Molecular Structure of the Cell-Attachment Protein of Reovirus: Correlation of Computer-Processed Electron Micrographs with Sequence-Based Predictions

R. D. B. FRASER,¹ D. B. FURLONG,² B. L. TRUS,³ M. L. NIBERT,² B. N. FIELDS,² AND A. C. STEVEN^{4*}

Fogarty International Center,¹ Computer Systems Laboratory, Division of Computer Research and Technology,³ and Laboratory of Physical Biology,⁴ National Institute of Arthritis, Musculoskeletal and Skin Diseases, Building 6, Room 114, Bethesda, Maryland 20892, and Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115²

Received 27 November 1989/Accepted 16 February 1990

The receptor-recognition interaction that initiates reovirus infection is mediated by the sigma 1 protein, located at the vertices of the icosahedral virion. We have applied computer-based image-averaging techniques to electron micrographs of negatively stained preparations of sigma 1 purified from virions (serotype 2 Jones). Combining these results with inferences based on the amino acid sequence has led to a molecular model in which the overall folding of the chains is described; its conformation embodies motifs, coiled-coil α -helices and nodular multichain elements rich in β -sheets, previously detected in the corresponding proteins of other viruses, but with some novel variations. Sigma 1 is a filamentous lollipop-shaped molecule with an overall length of ~ 48 nm; it has a flexible "tail," ~ 40 nm long by 4 to 6 nm wide, terminating at its distal end in a globular "head," ~ 9.5 nm in diameter. The purified protein is a tetramer (4 by 50 kilodaltons) consisting of two similarly oriented dimers bonded side by side and in register. For each chain, a cluster of hydrophobic residues at its amino terminus resides at the proximal end of the tail; next, an α -helical domain (residues 25 to 172) participates in a two-chained coiled coil, 22 nm long, with two such coiled coils pairing laterally to form the proximal half of the tail. The remainder of the tail (residues 173 to ~ 316) is less uniform in width and is expected to be rich in β -sheet; the interdimer bonding is evidently sustained through this portion of the molecule. Finally, the globular head consists of the carboxy-terminal domains (which contain the receptor-binding sites) folded into compact globular conformations; in appropriate side views, the head is resolved into two subunits, presumably contributed by the respective dimers. This model for how the four $\sigma 1$ polypeptide chains are threaded in parallel through the fiber is supported by the observed match between an empirical curvature profile, which identifies the locations of relatively flexible sites along the tail, and the flexibility profile predicted on the basis of the model. Appraisal of the interactions that stabilize the coiled coils suggests that (i) the α -helices are individually only marginally stable, a property that may be of significance with regard to the retracted conformation in which sigma 1 is accommodated in the intact virion, and (ii) the predominant interactions between the two coiled coils are likely to involve hydrogen bonding between patches of uncharged residues.

A function essential to all viruses is a mechanism whereby the virus may recognize a susceptible cell and attach specifically to it, thereby initiating the infection process. The molecular basis for this recognition event is stereochemical complementarity between some part of the virus and some surface component of the cell (the virus receptor). For small nonenveloped viruses, this interaction is mediated by a site on its capsid protein; for enveloped viruses, this interaction is mediated generally by its membrane glycoprotein(s); for larger icosahedral bacteriophages, it is mediated by specialized fibrous proteins that are clustered around a single specialized vertex of the icosahedral capsid; and for at least some of the more complex, nonenveloped, animal viruses (e.g. adenovirus, reovirus), the interaction is mediated by a fibrous protein that is located at each of the 12 vertices of the icosahedral capsid. An improved understanding of the specificity of the virus-host interactions is of interest not only since they represent an important and widespread class of macromolecular recognition reactions, but also from the viewpoint that these reactions may be targeted in an approach to antiviral drug therapy (38).

The cell-attachment protein of reovirus, called sigma 1 ($\sigma 1$), has been purified after being released from virions by exposure to elevated temperatures (3) or treatment with urea (40). The purified protein is a homo-oligomer, whose monomers have molecular masses of 49 to 51 kilodaltons (kDa) in different strains of reovirus (25). Electron microscopy of $\sigma 1$ serotype 2 (Jones) has revealed flexible, filamentous, asymmetric particles with a globular head at one end (12). Cloning-derived $\sigma 1$ sequences have been reported for all three serotypes of reovirus (2, 4, 22, 25), revealing a substantial degree of interstrain homology. An amino acid sequence located toward the amino terminus has been found to contain the classical heptad repeat (*a-b-c-d-e-f-g*, with residues *a* and *d* predominantly hydrophobic) indicative of a polypeptide chain with a propensity to form an α -helical coiled-coil rope (6), and this conformation has been proposed for the tail portion of sigma 1 (2, 12).

