Three-Dimensional Structure of the Adenovirus Major Coat Protein Hexon

MICHAEL M. ROBERTS, JANICE L. WHITE,* MARKUS G. GRÜTTER,† ROGER M. BURNETT‡

The three-dimensional crystal structure of the adenovirus major coat protein is presented. Adenovirus type 2 hexon, at 967 residues, is now the longest polypeptide whose structure has been determined crystallographically. Taken with our model for hexon packing, which positions the 240 trimeric hexons in the capsid, the structure defines 60% of the protein within the 150 × 10^6 dalton virion. The assembly provides the first details of a DNA-containing animal virus that is 20 times larger than the spherical RNA viruses previously described. Unexpectedly, the hexon subunit contains two similar β-barrels whose topology is identical to those of the spherical RNA viruses, but whose architectural role in adenovirus is very different. The hexon structure reveals several distinctive features related to its function as a stable protective coat, and shows that the type-specific immunological determinants are restricted to the virion surface.

ADENOVIRUSES ARE NONENVELOPED viruses that have been identified in several mammalian hosts (*1*). Human adenoviruses primarily cause respiratory diseases, have the capacity to induce tumors in rodents, and have provided a well-studied model system for molecular biology. The simplicity of the icosahedral shape of the virion is in contrast to its complex architecture involving at least ten different constituent proteins and a single copy of linear double-stranded DNA. The 252 capsomeres are visible on the virion surface in the electron microscope and are named pentons and hexons to reflect the number of their neighbors. Twelve penton complexes are located at the vertices so that their pentameric penton bases (*2*) complete the outer shell of 240 hexons, and their projecting trimeric fibers (*2*) give adenovirus its characteristic appearance. Adenovirus is large, with an insphere radius of 450 Å from the particle center to the outer facet surface defined by the hexons (*3*). The size has so far precluded a direct structural attack and resulted in our current approach combining x-ray crystallography and electron microscopy (*4*).

Adenovirus type 2 (ad2) hexon is a trimer (*5*) containing three polypeptide chains of 967 amino acids (109,077 daltons) (*6*). Ad2 hexon crystallizes in the cubic space group $P2_13$, a = 150.6 Å, with one subunit per asymmetric unit (*7*). Electron density maps have been calculated with x-ray diffraction data from native crystals and five heavy atom derivatives at *p*H 5.0 (*4*). A 6-Å model (*8*) showed that hexon consists of two distinct parts: a triangular "top" 64 Å in height and containing three towers, and a pseudo-hexagonal "base" 52 Å in height and containing a central cavity. The pseudo-hexagonal symmetry of the base provides two kinds of vertical hexon-hexon contact faces, which alternate around the central molecular threefold, or pseudo-sixfold, axis. These are the A face, situated under each tower, and the B face (Fig. 1). Electron micrographs of capsid fragments revealing the orientation of the triangular hexon tops have led to a capsid model with a small *p*3 crystal of 12 hexons, with A to B contacts and tops outermost, at each facet (*3, 4*) (Fig. 2). The precise details of this close-packed arrangement are being refined by further structural studies (*9*).

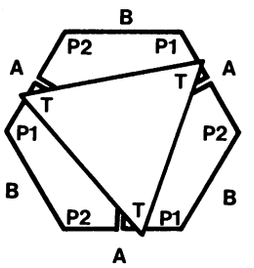

Fig. 1. A symbolic representation of the hexon trimer. A triangular top with three towers is superimposed on a pseudo-hexagonal base, with its vertices offset by 10° from the midpoints of the vertical A hexon-hexon basal contact face. Domains P1 and P2 lie in the base and domain T forms the tower, which lies above the subunit interface in the base (Fig. 5b). The A face is formed from parts of domains P1 in one subunit and P2 in a neighboring subunit. The B face is composed of the remainders of P1 and P2, both within one subunit.

