained a ~120 kb insert and both ends of the 30 kbp HeLa contig. When transfected into mouse A9 cells along with a neomycin phosphotransferase gene, this BAC clone, designated BAC-CXADR1, transformed a large fraction of G418-resistant colonies to express adenovirus receptor activity (see Figure 5). Restriction fragments of this BAC clone hybridized to CAR cDNA on Southern blots.

To determine the sequence of the BAC-CXADR1 insert, *Bam*HI and *Bg*III fragment subclones were prepared, and both strands of selected subclone inserts were sequenced using the dideoxy chain termination method and a combination of techniques including a nested deletion approach and primer walking as described elsewhere (Connelly et al. 1998; Anderson et al. 2001). Where overlapping fragments were not available, fragment order was determined by sequencing junction fragments produced by PCR. A 121,793 nucleotide insert sequence was derived that was consistent with restriction analyses of the original BAC clone and with restriction analyses of individual subclones (Figure 7).

The ~57 kbp CAR gene lies near the middle of the 122 kbp CXADR1 genomic fragment, suggesting that most of the signals required for regulating CAR expression lie within this segment. The overall G+C content for the CXADR1 insert is 40.1%, which is almost identical to the average of 40.89% for the entire sequenced portion of chromosome 21 (Hattori et al. 2000). Thirty-eight percent of the fragment consists of interspersed repeat sequences most of which are SINE (23%) or MIR (21.4%) sequences, followed by LINE (9.8%) elements. This compares to 10.8 percent SINE and 15.5 percent LINE, and 38 percent total repetitive elements for chromosome 21 as a whole (Antonarakis 2001). The CXADR1 segment contains 26 simple repeat sequences and 29 regions of low complexity. Inspection of the CXADR1 sequence immediately upstream of the homology to the 5' end of the CAR cDNA does not reveal a likely TATA- or CAT-box sequence. Indeed, this region is quite GC rich, and inspection of the CpG frequency revealed a region of high CpG content that overlaps *CXADR* exon 1 (Figure 8). The promoters of housekeeping genes are universally associated with CpG islands (Antequera and Bird 1999). Therefore, while experimental evidence for the location of the CXADR promoter is lacking, the available evidence is consistent with the existence of a housekeeping-type promoter located at or immediately upstream of the region that encodes the 5' end of the CAR cDNA.

Comparison of the current 2434 nucleotide CAR cDNA sequence (Genbank NM_001338) to the BAC-CXADR1 insert sequence (Genbank AF200465) revealed an exact match to seven regions spanning a ~53 kbp segment of the BAC insert with two exceptions, a *Eco*RI restriction site (GAATTC) at the beginning of the CAR cDNA

sequence that is almost certainly derived from the cloning vector, and a few "A-track" nucleotides at the 3' end. This result indicated that the CXADR1 insert contains the entire CAR gene, consistent with the fact that transfection experiments conferred Ad2 binding activity to mouse A9 cells and that the CAR gene is composed of seven exons (Bowles et al. 1999). This structure is not atypical of human housekeeping genes. The first ATG in the cDNA sequence is 54 nucleotides from the 5' end of the cDNA (excluding the *Eco*RI site) and initiates the 1098 nt CAR open reading frame.

The internal exon-intron boundaries all obey the AG/GT splice site rule (Jackson 1991); exons 2 through 6 have an average length of 158 bps; translation is predicted to begin in exon 1 and it terminates at a TAG codon 263 nucleotides into exon 7. However, four of the six introns are larger than average, especially intron 1, which is 33,886 nt, and intron 3, which is 11,780 nt. The average intron length for vertebrates is about 1200 nt (Hawkins 1988). The 3' ends of the published cDNA sequences end at or in a sequence of 13 A residues at the beginning of a 30 residue A-rich segment in the CXADR1 BAC that includes only 3 non-A nucleotides. This A-rich region represents one end of an Alu element. No polyA addition signal (AAUAAA) is found upstream of this polyA track, suggesting that priming of the cDNA may have initiated within the Alu element rather than at the true 3' end of the CAR mRNA.

CAR cDNA expression studies by Tomko *et al* revealed a range of CAR mRNA species in HeLa cells with sizes from 1.3 to 6 kb (Tomko et al. 1997). Our own Northern blot studies revealed a major CAR species of about 6 kb (data not shown). This length is surprising because it would make the CAR mRNA six times the length required to encode the open reading frame and one of the longest 3' UTR sequences known (Pesole et al. 1997). The average size for mammalian 3' UTR elements is about 740 nt. Comparison of a 4000 nucleotide segment of the BAC CXADR1 sequence distal to the beginning of exon 7 with the NCBI EST database revealed the presence of a number of cDNAs in the database that terminated at or just before nucleotide 87294 in the BAC CXADR1 sequence. Eleven nucleotides 5' of this site is an AAUAAA sequence associated with the addition of polyA tails. These observations suggest that the CAR mRNA terminates at or near nt 87294 in the CXADR1 sequence, thus making the CAR cDNA ~5610 nucleotides in length and the CAR 3' UTR 4409 nucleotides. Consistent