In the present study, we have applied computer processing techniques to improve the yield of structural information from negatively stained electron micrographs. Specifically, we have applied a straightening procedure (30) that allows one to compensate for the random curvature exhibited by individual fibers. After classifying the resulting images by

* Corresponding author.

multivariate statistical analysis (34, 36), image averaging was performed to improve interpretability. In a complementary approach, a more detailed analysis of the amino acid sequence was performed with a view to identifying such features as tracts of regular secondary structure, concentrations of particular amino acids, etc., which may denote domainal divisions. These two lines of evidence have been integrated into a self-consistent model of the $\sigma 1$ fiber.

MATERIALS AND METHODS

Purification of the protein. Reovirus serotype 2 (Jones) was propagated and purified, and $\sigma 1$ protein was, in turn, isolated from virions as described by Furlong et al. (12).

Electron microscopy. Negative staining of purified protein adsorbed to a thin carbon support film was carried out according to a modification of the method of Valentine and Green (35), as described earlier (12). Micrographs were recorded at a nominal magnification of $\times 50,000$ on a JEOL 100B electron microscope and calibrated relative to a carbon film line-grating replica of 21,600 lines per cm (Ladd Research Industries, Inc., Burlington, Vt.).

Image processing. Electron micrograph negatives were digitized by using a Perkin-Elmer 1010MG scanning microdensitometer. The step of the scanning raster was equal to the dimension of the square aperture and was chosen to be equivalent to 0.4 nm at the specimen. General image processing and display operations were performed using the PIC system of programs (33) implemented on a VAX 8350 computer (Digital Equipment Corporation, Maynard, Mass.) with a model 8500 image processor (Gould Imaging & Graphics, Fremont, Calif.). Photographic records of digital images were recorded by means of a model 4007 camera station (Matrix Instruments Inc., Northvale, N.J.).

Particles were selected for analysis according to the following criteria. They had to be (i) fully embedded in a smooth (i.e., nongranular) stain layer, (ii) not evidently kinked or broken (the straightening algorithm can only be expected to compensate reliably for continuous curvature), and (iii) free standing, i.e., not complexed with other particles, crossing them, or only partially present in the micrograph. Straightening was effected by cubic-spline interpolation onto a curvilinear coordinate system defined by the axis of the molecule under consideration (30). Straightened particles were brought into translational alignment by cross-correlation methods. A uniform normalization was then imposed, scaling the images to have the same average and the same dynamic power.

Prior to averaging, the images were further analyzed by statistical classification procedures in order to ensure that the particles to be averaged were genuinely alike. For this purpose, methods of multivariate statistical analysis were used, applying procedures described in detail elsewhere (34). In brief, the primary technique used was a form of supervised classification whereby sets of archetypal reference images were chosen, each set was separately averaged, and the remaining images were assigned to one of these reference sets, provided that they mapped more closely to one template than to the others by a margin that exceeded a preset level of significance. (Here interimage distances were quantified by the sum of the squares of the differences between corresponding pixels). Of 114 fiber images analyzed, 32 were determined to be type A and 34 were type B (see Results). Finally, all images in each set of interest were averaged. To ensure reproducibility of the results, similar analyses were conducted, using different template particles and initial

selections made by different observers. In addition, the data were subjected to factorial analysis (36), according to the principal components formalism (34).

Dimensions given for various structural features were determined according to the following edge-detection criterion (31). After averaging, transverse density traces were calculated from the digital images, in which the signal-to-noise ratios were further improved by lateral averaging over appropriate distances (i.e., ~ 18 nm in the distal tail regions and 1.5 nm for globular features). The width of the particle was determined as the spacing between the points on either side at which the density fell to 25% of the central maximum value.

A curvature profile was calculated for each particle as a by-product of the straightening procedure. At each point along its center line, parameterized by its cubic spline coefficients, the curvature was calculated as $\Delta\theta/\Delta s$, i.e., the change, from pixel to pixel, in the slope of the tangent to the curve. The average curvature profile was then calculated for the entire data set. To establish a baseline of significance for curvature values that are higher than fluctuations associated with random experimental errors, a randomly chosen particle was straightened eight different times and the curvature profile calculated on each occasion. The standard deviation for the eight curvatures thus obtained was calculated at each point. These data were then averaged over the length of the particle, and the resulting value was used to set the baseline described in the legend to Fig. 5c.

Computational analysis of amino acid sequences. Secondary structure predictions (5, 13), hydropathy profiles (19), flexibility profiles (18), calculation and display of various kinds of helical projections (see Results), and other sequence-related analyses were performed either on a VAX 8350 (see above) or on an IBM PC-AT microcomputer. Use was also made of the AASAP program kindly made available by D. A. D. Parry.

RESULTS

Ultrastructural analysis of purified $\sigma 1$ protein. A field of negatively stained fibers is shown in Fig. 1. Most of the fibers are randomly dispersed but we also occasionally observed spherical clusters resembling "dandelion clocks" (Fig. 2a), in which fibers radiate from a central point at which the tips of their tails meet. Similar aggregates have also been observed in $\sigma 1$ preparations derived from a recombinant vaccinia virus expression system (1). Other recurring morphological types include particles that appear to have a blob at both ends ($\sim 5\%$ overall; Fig. 2b); particles whose heads appear heart shaped and, in extreme cases, bilobed (5 to 10%; Fig. 2c); and particles which are sharply bent or kinked, most commonly close to the head-tail junction (15 to 20%; Fig. 2d), but sometimes near the midpoint of the tail (2 to 4%; Fig. 2e).