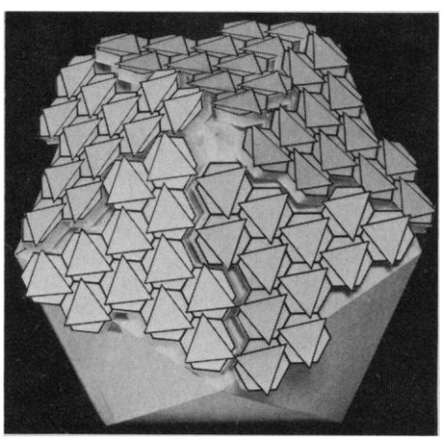

Fig. 2. A model adenovirus capsid, showing the organization of 12 hexons in small *p*3 nets on each icosahedral facet. The vacant position at each vertex is filled by a penton complex. The translational and icosahedral symmetry combine to give a high degree of identical chemical contacts for the four hexon molecules in topographically distinct positions (*3*).

An initial chain tracing was made from a 2.9-Å resolution electron density mini-map comprising 120 skew sections calculated on a 2 Å cm^{-1} scale normal to the molecular threefold axis (*4, 8*). The electron density was relatively clear in the base, where there is a greater proportion of secondary structure, but the weak regions at the top were difficult to interpret. Various segments of chain containing a total of 820 residues were traced, but identification of the residues and assignment of an overall topology was not immediately possible. This was so despite the apparent reliability of the phase information, as indicated by the small change in the figure of merit upon adding the fifth derivative (*8*). The main difficulty stemmed from the structure itself. Hexon has little secondary structure, its three chains are tightly interwoven at the subunit interfaces, and the faint electron density in the towers indicates mobility (*8*). In contrast, electron density maps of complete virus particles are superb as a result of averaging methods, and reveal coat proteins with disorder only at the NH_2-termini.

We recently found a common origin between the map and the amino acid sequence and have now solved the remaining problem of the overall relations among the previously traced chain segments. Segments of chain were tentatively assigned to a position in the sequence based on the shapes of side-chain density. Since the appearance of a particular residue was variable, the absence of density for glycine was an excellent positional indi-

M. M. Roberts, Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, 630 West 168 Street, New York, NY 10032.
J. L. White and M. G. Grütter, Abteilung Strukturbiologie, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.
R. M. Burnett, Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, 630 West 168 Street, New York, NY 10032, and Abteilung Strukturbiologie, Biozentrum der Universität Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

*Present address: Department of Biochemistry, University of Sheffield, Sheffield S10 2TN, England.
†Present address: CIBA-GEIGY AG, CH-4002 Basel, Switzerland.
‡To whom correspondence should be addressed at Columbia University.

cator. Additional factors were assessed as the positioned segments accumulated. For example, in β-sheets the hydrogen bonds of one strand must join to the neighboring strands at the proper locations. The hydrogen bonds within parallel β-sheets have a distinctly different pattern from those within antiparallel sheets, which are more commonly exposed to solvent (*10*). This was a useful guide in assigning chain direction, since hexon has little α-helix and its basal surfaces are almost exclusively β-sheet. Further clues were provided by the emergence of hydrophobic pockets, enclosed by β-strands of alternating hydrophobic and hydrophilic residues, for which proper closure favored correct sequence assignments. Additional confirmation was given by mercury binding near five of the seven cysteine residues, and the locations of residues differing in ad5 hexon (*11*). Sequence alterations are unlikely inside the molecule or in regions involved in hexon-hexon contacts.

Tentative stretches that violated any of the above criteria were viewed with reservation, as were those that matched different parts of the sequence equally well. Relocations were common for these stretches before the constraint of molecular connectivity imposed the correct solution. Plastic Nicholson molecular models were molded to fit the map features to ensure the presence of all residues and to aid in the estimation of their coordinates. The final tracing, which incorporated 859 residues (Figs. 3 and 4), showed that comparatively modest changes in the initial segments were necessary. There are five internal gaps (90 to 94, 194 to 213, 270 to 291, 360 to 364, and 444 to 454), and the NH_2- and COOH-termini are disordered (1 to 41 and 964 to 967). Although the three longer gaps potentially introduced ambiguities in subunit assignment, their position at the top surface of the towers and their relatively small length preserved contiguity (Fig. 4). The two short gaps in the base are probably artifacts that result from proximal heavy atom sites. The overall folding pattern is undoubtedly correct, but the difficult structure solution suggests that minor rearrangements may follow refinement.