with this observation, primers just within the predicted boundaries of CAR exon 7 amplified a ~4 kbp fragment from HeLa cell cDNA that was indistinguishable in size from the BAC amplified product, whereas PCR with a 3' primer just distal to the predicted 3' end of the CAR cDNA failed to yield a product (Freimuth and Anderson, unpublished data). There are no open reading frames in the 3' UTR longer than 100 codons, although two ATG-initiated open reading frames longer than 200 codons are found in the CXADR1 sequence that are not associated with known features (see Fig. B). The 3' UTR contains an AluY element and part of a FLAM C SINE element near the middle of the exon (at 85332-85624 bp and 85756-84030 bp in CXADR1, respectively). The orientation of the Alu element is such that its A track would serve as a template for RT-PCR amplification with an oligo-T primer, and this very likely is the mechanism by which the partial CAR cDNA was cloned (Bergelson et al. 1997; Tomko et al. 1997). Approximately 5% of human mRNAs contain Alu sequences (Yulug et al. 1995). It is not known if the unusually long CAR 3' UTR is physiologically important. The levels of CAR mRNA in many cells is quite low, however, and one function of the 3' UTR may be to regulate mRNA stability. Several sequence elements in the 3'UTRs of mRNAs, including a A+U-rich motif AUUUA, bind proteins that are thought to mediate mRNA degradation (Guhaniyogi and Brewer, 2001). The 4674 nucleotide exon 7 of CAR has 18 of these elements, 11 of which are distal to the Alu A-track. This number of AUUUA elements is unusual in mRNA sequences. For example, the entire 12,500 nucleotide-long mRNA for DNA-PKcs, the catalytic subunit of a protein kinase involved in DNA repair, which is quite stable, contains only 12 AUUUA elements. Nevertheless, 18 is about the number expected for a 4,600 nt segment of the nucleotide composition of exon 7, which like many mRNA 3'UTR segments, is A+U rich (Guhaniyogi and Brewer 2001).

Comparison of the BAC insert sequence with the Genbank database also revealed exact matches to presumptive exons 4 and 5 of the human *ANA* (TOFA/BTG3) cDNA (Guéhenneux et al. 1997; Yoshida et al. 1998). The *CXADR* and *ANA* genes therefore are transcribed towards each other (Fig. 7). *ANA* was mapped to chromosome 21 q11.2 to q12.1, consistent with the location of the CAR gene. An EST database analysis of the segment in the CXADR1 sequence distal to the reported 3' end of the 1210 nucleotide *ANA* cDNA (Genbank D64110) also revealed a set of cDNAs that terminated at or near

nucleotide 110,834, that is 83 nucleotides distal to the last reported nucleotide of the *ANA* cDNA. Fourteen nucleotides proximal to this site is an AATAAA sequence, consistent with position 110,834 being the approximate 3' end of the *ANA* mRNA. Between this site and the predicted end of the *CAR* mRNA are 23,539 nucleotides of intergenic sequence with an average G+C content of 38%.

Comparison of the remainder of the *CXADR1* sequence with the NCBI EST database revealed one additional area between nt 17,634 and 18009 that is similar to a number (>100) of EST clones homologous to ribosomal protein L39. Because none of the EST sequences provided an exact match to the *CXADR1* sequence, this segment may represent a fragment of a pseudogene rather than an expressed sequence related to the L39 protein.

5.3 *CXADR* Polymorphism Analysis

Although no human disease has been associated with the *CAR* locus, it is possible that individuals might differ in their sensitivity to adenovirus or coxsackievirus infections because of variations in the *CAR* sequence. As noted above, the 345 amino acid *CAR* protein is composed of a signal sequence (amino acids (aa) 1-19), two extracellular IG-like domains (aa34-127, 155-219), a transmembrane region (aa238-258), and a cytoplasmic domain (aa259-365). Inspection of the *CXADR* gene structure revealed that the signal sequence is encoded primarily in exon 1 (aa1-14), the first IG-like domain is encoded by exons 2 and 3 (aa15-69 and aa 70-138), the second IG-like domain is encoded by exons 4 and 5 (aa139-190 and aa 191-231), and the transmembrane region and first part of the cytoplasmic domain by exon 6 (aa232-277). The *CAR* protein domain organization and splicing pattern also are seen in a small subset of other Ig superfamily members termed the CTX family (Chretien et al. 1998). CTX family proteins also are found in lower vertebrates, and some family members appear to function in cell-cell adhesion during early stages of development. The human genome encodes a second CTX family protein, the A33 antigen, which is the closest known sequence homologue to *CAR* (Heath et al. 1997; Chretien et al. 1998). A33 is expressed on many colon and intestinal tumor cells (Barendswaard et al. 1998). Differential regulation of *CAR* in tumor vs normal cells, however, has not been reported.

About a dozen polymorphic nucleotides (SNPs) in the gene have been identified by comparing sequences from different individuals and EST sequences in the public databases, and a chromosome 21-specific SNP web database was established as a joint effort between the Division of Medical Genetics of the University of Geneva Medical School and the Swiss Institute of Bioinformatics (<http://csnp.unige.ch/>) (Deutsch et al. 2001). Unfortunately, at the current time, all of the CAR-related SNPs in that database are in the 3' UTR except one, which is a silent change in the third position of a proline codon (CAR amino acid 40) in exon 2.