Images of individual fibers were selected for digital analysis according to the criteria described in Materials and Methods and then computationally straightened and aligned. By visual appraisal, the straightened particles seemed to include two obvious classes, particles with relatively broad and uniform tails (type A; Fig. 3a) and particles whose tails were perceptibly narrower at the proximal end (type B; Fig. 3b). A few particularly clear examples of both kinds of fibers were picked out by two observers independently, with $\sim 90\%$ consistency. The two sets were averaged separately and used as references to sort the remaining particles according to a supervised classification procedure (see Materials and Methods).

FIG. 1. Field of purified reovirus $\sigma 1$ fibers, negatively stained with uranyl formate. Bar = 50 nm.

In a second independent classification, the data were also analyzed according to the principal components formalism of multivariate statistics (34). In the two-factor plot shown in Fig. 4, each point represents an image of a fiber designated as type A or B according to the above classification. If there were to be only one class of images apart from random variations (i.e., the inferred classification was, in fact, spurious), the points would be randomly distributed in this plane. If, on the other hand, there were to be more than one class of images, each such class should segregate into a particular part of the plane. In fact, the two putative classes

do segregate in this way (Fig. 4), formally validating the initial classification, although the noise level is so high as to preclude their drawing fully apart into two well-resolved clusters. Accordingly, the two classes of images were averaged separately (Figs. 3c and d). For comparison and to give a completely objective representation, a global average image was also computed (Fig. 3e).

Morphology of the reovirus fiber. The overall length of the particle, as judged from the continuous stain-excluding profile, is 48 nm, and possibly is as great as 51 nm if the relatively diffuse extension at the proximal tip is included.

FIG. 2. Gallery of $\sigma 1$ fibers, showing illustrative examples of recurring morphological types. (a) Aggregates in which fibers bind by the tips of their tails; (b) fibers which appear to have bound additional molecules to their tips and so appear double headed; (c) fibers whose heads appear bilobed or heart shaped; (d) fibers which are kinked immediately adjacent to their heads; (e) fibers which have kinks near the midpoints of their tails. The low contrast of the original negatives was enhanced computationally after digitization. Bar = 50 nm.

FIG. 3. Computer-processed electron micrographs of negatively stained $\sigma 1$ fibers. (a and b) Typical examples of individual fibers after straightening and translational alignment, with panels a and b corresponding to two different morphological types of fiber (types A and B, respectively, see Results). In panel c, the top two images represent independent averages over two sets of fibers of type A, which are combined in the bottom image. Panel d shows corresponding results for fibers of type B. (e) Globally averaged fiber image, without regard to subclasses. Bar = 20 nm.

These values are considerably greater than the earlier measurements of fiber length (12). We infer that the difference is due largely to the straightening procedure used in the present analysis, which increases end-to-end separation while conserving contour length.

Morphologically, the fiber consists of a tail (T) and a head (H) (Fig. 5), with the former partitioning into T(i), a terminal appendage; T(ii), a rodlike zone ~ 22 nm long of uniform width that is resolved laterally into two equal rods in the type A fiber and is distinctly narrower in the type B fiber; T(iii), in which the rod widens into an extended nodule ~ 9.5 nm long by ~ 5.6 nm wide, ending in a constriction of reduced diameter; and T(iv), a smaller nodule, ~ 5.1 nm across,

FIG. 4. An illustrative set of 40 $\sigma 1$ particle digital images were cropped to minimal dimensions (120 by 20 pixels from 160 by 48 pixels) to reduce extraneous noise from surrounding background, and the set was analyzed by principal components factorial analysis. The two-factor plot, i.e., the mapping of these images according to the two major sources of interimage variability, is shown here. The particles were assigned to sets A and B according to a supervised classification procedure (see Materials and Methods). In this map, which is a more objective basis for classification, one sees a pronounced segregation of the two types across the plane, with only a few ambiguous cases.

terminating in a second constriction immediately before the head. Features T(iii) and T(iv) are more clearly visualized in the type B particle (cf. Fig. 3c and d).

Sites that exhibit enhanced curvature. Most fibers exhibit some curvature, and many are sharply kinked. When kinks occur, they are most commonly observed in the T(iv) region (Fig. 2d) or near the midpoint of the tail, i.e., near the junction between T(ii) and T(iii) (Fig. 2e). In a more statistically representative approach to identifying sites in the molecule that are relatively flexible (or, at least, deformable), a curvature profile was calculated for each particle as a by-product of the straightening procedure. The average curvature profile was then calculated for the entire data set (Fig. 5c). Peaks in this profile denote sites that are particularly liable to bending and therefore represent the locations of relatively labile or flexible structures. A broad maximum is found in the T(iv) region, with several other secondary maxima in other parts of the tail. Interestingly, the T(iv) maximum and the one located between T(ii) and T(iii) (Fig. 5c) correlate with the sites at which kinks are most frequently observed and thus confirm the significance of those observations.