The hexon polypeptide forms only 8% α-helix and 22% β-sheet, somewhat lower than the estimates of 27% and 26% obtained from circular dichroism (*12*). The extensive contact between the subunits in the hexon base develops into an extraordi-

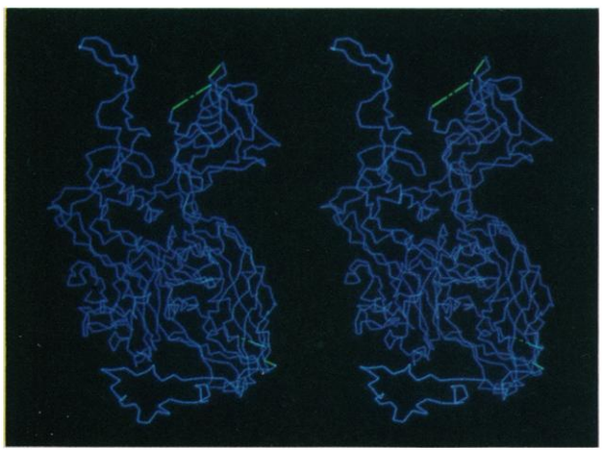

Fig. 3. A stereo pair (*17*) showing the main chain (blue) connecting the Cα positions for one subunit of the hexon trimer. Green dashes represent gaps in the tracing. The view is from the side and normal to the molecular threefold axis.