To more systematically examine potential genetic diversity in CAR, we examined the sequences of exons 2-6 encoding all of the mature extracellular receptor protein from a number of different individuals by resequencing PCR amplified genomic segments containing these exons (Figure 9). DNAs from 90 individuals representative of the normal U.S. population were obtained from the Coriell Institute Polymorphism Discovery Resource (Collins et al. 1998). Both strands of each amplified fragment were then sequenced by the dideoxy method using an automated sequencer, and the nucleotide traces were manually inspected for potential polymorphisms. For exons 2 and 3, which encode CAR D1 (the region of CAR that interacts with both CVB and adenovirus), DNAs from 90 individuals were examined (Table 1); however, only two potential polymorphisms were identified. One of these (SNP 1) occurs in the third base of the codon for amino acid Gln111 (CAA). In the 180 chromosomes analyzed, A appeared at this position 170 times while G appeared 10 times for a minor allele frequency of 5.5%. Four individuals were heterozygotes while 3 were homozygous for the minor allele G. The second potential SNP is located 67 nucleotides distal to the 3' exon 3 boundary and was observed as a C/T heterozygote in only one individual. We did not detect polymorphisms at Pro40 (see above), indicating that this potential SNP is either very rare or represents a sequencing error.

Eight additional SNPs were identified in amplicon segments 4, 5 and 6, but only one of these was in exon sequence (SNP 8 in exon 5, see Table 1). The variant sequence was observed in only one of 24 individuals analyzed as a A/T heterozygote in the third nucleotide of the codon for Thr193 (ACT), which appears in the second Ig-type domain (CAR D2). Three other SNPs (SNPs 4, 5, and 10, Table 1), all in intron sequence, were

observed in only one individual, while the minor T allele of SNP 7, located proximal to the 5' boundary of exon 5, was observed in two individuals, one of whom was homozygous, giving an allele frequency of 6%. Of the remaining three SNPs, SNP 4 is in intron 3 near the 5' end of the exon 4 segment, and the minor allele was found in 3 of 24 individuals. SNP 6 is just proximal to exon 5, and the minor allele appeared in 10 of 24 individuals giving an allele frequency of 25%. SNP 9 occurs just distal to exon 6 and appeared with a minor allele frequency of 16%.

One sequence length and composition polymorphism was observed (designated A in Figure 9) just distal to exon 4. Our reference sequence (Genbank AF200465) had the simple sequence $A_{10}T_5$ at this position, and individuals differed in the number of As and Ts in this segment. Because both strands cannot be sequenced through the PCR amplified product from heterozygotes, we did not attempt to quantitate the composition and frequency of this polymorphism.

Although our analysis did not include the CAR leader sequence nor the second half of the cytoplasmic domain, we conclude from our analysis that coding change SNPs in CAR that affect adenovirus binding are likely to be very rare. This may suggest an important or essential role for CAR in cell metabolism; however, because of the small size of the coding region of the gene, at most only about one cSNPs would be expected for the sample size analyzed.

6. Concluding Remarks

From these analyses it appears that CAR has several favorable characteristics that may account for why both coxsackie B viruses and adenoviruses have selected it as their major attachment receptor. First, CAR is well conserved in vertebrate animals, so it has been a constant target during the evolution of virus-host systems. This high level of conservation parallels the low frequency of genetic diversity of the CAR gene in humans. Second, CAR has a fairly wide tissue distribution; thus receptor availability is not a major factor in determining virus tropism. Third, in comparison to antibodies, CAR is a fairly small and compact molecule which may allow for interaction with binding sites on viral proteins that are relatively inaccessible to antibodies. Fourth, the possible dimerization of

CAR either before or after interaction with virus could result in increased avidity of virus binding to the host cell membrane.

In the near future we can expect that additional cellular receptors other than CAR will be identified that interact with non-CAR binding serotypes such as those in subgroup B. Structural studies of these knob-receptor complexes should provide further insights into the constraints on altering receptor specificity and the range of possible receptor types with which knob might interact after alteration by *in vitro* evolution methods (Pereboev et al. 2001). Further study of these systems will provide a wealth of information on factors that govern molecular recognition, and they also may provide a means to alter the receptor specificity of adenoviral vectors for use in gene therapy and other applications.

Literature Cited

- Anderson, CW et al. (2001) Frameshift mutation in *PRKDC*, the gene for DNA-PKcs, in the DNA repair-defective, human, glioma-derived cell line M059J. *Radiat Res* **156**: 2-9.
- Antequera, F and Bird, A (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr Biol* **9**: R661-7.
- Antonarakis, SE (2001) Chromosome 21: from sequence to applications. *Curr Opin Genet Dev* **11**: 241-6.
- Bai, M et al. (1993) Mutations that alter an Arg-Gly-Asp (RGD) sequence in the adenovirus type 2 penton base protein abolish its cell-rounding activity and delay virus reproduction in flat cells. *J Virol* **67**: 5198-205.
- Barendswaard, EC et al. (1998) Rapid and specific targeting of monoclonal antibody A33 to a colon cancer xenograft in nude mice. *Int J Oncol* **12**: 45-53.
- Bell, AC et al. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic. *Science* **291**: 447-50.
- Benihoud, K et al. (1999) Adenovirus vectors for gene delivery. *Curr Opin Biotechnol* **10**: 440-7.
- Bergelson, JM et al. (1997) Isolation of a common receptor for Coxsackie B viruses and adenoviruses 2 and 5. *Science* **275**: 1320-3.