Inferences from the amino acid sequence. The amino acid sequence of $\sigma 1$ from the serotype 2 D5/Jones strain of reovirus has been derived from the cloned gene sequence (4). This sequence, as amended at the carboxy terminus by the inclusion of an additional 62 residues (25), was analyzed for conformation prediction by using the procedures of Chou and Fasman (5) and Garnier et al. (13), which yielded very similar results. In addition, the hydropathy profile (19) was calculated, and a search was also made for the heptad pattern characteristic of α -helical coiled-coil rope structures. Our results confirm and extend the findings of Furlong et al. (12) as follows.

The sequence can be delineated into four distinct segments, S(i) to S(iv), as shown in Table 1 and Fig. 6. The N-terminal segment S(i), consisting of residues 1 to 24, contain a hydrophobic portion (residues 1 to 16) with a strong β prediction followed by a hydrophilic portion (residues 17 to 24) with a strong turn prediction.

Residues 28 to 172 show evidence of a heptad periodicity (2, 12). A discontinuity in the periodicity occurs around residue 150. Similar discontinuities have been previously detected in the sequences of other α -helical, coiled-coil proteins (e.g., 20, 21, 27, 31), and their implications for interruptions in the regularity of the coiled coils have been discussed (21, 27). The boundaries of this segment are not precisely defined by sequence considerations, but its maxi-

FIG. 5. (a) Globally averaged image of the sigma 1 fiber, showing the subdivision into morphological zones T(i) to T(iv) and H. Plotted in parallel on an expanded scale (arrows mapping panel a onto panel b) are the laterally integrated, stain-excluding density (see Results) (b); a profile showing the average curvature (calculated during the computational straightening procedure) as a function of position along the fiber (c). It extends only as far as the center of H, which was taken as point of reference in combining the traces. The dashed line represents the level of random experimental errors, i.e., a baseline for significance. (d) A theoretical flexibility trace calculated for the tail region of the fiber on the basis of our proposed model for how the $\sigma 1$ chains pass through it (Results), using the algorithm of Karplus and Schulz (18). The arrowheads in panels c and d mark the principal maxima of the respective curves.

mum length is probably 150 residues (25 to 174). We have designated this segment S(ii), subdivided into S(ia) and S(ib) by the break at residues 149 and 150.

An unusual feature of segment S(ii) is its relatively weak α -helical predictions (Fig. 6), apart from an initially strong prediction for its N-terminal part (residues 25 to 38), which is also very hydrophilic. The weak α -helix prediction over the next 64 residues of S(ia) (residues 39 to 102) is a consequence of the unusually high contents of glycyl and seryl residues. The destabilizing effect of glycyl residues is attributable to the additional conformational space available due to the small side chain (9, 11, 24). The α -helix potential of

residues 117 to 174 is weakened by their high content of threonyl and seryl residues. Much of the S(ii) coiled-coil segment, therefore, may be only marginally stable.

The remainder of the sequence (residues 175 to 462) is predicted to consist mainly of β -sheet alternating with random coil. The portions of the sequence predicted to have a β -sheet conformation are generally also hydrophobic (Fig. 6). The first part of this chain section (residues 175 to 247), which we have called S(iii), contains no proline and is successively rich in glycyl residues (177 to 194), aromatic residues (197 to 232), and hydrogen-bonding residues (234 to 245). Aromatic residues are believed to confer rigidity [20, 21], and this suggests that much of the S(iii) segment has a rigid β -sheet structure. The remainder of the sequence (residues 248 to 462) we have designated S(iv). It contains long runs of chain predicted to have a high content of β -sheets; these are predominantly hydrophobic, and thus are likely to be sequestered from the solvent. Moreover, a relatively high incidence of hydrophobic residues is indicative of a sequence that is more likely to assume a globular folding than an extended filamentous structure.

Structure of the fiber: correlation of ultrastructural and

TABLE 1. Tentative attribution of $\sigma 1$ sequence segments to morphological features of the reovirus fiber (Fig. 5)

Segment	Residue	No.	Probable location
S(i)	1-24	24	T(i)
S(ia)	25-149	125	T(ii)
S(ib)	150-174	25	T(ii)
S(iii)	175-247	73	T(iii)
S(iv)	248-462	215	T(iv), H

FIG. 6. Predicted secondary structure and hydropathy profile for the sigma 1 protein of reovirus type 2 (Jones), calculated from the sequence determined by Nibert et al. (25). The secondary structures were calculated by the Chou-Fasman algorithm (5). The hydropathy profile is according to Kyte and Doolittle (19), with hydrophobicity shown as positive and hydrophilicity as negative. Also marked are the heptad-rich region with its interruption (arrow) and the division of the sequence into the zones defined on the basis of their sequence properties (see Results).