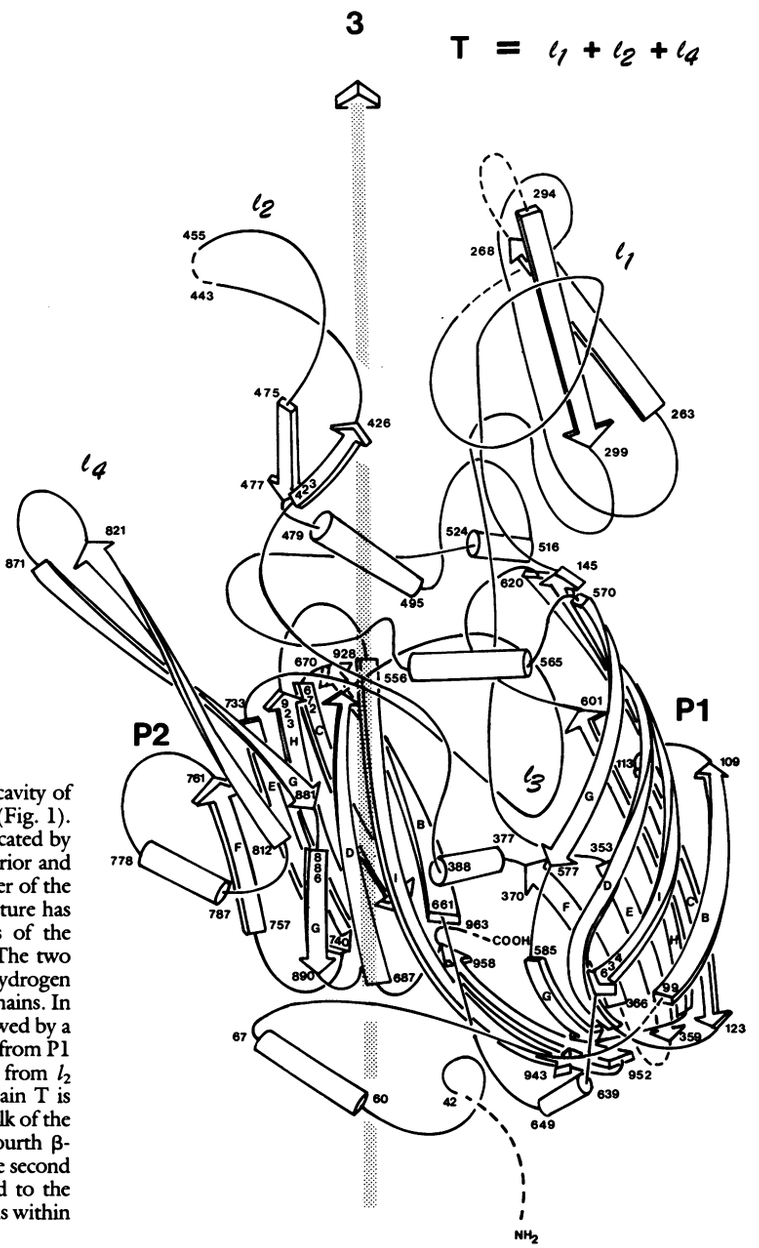

Fig. 4. A sketch of the hexon subunit viewed from within the central cavity of the trimeric molecule, looking toward the B hexon-hexon contact face (Fig. 1). Most of the chain lies behind the vertical threefold molecular axis (indicated by 3). The bottom of the molecule lies toward the adenovirus capsid interior and the top forms its external surface. The NH_2-terminus is at bottom center of the picture, with the COOH-terminus slightly above. The secondary structure has been overemphasized for clarity by drastically reducing the lengths of the connecting loops. Gaps in the tracing are indicated by dashed lines. The two domains P1 and P2 overlap through a very long β-strand forming hydrogen bonds with both domains. Several structural features appear in both domains. In each β-barrel, an α-helix precedes the first strand, and an α-helix is followed by a tower loop between the fifth and sixth strands. These loops are l_2 rising from P1 and l_4 from P2, and both contain two antiparallel β-strands, those from l_2 forming the β-constriction about the molecular axis. The tower domain T is formed from three loops (l_1, l_2, l_4), each from a different subunit. The bulk of the tower is provided by l_1, an excursion in P1 between the third and fourth β-strands. Its counterpart in P2 is l_3; although not as extensive, it is still the second longest connecting bridge of this barrel (Fig. 6). Loop l_3 is restricted to the upper part of the base, where it stabilizes the interaction of the P domains within one subunit.

nary interpenetration in the towers, each of which is formed from three different loops, one from each subunit (Figs. 4 and 5). This feature explains hexon's extreme resistance to denaturation (*1*), which is consistent with its role as a structural building block (*3*, *4*) and favors its description as a complete molecule. The trimer is formed from three copies of each of the two pedestal domains, P1 and P2, in the base and domain T in the tower (Figs. 1, 4, and 5). Domains P1 and P2 are very similar, each consisting of an eight-stranded flattened β-barrel arranged in the same "jellyroll Greek key" topology (*10*) as that found in the coat proteins of small spherical plant viruses (*13*, *14*), rhinovirus (*15*), and poliovirus (*16*). This precise topology is unique to viral proteins. Domains P1 and P2 are related almost precisely by a pseudo-sixfold axis (*17*) so that there is an almost exact hexamer of barrels. Each tower domain is formed not only by loops l_1 and l_4 rising from P1 and P2 in adjacent subunits at their interface, but also by loop l_2 rising from P1 in the opposed third subunit through the "β-constriction" (Figs. 4 and 5). The β-constriction arises from three l_2 loops that narrow to form an antiparallel six-stranded β-tube around the central threefold axis.

The function of the tower may be to increase the area of the interface between two hexon subunits and to anchor this pair to the third subunit to stabilize the whole molecule. A lower molecular height provides sufficient strength within the covalently bonded subunit, and so the hexon's curious shape reflects its function as a highly stable building block (*3*, *4*). The degree of interweaving between the hexon subunits is quite extraordinary as it involves internal loops rather than termini. This could be the basis for the hexon's unique requirement for another protein (adenovirus 100K) in its folding pathway (*18*). The β-constriction is analogous to the interwoven NH_2-termini around the threefold axes in tomato bushy stunt virus and southern bean mosaic virus (*13*, *14*) and around the fivefold axes in rhinovirus and poliovirus (*15*, *16*). In each virus, a β-structure stabilizes the association of subunits into larger units. This mechanism, which decreases the number of independent units and so gives greater strength to the whole assembly, is emerging as a general feature of viral architecture. The NH_2-terminal chain of hexon, from residues 42 to 89, is tucked under a neighboring subunit and the final 41 residues are invisible (Figs. 4 and 5). The disordered NH_2-terminus would lie to the interior of the capsid, as in the RNA viruses. Its role in adenovirus may be to fasten hexon to the underlying protein shell formed from polypeptide VI (*1*). Such "anchoring" by a terminus may also be general, as the NH_2-termini in the RNA viruses apparently interact with the nucleic acid core.