- Bewley, MC et al. (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**: 1579-83.
- Boulanger, PA and Puvion, F (1973) Large-scale preparation of soluble adenovirus hexon, penton and fiber antigens in highly purified form. *Eur J Biochem* **39**: 37-42.
- Bowles, KR et al. (1999) Genomic organization and chromosomal localization of the human Coxsackievirus B-adenovirus receptor gene. *Hum Genet* **105**: 354-9.
- Chretien, I et al. (1998) CTX, a *Xenopus* thymocyte receptor, defines a molecular family conserved throughout vertebrates. *Eur J Immunol* **28**: 4094-104.
- Collins, FS et al. (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* **8**: 1229-31
- Connelly, MA et al. (1998) The promoters for human DNA-PK_{CS} (*PRKDC*) and *MCM4*: divergently transcribed genes located at chromosome 8 band q11. *Genomics* **47**: 71-83
- Defer, C et al. (1990) Human adenovirus-host cell interactions: comparative study with members of subgroups B and C. *J Virol* **64**: 3661-73.
- Deutsch, S et al. (2001) A cSNP map and database for human chromosome 21. *Genome Res* **11**: 300-7.
- Dmitriev, I et al. (2000) Ectodomain of coxsackievirus and adenovirus receptor genetically fused to epidermal growth factor mediates adenovirus targeting to epidermal growth factor receptor-positive cells. *J Virol* **74**: 6875-84.
- Dmitriev, I et al. (1998) An adenovirus vector with genetically modified fibers demonstrates expanded tropism via utilization of a coxsackievirus and adenovirus receptor-independent cell entry mechanism. *J Virol* **72**: 9706-13.
- Douglas, JT et al. (1996) Targeted gene delivery by tropism-modified adenoviral vectors. *Nat Biotechnol* **14**: 1574-8.
- Durmort, C et al. (2001) Structure of the fiber head of Ad3, a non-car-binding serotype of adenovirus. *Virology* **285**: 302-12.
- Freimuth, P (1996) A human cell line selected for resistance to adenovirus infection has reduced levels of the virus receptor. *J Virol* **70**: 4081-5.
- Freimuth, P et al. (1999) Coxsackievirus and adenovirus receptor amino-terminal immunoglobulin V-related domain binds adenovirus type 2 and fiber knob from adenovirus type 12. *J Virol* **73**: 1392-8.

- Guéhenneux, F et al. (1997) Cloning of the mouse *BTG3* gene and definition of a new gene family (the *BTG* family) involved in the negative control of the cell cycle. *Leukemia* **11**: 370-5
- Guhaniyogi, J and Brewer, G (2001) Regulation of mRNA stability in mammalian cells. *Gene* **265**: 11-23.
- Haisma, HJ et al. (2000) Targeting of adenoviral vectors through a bispecific single-chain antibody. *Cancer Gene Ther* **7**: 901-4.
- Hattori, M et al. (2000) The DNA sequence of human chromosome 21. The chromosome 21 mapping and sequencing consortium. *Nature* **405**: 311-9
- Hautala, T et al. (1998) An interaction between penton base and alpha v integrins plays a minimal role in adenovirus-mediated gene transfer to hepatocytes in vitro and in vivo. *Gene Ther* **5**: 1259-64.
- Hawkins, JD (1988) A survey on intron and exon lengths. *Nucleic Acids Research* **16**: 9893-9908
- Heath, JK et al. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc Natl Acad Sci U S A* **94**: 469-74.
- Henry, LJ et al. (1994) Characterization of the knob domain of the adenovirus type 5 fiber protein expressed in *Escherichia coli*. *J Virol* **68**: 5239-46.
- Honda, T et al. (2000) The coxsackievirus-adenovirus receptor protein as a cell adhesion molecule in the developing mouse brain. *Brain Res Mol Brain Res* **77**: 19-28.
- Hong, JS and Engler, JA (1996) Domains required for assembly of adenovirus type 2 fiber trimers. *J Virol* **70**: 7071-8.
- Hubbard, SJ and Argos, P (1994) Cavities and packing at protein interfaces. *Protein Sci* **3**: 2194-206.
- Jackson, IJ (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* **19**: 3795-8.
- Kirby, I et al. (2000) Identification of contact residues and definition of the CAR-binding site of adenovirus type 5 fiber protein. *J Virol* **74**: 2804-13.

Kirby, I et al. (2001) Adenovirus type 9 fiber knob binds to the coxsackie B virus-adenovirus receptor (CAR) with lower affinity than fiber knobs of other CAR-binding adenovirus serotypes. *J Virol* **75**: 7210-4.

Kwong, PD et al. (1998) Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**: 648-59.

Lortat-Jacob, H et al. (2001) Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. *J Biol Chem* **276**: 9009-15.

Louis, N et al. (1994) Cell-binding domain of adenovirus serotype 2 fiber. *J Virol* **68**: 4104-6.

Mayr, GA and Freimuth, P (1997) A single locus on human chromosome 21 directs the expression of a receptor for adenovirus type 2 in mouse A9 cells. *J Virol* **71**: 412-8

Myszka, DG et al. (2000) Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci U S A* **97**: 9026-31.

Nemerow, GR (2000) Cell receptors involved in adenovirus entry. *Virology* **274**: 1-4.

Pereboev, A et al. (2001) Phage display of adenovirus type 5 fiber knob as a tool for specific ligand selection and validation. *J Virol* **75**: 7107-13.

Persson, R et al. (1985) Virus-receptor interaction in the adenovirus system: characterization of the positive cooperative binding of virions on HeLa cells. *J Virol* **54**: 92-7.

Pesole, G et al. (1997) Structural and compositional features of untranslated regions of eukaryotic mRNAs. *Gene* **205**: 95-102

Philipson, L et al. (1968) Virus-receptor interaction in an adenovirus system. *J Virol* **2**: 1064-75.