sequence-based information. (i) **N termini are located at the proximal end of the fiber.** In a previous study, it was deduced that the N termini are located at or near the proximal end of the fiber (12) on the premise that the heptad-rich N-terminal sequence could form coiled coil(s) that make up the tail, although the heptad-containing sequence seemed too short to account for its full length. The present results clarify this correlation. Although the full length of the tail is substantially greater than was previously thought (12), the processed micrographs (Fig. 3) reveal that it comprises two distinct pieces, and the length of the proximal portion T(ii), at ~22 nm, correlates very closely with the expected length of the coiled coils, viz., 150 residues (25 to 174) at 0.1485 nm axial rise per residue (6), giving 22 nm.

A second line of evidence also supports this conclusion. The cluster of hydrophobic residues at the amino terminus readily accounts for the dandelion clock aggregates (Fig. 2a). In other words, these aggregates are stabilized by this hydrophobic interaction and so are essentially micellar in character. Similarly, we consider it likely that most if not all of the two-headed particles (Fig. 2b) represent fibers that have bound some other molecule(s) to this "sticky" patch at their proximal tips. The alternative interpretation, that they represent a side-by-side association of two oppositely oriented fibers, is implausible, both because the tails of these particles are often asymmetric (i.e., type B), which this explanation does not allow, and because the increase that would be expected in the width of their tails relative to a head is not observed.

(ii) **Purified $\sigma 1$ is a tetramer.** Earlier studies have concluded that $\sigma 1$ is most likely either a dimer (40) or a tetramer (3). Our ultrastructural data support the tetramer hypothesis on the following grounds. The two views of the T(ii) segment (Fig. 2c and d) correspond in terms of both length and lateral dimensions, with a pair of two-stranded coiled coils in a side-by-side configuration as viewed in the eclipse position (type B) and in the en face position (corresponding to a 90°

rotation around its axis) in which they are resolved from each other (type A). If, on the other hand, type A were to be a dimer of B-type molecules associated in parallel and in register, its head region should be appreciably bulkier than that of the A-type particle, which is not the case (cf. Fig. 3c and d).

The respective widths of the two views of the distal tail are 4.6 nm (type A) and 3.1 nm (type B). When compared with the corresponding measurement (2.0 nm) for the proximal half fiber of bacteriophage T7, which is known to be a single three-stranded coiled coil (31), these data strongly indicate that the $\sigma 1$ particle is too wide for a single two-stranded coiled coil. Thus these images imply that two two-stranded ropes are associated side by side in T(ii), with little if any supercoiling around their common axis. We conclude that the two morphological types (A and B), represent side views of the same particle, differing by ~90° in azimuth, and that the remainder are side views from angles that are significantly different from either of these positions. The seemingly anomalous results that B-type view of $\sigma 1$ appears wider than the T7 tail fiber may be attributed to (i) lateral smearing caused by some angular dispersion about this position among the various B-type molecules averaged and (ii) the fact that the $\sigma 1$ molecules exhibit greater curvature, which may not be entirely corrected on the scale of a fraction of a nanometer.

The center-to-center spacing between the two parallel rods resolved laterally in the T(ii) segment of the type A fiber (a spacing that is less subject to uncertainty arising from limited resolution than are the width measurements) is 2.2 nm. This spacing is entirely compatible with the separation to be expected between two parallel coiled coils and is not reconcilable with the spacing between the two strands of a single coiled coil (1.0 nm [11]), which, in any case, would not be resolved under these conditions.

(iii) **Demarcation of other morphological domains.** The correspondence between the morphological zones T(i) to

T(iv) and H (Fig. 5a) and the sequence segments S(i) to S(iv) was investigated by plotting the laterally integrated optical density across the image in Fig. 3e against distance along the center line (Fig. 5b) and then allocating the residues in proportion to the cumulative area under the curve. These results cannot be taken too literally, since they may be biased by local concentrations of positive staining or the possibility that disordered outcrops may not be visualized. With these qualifications, however, they do allow a useful approximation to be made of how the various sequence segments are distributed. The most important findings were that the number of residues allocated to T(ii) was close to the number in S(ii), the number allocated to T(iii) was close to the number in S(iii), the number allocated to T(iv) + H was close to the number in S(iv), and approximately one-third of the mass resides in the head.

(iv) Correlation of curvature with flexibility. We have also calculated a flexibility profile for the fiber (Fig. 5d). This curve describes the local variations in flexibility that are predicted by our hypothesis for how the $\sigma 1$ chains pass through the fiber. Accordingly, comparison of this curve with the empirical curvature profile (Fig. 5c) constitutes a quantitative test of that model. Using the flexibility indices for each amino acid assigned by Karplus and Schulz (18) on the basis of crystallographic temperature factors, the profile was first calculated from the primary sequence by using uniform sampling, i.e., with equal steps in the horizontal dimension for each amino acid. It was then resampled to allow for the nonuniform distribution of mass along the fiber, i.e., to maintain consistency with the longitudinal distribution of mass calculated from the filtered image (Fig. 5b).