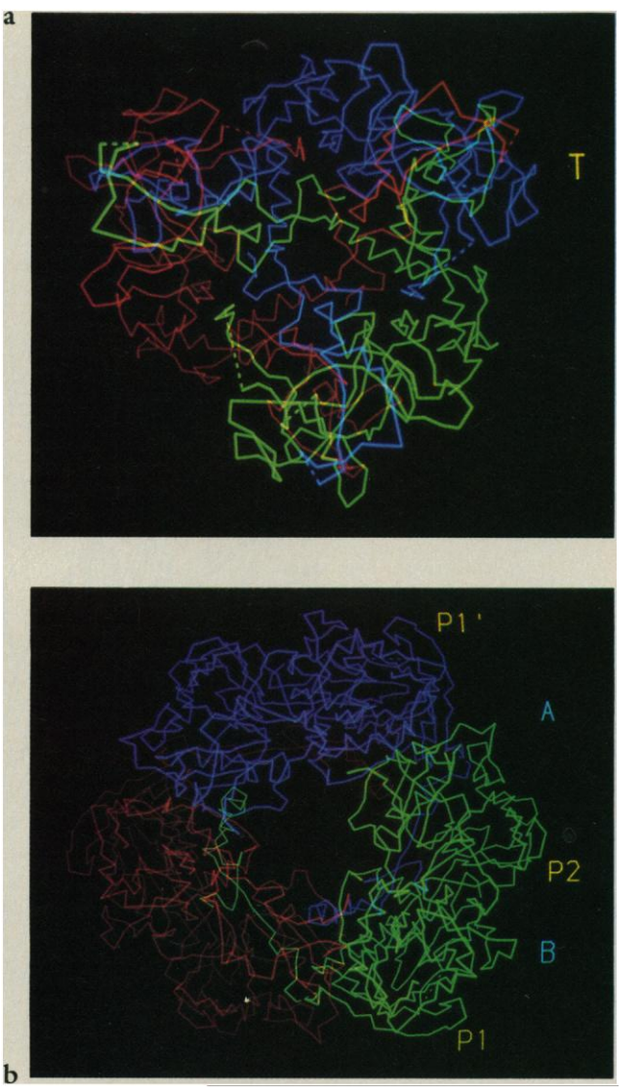

Fig. 5. Views of the top (a) and base (b) of the hexon trimer in the same relative orientations and viewed down the molecular threefold axis. The top of the molecule has been separated from the base by cutting the three tower loops just above the β-barrels (l_1: 146 to 334; l_2: 417 to 534; l_4: 819 to 875) (Fig. 4). Each subunit is differently colored, and dashed lines indicate connections across gaps in the traced sequence. (a) The top contains three towers, each formed by contributions from all three subunits, making the top of hexon a highly interpenetrated structure. A distinctive feature is the symmetric "β-constriction" seen at the center, consisting of six antiparallel β-strands formed by three l_2 loops. These strands spiral upward in a right-handed fashion from the center into the towers. (b) In the base, the interface between two neighboring subunits is more distinct, with little interpenetration except for the NH_2-terminal chain, which wraps around the base of a neighboring subunit before disappearing from view. The two different corners of the pseudohexagonal base are formed by P1 and P2, with the corner angle of 120° defined by the offset of the two sets of four vertical strands of each β-barrel. The scale factor for (a) is about 20% larger than that for (b).

The similarity of domains P1 and P2 suggests gene duplication, but this must have been an ancient event as there is no obvious homology even between residues in related spatial positions in the two β-barrels (*17*). However, the similarity is emphasized by the occurrence of similar structural features, such as α-helices and long loops, between corresponding strands (Fig. 6). In contrast, despite the basic similarity of their β-barrels, each RNA virus exhibits a separate pattern of additional features, although within rhinovirus or poliovirus the different β-barrels are related. It is noteworthy that in adenovirus the β-barrel axes are perpendicular rather than parallel to the virion surface, so that the β-barrels form the walls of the tubular hexon base. In both cases the β-barrel provides a shell of protein. In the RNA viruses the shell is the capsid, whereas in adenovirus the shell is the tubular base. Thus, although β-barrels are so far universally found in spherical viruses, their role apparently is not to form complete structure units (*3*, *4*), but to act as elementary components in their construction. Support for this primitive function is provided by the close topological similarity between all the viral β-barrels, strongly suggesting their evolution from a common precursor, and by the preservation of curvature implied by corresponding strands forming the outer surface of both the RNA virions and the hexon molecule.