Robbins, PD et al. (1998) Viral vectors for gene therapy. *Trends Biotechnol* **16**: 35-40.

Robison, CS and Whitt, MA (2000) The membrane-proximal stem region of vesicular stomatitis virus G protein confers efficient virus assembly. *J Virol* **74**: 2239-46.

Roelvink, PW et al. (1996) Comparative analysis of adenovirus fiber-cell interaction: adenovirus type 2 (Ad2) and Ad9 utilize the same cellular fiber receptor but use different binding strategies for attachment. *J Virol* **70**: 7614-21.

- Roelvink, PW et al. (1998) The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. *J Virol* **72**: 7909-15.
- Roelvink, PW et al. (1999) Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. *Science* **286**: 1568-71.
- Santis, G et al. (1999) Molecular determinants of adenovirus serotype 5 fibre binding to its cellular receptor CAR. *J Gen Virol* **80**: 1519-27.
- Schnell, MJ et al. (1996) Foreign glycoproteins expressed from recombinant vesicular stomatitis viruses are incorporated efficiently into virus particles. *Proc Natl Acad Sci U S A* **93**: 11359-65.
- Seth, P et al. (1994) Mechanism of enhancement of DNA expression consequent to coinfection of a replication-deficient adenovirus and unmodified plasmid DNA. *J Virol* **68**: 933-40.
- Stites, SE (1997) Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**: 1233-1250
- Stouten, PF et al. (1992) New triple-helical model for the shaft of the adenovirus fibre. *J Mol Biol* **226**: 1073-84.
- Tainer, JA et al. (1982) Determination and analysis of the 2 A-structure of copper, zinc superoxide dismutase. *J Mol Biol* **160**: 181-217.
- Tomko, RP et al. (1997) HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc Natl Acad Sci U S A* **94**: 3352-6.
- Trepel, M et al. (2000) Molecular adaptors for vascular-targeted adenoviral gene delivery. *Hum Gene Ther* **11**: 1971-81.
- van Oostrum, J and Burnett, RM (1985) Molecular composition of the adenovirus type 2 virion. *J Virol* **56**: 439-48.
- van Raaij, MJ et al. (2000) Dimeric structure of the coxsackievirus and adenovirus receptor D1 domain at 1.7 Å resolution. *Structure Fold Des* **8**: 1147-55.
- van Raaij, MJ et al. (1999) Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. *Virology* **262**: 333-43.

van Raaij, MJ et al. (1999) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**: 935-8.

Wang, X and Bergelson, JM (1999) Coxsackievirus and adenovirus receptor cytoplasmic and transmembrane domains are not essential for coxsackievirus and adenovirus infection. *J Virol* **73**: 2559-62.

Wickham, TJ et al. (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**: 309-19.

Wilson, JM (1996) Adenoviruses as gene-delivery vehicles. *N Engl J Med* **334**: 1185-1187

Xia, D et al. (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 Å resolution. *Structure* **2**: 1259-70.

Yoshida, Y et al. (1998) *ANA*, a novel member of *Tob/BTG1* family, is expressed in the ventricular zone of the developing central nervous system. *Oncogene* **16**: 2687-93

Yulug, IG et al. (1995) The frequency and position of *Alu* repeats in cDNAs, as determined by database searching. *Genomics* **27**: 544-8

Figure Legends:

Figure 1: Ribbon diagrams of Ad5 knob monomer and trimer. **A**, knob monomer, with β strands labeled according to the nomenclature adopted by Xia *et al.* The R-sheet is composed of β -strands D, G, H and I, the V sheet is composed of β -strands A, B, C and J. **B**, knob trimer, as viewed from the cellular receptor.

Figure 2: Diagrams of CAR D1 and the Ad12 knob-CAR D1 complex. **A**, ribbon diagram of the CAR D1 domain with β strands labeled. Strands C, C' and C'' contain residues important for interfacial contacts with knob. **B**, the Ad12 knob-CAR D1 complex viewed down the fiber shaft. Ad12 knob trimer is shown as a ribbon structure and CAR D1 monomers as a backbone structure. The AB loop, which contains ~50% of the interfacial residues from knob, is shown in black.

Figure 3: Molecular surface representations of the interface between Ad12 knob-CAR D1 complex. **A**, surface diagram of two knob monomers in complex with a monomer of

CAR D1 (CAR D1 shown as a ribbon structure). The light grey monomer of knob contains the majority of interfacial contacts with the AB, CD and DG loops all interacting with CAR D1, the dark grey monomer contains the FG loop which interacts with CAR D1. **B**, surface diagram of two adjacent Ad12 knob monomers (same view as A) colored on a sliding scale from white (contact) to grey (no contact). Residues 415 and 418 are contained within the AB loop, and residues 487 and 497 are contained within the DG loop. The grey depression shown between residues 415, 418, 487 and 497 forms part of the large interfacial cavity that is observed in the Ad12 knob-CAR D1 complex. **C**, surface diagram of the CAR D1 surface that interacts with Ad12 knob. The surface is colored on a sliding scale: dark grey (contact) to light grey (no contact). **D**, sequence conservation diagram of two knob monomers viewed at the CAR D1 interface (same view as A). Molecules are highlighted on a sliding scale from white (highest relative level of amino acid sequence conservation) to dark grey (lowest relative level of conservation).