In view of the very different ways in which the curvature and flexibility profiles are calculated, we do not expect the two curves to be superimposable. Rather, we look for a reasonable measure of correspondence between the respective locations of their major peaks, and in this regard, the match between the two curves is remarkably good (cf. Fig. 5c and d). The highest peak in the flexibility profile, Asn-Ser-Thr-Gly-Ser-Thr (residues 269 to 274), coincides with the predominant kink site, although the breadth of the corresponding peak in the curvature plot suggests that there may also be other flexible peptides nearby. The second highest flexibility peak is very close to the amino-terminal (proximal) tip. This location is so close to the end of the particle that instances of intramolecular kinking there would be hard to recognize as such, since the pieces proximal to such kinks would be so small. However, the existence of a flexible hinge in this Gly-rich segment (residues 17 to 21 and Gly-Asn-Gly-Glu-Ser) is consistent with the observed variability in appearance of two-ended fibers (Fig. 2b) which we have interpreted as having other molecules bound nonspecifically to their sticky ends (see i above). A peptide with notably similar properties, Gly-Asn-Gly-Leu-Asn-Lys-Thr-Gly-Asn-Ser (residues 177 to 186), is also found immediately after the heptads of the rod domain and accounts for the kink site between T(ii) and T(iii).

Coinciding maxima are also found at a point about a quarter of the way along the rod domain (although we did not notice a significant amount of kinking at this site). Finally, a further respect in which the two curves match is that, between this site and the end of the rod domain, no significant maxima are observed in either case, indicating that this part of the fiber is relatively rigid. We conclude, therefore, that the substantial measure of agreement observed between these two curves supports our proposal for how the $\sigma 1$ chains pass through the fiber.

A high incidence of Gly is frequently associated with sites of enhanced mobility or flexibility (16–18). It is, moreover, noteworthy that, in addition to Gly, the three kink-prone sites are predominantly occupied by Asn, Ser, and Thr residues, although these are by no means the only residues deemed flexible according to the temperature-factor criteria (18).

Interactions between coiled coils in the T(ii) tail segment. We have studied the distribution of the surface residues in a two-stranded rope formed from parallel $\sigma 1$ chains in register in order to obtain information about the way in which two such ropes may be bonded together. This distribution can conveniently be studied by the use of radial projections (Fig. 7) in which the axis for the projection is taken as the axis of the rope. Two types of projection were used. In the first, only the most superficial residues were included (residues occupying the *b*, *c*, and *f* positions in the heptad); in the second, the partially buried residues in *e* and *g* were also included.

A striking feature of the first type of projection is the occurrence of uncharged surface patches rich in hydrogen-bonding groups (28) and containing occasional hydrophobic residues. A systematic appraisal of possible interactions between two two-strand coiled-coil ropes of S(ii) segments was carried out by marking off surface zones in the projections corresponding to $\pm 45^\circ$ from the line of contact. Following the general technique used by Crick (6), the projection of the second rope was inverted and superposed on the projection of the first rope and a search was made for favorable interactions in the range of $\pm 45^\circ$ from the line of contact. In this search, it was assumed that the two ropes are equivalent, implying that they are related by a dyad along the center line of the particle. In effect, this means that the azimuth of the line of contact must be the same for both ropes, and so the only parameters that may be varied are the azimuth of the line joining the two rope axes and the pitch of the coiled coil. An additional refinement was the use of a projection onto the plane of contact (Fig. 8), which is a more stringent test of the proximity of two potentially interacting side chains.

In the configuration found to have the greatest potential for interactions between S(ii) segments (which have an uninterrupted run of 125 residues with a heptad periodicity), the azimuth was such that Ser-96 (an *f* position) lay on a line joining the axes of the two ropes (Fig. 9) and the pitch of the coiled coil was 14 nm (26). A list of the favorable interactions in this configuration is given in Table 2. Of the 18 interactions, 14 are hydrogen bonds, 2 are hydrophobic interactions, and 2 are ionic.

DISCUSSION

Oligomeric character of $\sigma 1$. Previous studies have indicated that the $\sigma 1$ fiber is most likely a dimer (40) or a tetramer (3). Assuming that the protein is equally represented at each of the 12 vertices, data on the protein stoichiometry of the virion from quantitative gel electrophoresis (29) suggest that it is a dimer on the basis of Coomassie blue staining, whereas the tetramer is favored on the basis of labeling with ^3H or ^{14}C . In chemical cross-linking experiments with purified $\sigma 1$ (in which rather small fractions of total protein were converted into covalently linked oligomers), the highest oligomers detected were tetramers (3, 40). These data were interpreted as implying that the fiber is a dimer (40) (taking the observed tetramer to be an intermolecular cross-link) or, alternatively, a tetramer (3). The

FIG. 7. (a) Exscribed radial projection of the surface residues in heptad positions *b*, *c*, and *f* for the two-strain coiled-coil rope segment S(iiia) comprising residues 25 to 149. Note how uncharged residues occur in patches separated vertically by approximately $P/2$, where P is the pitch of the coiled coil (assumed to be 14 nm in this case). (b) Method of constructing such a radial projection for a two-stranded coiled-coil rope. The position of each residue is calculated and projected onto an exscribed cylinder. When cut along the 0 to 360° line and opened out, a projection of the type shown in panel a is obtained.