The structure reveals the basis for several earlier biochemical observations. The sequence suggested that the middle third of the polypeptide chain is different from the first and last third portions (*19*). Proline occurs more frequently in the first and last thirds, and trypsin-sensitive sites are located in the first third. The first third corresponds to a region containing the first portion of P1 and l_1, while the last third includes all of P2

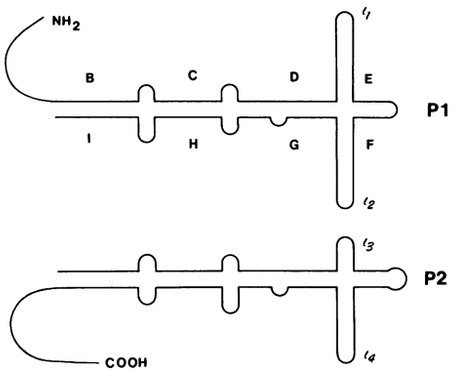

Fig. 6. A schematic representation of the P1 and P2 domains. The strands have been lettered according to the convention of tomato bushy stunt virus (*13*). The G strands show a discontinuity in their hydrogen bonding (see Fig. 4). The intervening small loops between the strands have lengths corresponding to less than 10, or 10 to 20 residues, and the lengths of the longer loops are 207 (l_1), 199 (l_2), 35 (l_3), and 109 (l_4) residues. Loops l_1, l_3 and l_2, l_4 emerge from the β-barrels in corresponding positions. Intervening loops in the RNA viruses occur between strands C:D, E:F, and G:H, so that they emerge from their β-barrels at the opposite end from that used in hexon.

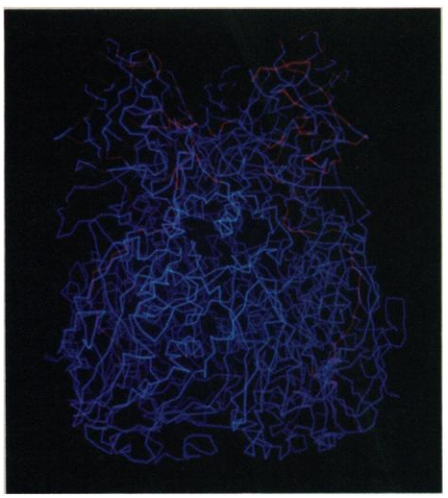

Fig. 7. A side view of the hexon trimer, showing regions of amino acid homology (blue) and heterology (red) between hexons from adenovirus types 2 and 5. The heterologous regions all lie on exposed regions at the top of the molecule, with the exception of a small strand on the surface of the base (138 to 148). The capsid model (Fig. 2) indicates that this stretch, as with the other heterologous regions, is not involved in hexon-hexon contacts.

and l_4. In both regions, prolines play a crucial role in chain folding. Six of the eight strands in each β-barrel, and both strands in l_1, are terminated by proline at one or both ends. Proline permits very short bridges between strands (*10*), and four out of the eight Pro-X-X-X-Pro sequences are used in this manner. A total of 19 proline pairs are separated by five or fewer residues: l_1 contains eight while P1 and P2 each contains three. The proline-poor middle third of the sequence corresponds to l_2 and the final part of P1, which together contain five of the hexon's seven α-helices. Proline pairs have been found to form bridges in both viral and nonviral protein structures, and are particularly frequent in viral sequences.

The trypsin-sensitive sites are exposed on the surface of the molecule as expected (*19*) but the small piece of chain between the sites at Arg^{142} and Arg^{165} is within the molecule, at the beginning of the l_1 loop (Fig. 4), and probably would not be released upon digestion. The chain connecting these sites contains a highly acidic stretch of 16 consecutive Glu or Asp residues (146 to 161), which are surrounded by 10 adjacent basic residues from elsewhere in the chain. This region may provide the conformational change that causes cracking of the crystals if they are raised to *p*H 5.5, the point at which dissociation of hexon shells also occurs (*8*). The regions of heterology between ad2 and ad5 hexon, the two available human adenovirus sequences (*6*, *11*), are restricted to the top of the molecule (Fig. 7). The capsid model (*3*) (Fig. 2) shows that changes here provide an altered virion surface without disturbing residues involved in hexon-hexon contacts.