Figure 4: Amino acid sequence alignment of residues within the AB loop. Ad2, Ad9 and Ad12 interact with CAR, while Ad3 is suggested to recognize an unknown cellular receptor. Stars indicate conserved residues, and dots indicate similar residues. Highlighted in light grey are two residues of Ad3 knob that are thought to interfere with CAR binding, Lys138 and Glu140. Underlined are Ad12 knob residues in contact with CAR D1 in the Ad12 knob-CAR D1 complex.

Figure 5. Expression levels of human CAR protein and mRNA in mouse A9 cells transfected with BAC-CXADR. **A**, monolayers of mouse A9 cell subclones stably transfected with BAC-CXADR (panels 1-7), parental A9 cells (panel 8) and HeLa cells (panel 9) grown on glass coverslides were incubated sequentially with biotinylated Ad12 knob then fluorescein-conjugated anti-biotin antibodies. Bound antibody, an indirect indicator of CAR protein expression, was then visualized by fluorescence microscopy. **B**, Northern blot of total RNA and **C**, Southern blot of *Eco*RI-digested genomic DNA from transfected A9 cell subclones 1-7, using [³²P]-labeled human CAR cDNA as hybridization probe (only a single representative band from each analysis is shown).

Figure 6. Structure of full-length CAR and CAR D1D2 bound to coxsackievirus B3. **A** and **B**, cryoEM reconstructions of coxsackievirus B3 (CVB3) bound to full-length CAR and the complete CAR ectodomain (CAR D1D2), respectively. (Reproduced (permission pending) from He *et al* (2001) Interaction of coxsackievirus B3 with the full length coxsackievirus-adenovirus receptor. Nature Structural Biology 8(10) (in press).

Figure 7. Schematic diagram of the DNA insert contained in BAC *CXADR1*. The *Bam*HI restriction map for the 122 kb insert of BAC *CXADR1* is shown on the top line, oriented with the left and right ends pointing towards the centromere and telomere, respectively, of human chromosome 21. The size of the *CXADR* primary transcript and the direction of transcription are shown below the restriction map. Also indicated is the 3' end of the ANA gene. The nucleotide sequence of the BAC *CXADR1* insert was determined (Genbank accession number AF200465).

Figure 8. CpG and C+G Content of the *CXADR* gene promoter region. The number of "C" and "G" residues and CpG dinucleotides in a 200 nucleotide window was determined for the *CXADR1* insert. Shown is the region around the *CXADR* promoter, which represent the only major CpG island found in the insert. Peak CpG values occur at nucleotides ~30,030 and ~30,510 with respect to the beginning of the *CXADR1* insert. The 5' end of the CAR cDNA is at nucleotide 30,239.

Figure 9. Polymorphism analysis of *CXADR1* exons 2 through 6. Segments containing exons 2 through 6 from the genomes of DNAs from the Polymorphism Discovery Resource (Collins et al. 1998) were amplified by PCR and sequenced on both strands. Each exon is represented by a box that is positioned with respect to a central line representing the PCR amplified segment. The average region sequenced on both strands is represented by double arrows; the nucleotide length of this region and the length of the amplified region are given at the right. The positions of polymorphic nucleotides are indicated by vertical lines; numbers correspond to SNPs as described in Table 1; "A" refers to a sequence length/composition polymorphism (see text). 2081 nucleotides of the *CXADR1* gene were analyzed of which 790 corresponded to exon sequence. Eleven

potential polymorphisms were identified of which two were in exon sequence. Neither of these would affect the amino acid sequence of CAR.

Table 1. Single nucleotide polymorphisms in *CXADR* exons 2 through 6.

No.	Location ^a	AF200465 (bp)	Sequence ^b	Allele Distribution ^c	Sample Size
1	Exon 3	69,067	TTACA <u>A</u> CTGTC	AA 83 AG 4 GG 3	90
2	Intron 3	69,215	TGTAG <u>T</u> AGCAG	TT 86 TC 4 CC 0	90
3	Intron 4	76,354	CATGCC <u>C</u> ATTGA	CC 23 CT 1 TT 0	24
4	Intron 4	77,623	AGCAG <u>C</u> TACAC	CC 21 AG 0 GG 3	24
5	Intron 4	77,696	CCTTG <u>A</u> TACAT	AA 23 AT 0 TT 1	24
6	Intron 4	77,802	TTGTT <u>A</u> AATTT	AA 14 AT 8 TT 2	24
7	Intron 4	77,844	CAATAC <u>C</u> ATACT	CC 22 CT 1 TT 1	24
8	Exon 5	77,895	ATGACT <u>T</u> TCATC	TT 23 AT 1 AA 0	24
9	Intron 6	78,698	TGTATA <u>A</u> CCAAA	AA 17 AG 6 GG 1	24
10	Intron 6	78,856	GTTGT <u>G</u> CCCCG	GG 23 GT 0 TT 1	24