FIG. 8. (a) Contact surface projection of a two-stranded coiled coil of the sigma 1 segment S(ia) (residues 25 to 149). The azimuthal origin is such that Ser-96 projects onto the line of contact (see Fig. 9). Such projections are useful in studying the interactions between coiled-coil ropes. (b) To construct the projection, the coordinates of each residue are calculated and projected on to the plane of contact. Residues outside the range $\pm p_0$ from the plane are omitted.

ultrastructural evidence presented here firmly favors the tetramer hypothesis.

Further support for the conclusion that the fiber is a tetramer comes from measurements made from scanning transmission electron micrographs of the heads of fibers projecting from reovirus infectious subviral particles (D. Furlong and J. Wall, unpublished results). The average head mass is 74.5 kDa (± 18 kDa standard deviation; ± 2.5 kDa standard error of the mean; $n = 78$). According to the

FIG. 9. The azimuthal arrangement deduced for the two two-strand ropes formed from the S(ia) segment of the sigma 1 protein (residues 25 to 149). The z values and the azimuth values are defined in the legend to Fig. 7.

stain-exclusion profile of the fiber (Fig. 5b), the head accounts for one-third of the mass, a conclusion that is in good agreement with the head-tail demarcation point deduced from a detailed analysis of the primary sequence (25). On this basis, the mass predicted from the model is 68 kDa for a tetramer and 34 kDa for a dimer. Accordingly, the scanning transmission electron microscopic mass data indicate that the fiber is a tetramer in situ as well as in the purified state.

Secondary structure of $\sigma 1$. According to our predictive analysis, $\sigma 1$ is expected to have an α -helix content of 40% and a β -sheet content of 36% if we assume that the heptads of segment S(ii), with their propensity to form coiled coils, override the tentative α -helix prediction for this part of the molecule but accept the predictions of the algorithms (5) elsewhere. Very similar numbers, 44 and 40%, respectively, are obtained with another method (13). Circular dichroism measurements have been recorded from $\sigma 1$ purified from urea-disrupted type 3 (Dearing) reovirus from which an α -helix content of 52% and a β -sheet content of 19% were calculated (40). Thus, there is good agreement as to α -helix content but discrepancy with regard to β -sheet. The basis for this discrepancy is not apparent. On one hand, current secondary structure prediction algorithms are known to have limited accuracy. On the other hand, it is highly pertinent

TABLE 2. Interactions capable of stabilizing the association of two two-strand coiled-coil ropes formed from the sigma 1 protein segment S(ia)

Interaction ^a	Type
Ser-39-Ser-39'	H bond
Asn-43-Asn-43'	H bond
Ser-51-Ser-51'	H bond
Gln-55-Gln-55'	H bond
Arg-62-Arg-62'	H bond
Ser-85-Ser-85'	H bond
Ser-92-Ser-92'	H bond
Asp-103-Arg-104'	Ionic
Arg-104-Asp-103'	Ionic
Val-107-Val-107'	Hydrophobic
Arg-111-Arg-111'	H bond
Leu-123-Leu-123'	Hydrophobic
Ser-134-Ser-134'	H bond
Thr-137-Asn-138'	H bond
Asn-138-Thr-137'	H bond
Ser-142-Thr-145'	H bond
Thr-145-Ser-142'	H bond
Thr-149-Thr-149'	H bond

^a Prime denotes residues in the second interacting chain.

that in the comparable nodules of two other fibrous viral receptor-recognition proteins which were predicted to contain β -sheet, adenovirus fiber and T4 distal tail fiber, the presence of this conformation was confirmed by electron diffraction from paracrystals (8, 15). Hopefully, it will prove possible to produce $\sigma 1$ in sufficient quantities to make similar experiments feasible and thus clarify the current anomaly as to the β -sheet content of the fiber. However, the consideration that the above comparison is between $\sigma 1$ proteins from different strains of reovirus, subjected to different purification procedures, may limit the extent to which a real discrepancy exists, and we shall accept the predicted β -sheet content for the purposes of the following discussion.