Model building of the structure will yield detailed information about the hexon-hexon capsid interactions, whose overall features have been derived by electron microscopy. As ad5 hexon crystallizes isomorphously to ad2 hexon, and diffracts synchrotron radiation to 1.9-Å resolution, difference maps will also reveal the structural nature of the immunological differences between the two adenovirus species (Fig. 7). It is already clear that the type-specific differences are restricted to the outside of the capsid, and that the hexon-hexon contact regions in the base are completely conserved (Fig. 7). The emerging structure of adenovirus should act as a stimulus for further biochemical and immunological studies on this complex particle.

REFERENCES AND NOTES

1. H. S. Ginsberg, Ed., *The Adenovirus* (Plenum, New York, 1984).
2. J. van Oostrum and R. M. Burnett, *J. Virol.* **56**, 439 (1985).
3. R. M. Burnett, *J. Mol. Biol.* **185**, 125 (1985).
4. ———, in *Biological Macromolecules and Assemblies*, vol. 1, *Virus Structures*, F. A. Jurnak and A. McPherson, Eds. (Wiley, New York, 1984), p. 337.
5. M. Grütter and R. M. Franklin, *J. Mol. Biol.* **89**, 163 (1974).
6. G. Akusjärvi *et al.*, *J. Biol. Chem.* **259**, 13976 (1984).
7. R. M. Franklin, U. Pettersson, K. Åkervall, B. Strandberg, L. Philipson, *J. Mol. Biol.* **57**, 383 (1971).
8. R. M. Burnett, M. G. Grütter, J. L. White, *ibid.* **185**, 105 (1985).
9. J. van Oostrum and R. M. Burnett, *Ann. N.Y. Acad. Sci.* **435**, 578 (1984); J. van Oostrum, P. R. Smith, M. Mohraz, R. M. Burnett, unpublished observations.
10. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
11. R. Kinloch, N. Mackay, V. Mautner, *J. Biol. Chem.* **259**, 6431 (1984).
12. L. A. Day, R. M. Franklin, U. Pettersson, L. Philipson, *Eur. J. Biochem.* **29**, 537 (1972).
13. A. J. Olson, G. Bricogne, S. C. Harrison, *J. Mol. Biol.* **171**, 61 (1983).
14. M. G. Rossmann *et al.*, *ibid.* **165**, 711 (1983).
15. M. G. Rossmann *et al.*, *Nature (London)* **317**, 145 (1985).
16. J. M. Hogle, M. Chow, D. J. Filman, *Science* **229**, 1358 (1985).
17. P1 was superposed on P2 by a least-squares minimization of the C_α-C_α distances of spatially corresponding residues using the program RIGID (R. M. Burnett). For 96 such pairs, the average separation was 3.5 Å and the superposition required a rotation of 57.4° about an axis inclined at 25.0° to the molecular threefold axis. The matching was aided by displaying the structures using the computer graphics programs GRAMPS [T. J. O'Donnell and A. J. Olson, *Comput. Graph.* **15**, 133 (1981)] and GRANNY [M. L. Connolly and A. J. Olson, *Comput. Chem.* **9**, 1 (1985)]. These programs were also used to produce the color pictures.
18. C. L. Cepko and P. A. Sharp, *Cell* **31**, 407 (1982).
19. H. Jörnvall *et al.*, *J. Biol. Chem.* **256**, 6181 (1981).

20. We thank Richard Feng and Perry Kim of the Bronx High School of Science Physical Science Project for their computational assistance in recording and checking the measured coordinates. We also thank Jan van Oostrum and Janet L. Smith for many helpful discussions. Supported by grants from the Swiss National Science Foundation, Public Health Service grant AI 17270 from the National Institute of Allergy and Infectious Diseases, and an Irma T. Hirschl Career Scientist Award to R.M.B.

5 November 1985; accepted 7 April 1986