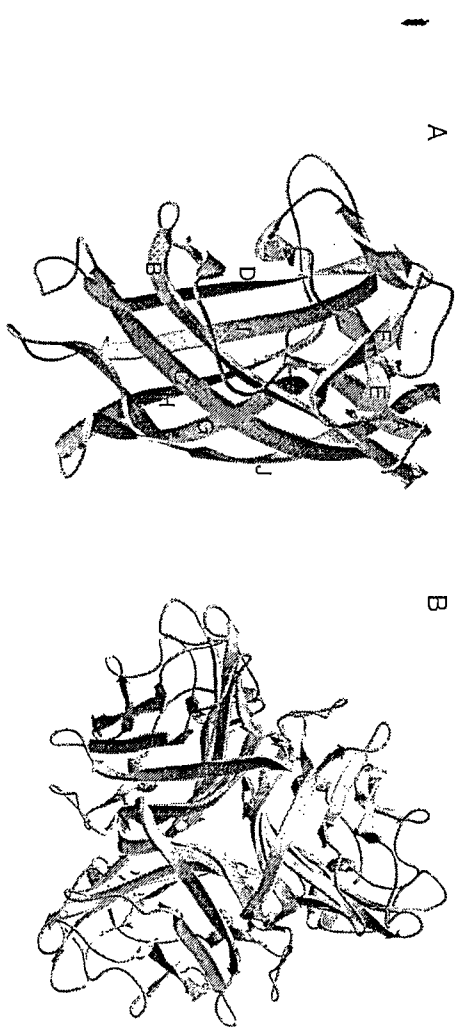
^aPosition of the SNP with respect to CXARD1 gene features is given (see also Figure C).

^bNucleotide sequence surrounding the SNP is given; central polymorphic nucleotide is in **bold** and is underlined.

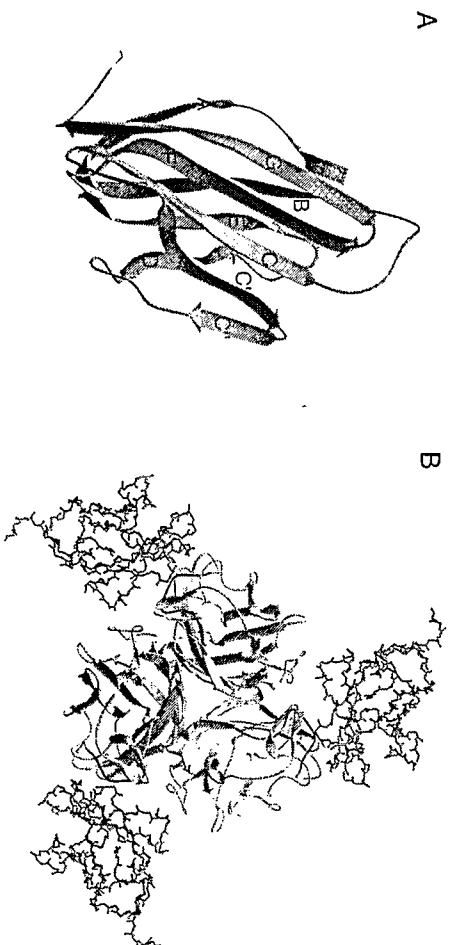
^cThe distribution of alleles among the analyzed individuals is given; e.g. for SNP1 in exon 3, 83 individuals were homozygous for the major allele and had A at the polymorphic position in both chromosomes; 4 individuals were heterozygous with A in one chromosome and G in the other; 3 individuals were homozygous for the minor allele and had G at the polymorphic position on both chromosomes.

^dNumber of individuals analyzed.

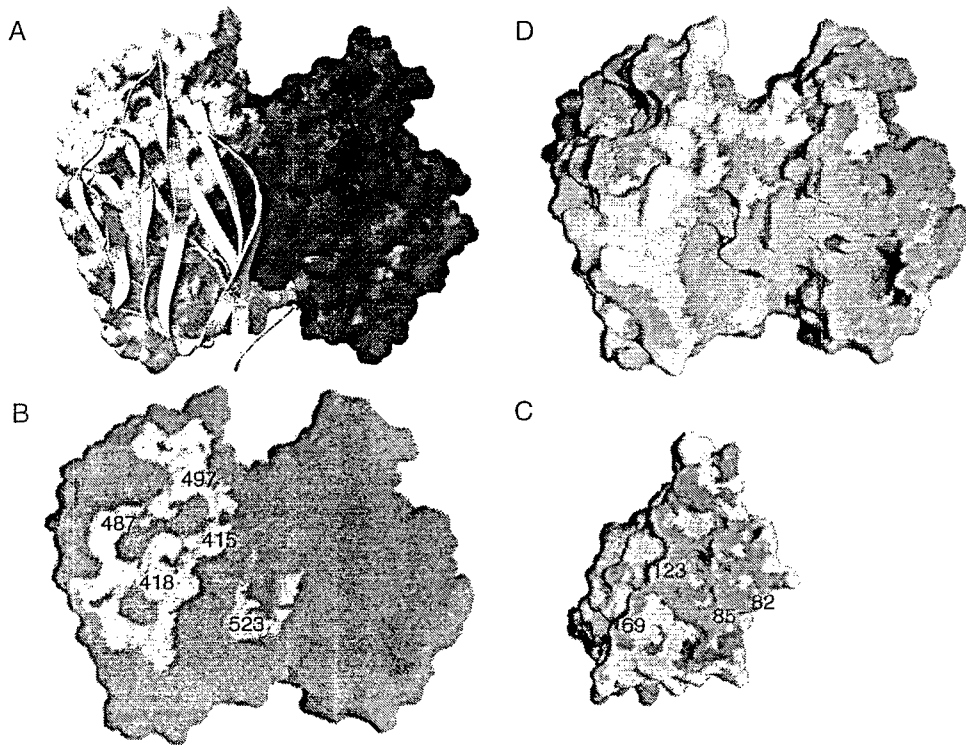
Freimuth Figure 1.



FREIMUTH FIGURE 2.



FREIMUTH FIGURE 3.

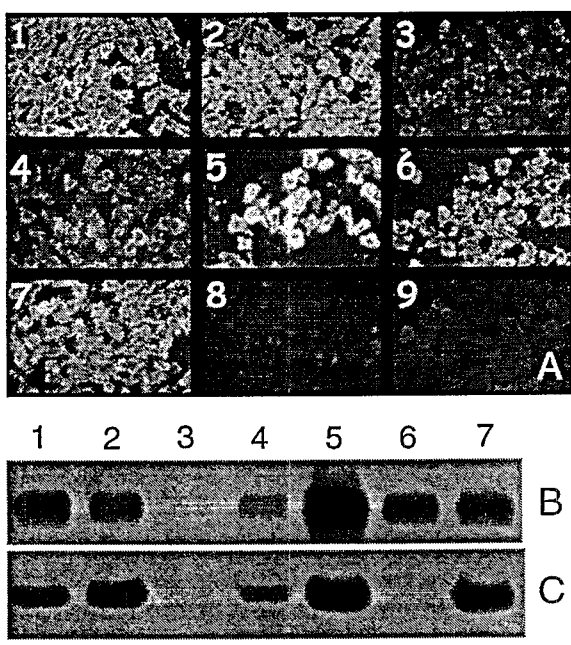


FREIMUTH FIGURE 4.

Ad2 LTLWTFPSPNCRH
Ad9 RTLWTFPTSPNCKID
Ad12 LTLWTFPPPNCSLI
Ad3 NTLWTGPKPEANCIE
**** * . . ** :

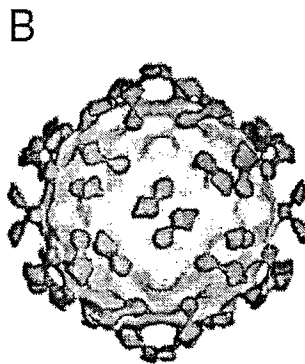
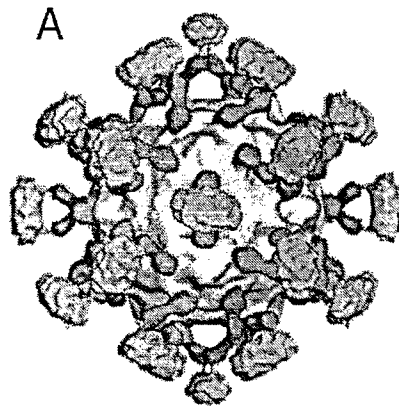
~~Figure 5 cell.tif~~

FRIMUT17 FIGURE 5



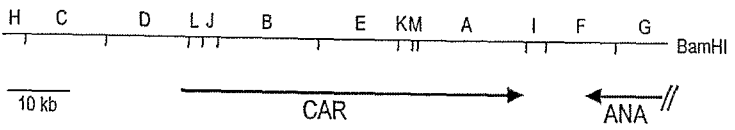
FREEMAN FIGURE 6

Fig. 6



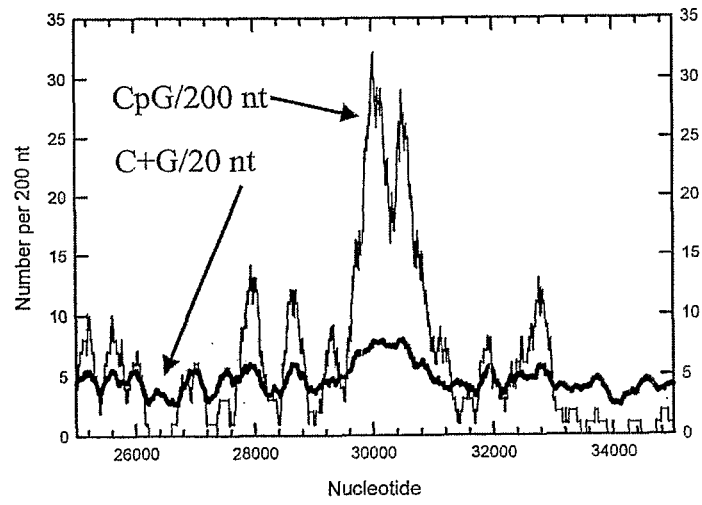
Freimuth Figure 7.

Fig 7



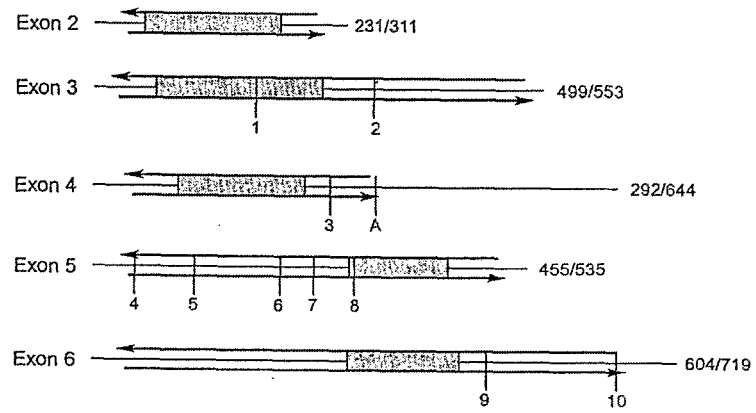
FREIMUTH FIGURE 8.

F-5 8



FREEMUTH FIGURE 9.

Fig 9



Acknowledgement

This work was supported in part by a National Institutes of Health grant (AI36251) and the Office of Biological and Environmental Research of the U.S. Department of Energy.