Structural motifs in $\sigma 1$. It is noteworthy that the deduced structure of the $\sigma 1$ fiber embodies the same motifs already detected in the molecular counterparts of other viruses, although with some novel variations on familiar themes. Previously, α -helical coiled-coil domains have been detected in the influenza virus hemagglutinin (39), the peplomer protein of coronavirus (7), and the tail fibers of bacteriophages T7 (31) and T3 (32); filamentous nodular domains with high contents of strands of β -sheet occur in the fibers of adenovirus (15) and phages T4 (8) and T7 (31). In terms of overall molecular organization, $\sigma 1$ is strikingly reminiscent of the T7 fiber. In both cases, the N terminus secures the fiber to the virion, followed by an α -helical coiled-coil segment and then the C-terminal part of the molecule (in which the receptor-binding sites are thought to reside [23, 31]), which takes the form of a string of nodules. However, unlike the T7 fiber (and most other such proteins characterized to date, which are mainly homotrimers), $\sigma 1$ appears to be a homotetramer.

Extended and retracted conformations. Furlong et al. (12) have presented evidence that $\sigma 1$ can exist in two structural states. In the first, the fibers are flush with the outer capsid surface; in the second, they extend from the vertices. In this state, the fibers closely resemble the purified protein, whose full length we have determined to be 48 to 50 nm. Such a long rod seems unlikely to extend all the way into the viral interior, in which case $\sigma 1$ should assume a more compact, folded conformation in the retracted state (12). Transition between the two states might be part of the infection process (12), although it might otherwise represent a nonphysiological event involving denaturation of a labile part (or parts) of the molecule, which would be induced by the two treatments which have been shown to detach $\sigma 1$ from the virion, viz., heating (3) or treatment with urea (40). However, we note that any such denaturation is likely to be localized, since the purified protein is competent to bind to receptors (23, 40). In either case, the present results identify portions of the molecule that are candidates to be involved in this transition, i.e., the T(ii) coiled-coil region (whose α -helices are suspected to be only marginally stable) and the points of enhanced flexibility in the tail (Fig. 5c), particularly the observed kink sites (Fig. 2d and e).

Intercoiled-coil interactions. The evidence summarized above indicates that the $\sigma 1$ fiber is a double dimer and thus observed a different symmetry than the other viral surface protein known to be a homotetramer, viz., the influenza virus neuraminidase which is a symmetric (p4) molecule (37). The $\sigma 1$ dimers appear to be bonded together to register throughout the tail region. In the coiled-coil region, T(ii), our analysis suggests that the interactions between the two ropes may be dominated by hydrogen-bonding interactions between patches of uncharged residues. There are some interesting parallels with the intermediate filament (IF) proteins

(10, 27). These also form two-stranded coiled coils that, in turn, pair to form four-chain complexes in which bonding between similar patches may be implicated (R. D. B. Fraser, unpublished results). However, there is the interesting distinction that in IF proteins, the paired coiled coils are thought to be antiparallel (14). In this respect, one might conjecture that the polarity of the $\sigma 1$ fiber may be imposed by interdimer interactions involving the T(iii) and T(iv) domains which have no counterparts in IF proteins.

ACKNOWLEDGMENTS

This work was performed while one of us (R.D.B.F.) was a Fogarty Scholar-in-Residence at the National Institutes of Health. We gratefully acknowledge T. Dermody, who made available his sequence information prior to publication, and D. A. D. Parry (Massey University, Palmerston North, New Zealand) for helpful suggestions concerning sequence analysis. We thank M. E. Bisher for help with photography.

This research was supported in part by Public Health Service grants IR37A113178 and NS16998 from the National Institutes of Health to B.N.F., D.B.F., and M.L.N.

LITERATURE CITED

1. Banerjee, A. C., K. A. Brechling, C. A. Ray, H. Erikson, D. J. Pickup, and W. K. Joklik. 1988. High-level synthesis of biologically active reovirus protein $\sigma 1$ in a mammalian expression vector system. *Virology* 167:601-612.
2. Bassel-Duby, R., A. Jayasuriya, D. Chatterjee, N. Sonenberg, J. V. Maizel, and B. N. Fields. 1985. Sequence of reovirus haemagglutinin predicts a coiled-coil structure. *Nature (London)* 315:421-423.
3. Bassel-Duby, R., M. L. Nibert, C. J. Homcy, B. N. Fields, and D. G. Sawutz. 1987. Evidence that the sigma 1 protein of reovirus serotype 3 is a multimer. *J. Virol.* 61:1834-1841.
4. Cashdollar, L. W., R. A. Chmelo, J. R. Wiener, and W. K. Joklik. 1985. Sequences of the S1 genes of the three serotypes of reovirus. *Proc. Natl. Acad. Sci. USA* 82:24-28.
5. Chou, P. Y., and G. D. Fasman. 1974. Conformational parameters for amino acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry* 13:211-222.
6. Crick, F. H. C. 1953. The Fourier transform of a coiled-coil. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* 6:685-688.
7. De Groot, R. J., W. Luytjes, M. C. Horzinek, B. A. M. van der Zeijst, W. J. M. Spaan, and J. A. Lenstra. 1987. Evidence for coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* 196:963-966.
8. Earnshaw, W. C., E. B. Goldberg, and R. A. Crowther. 1979. The distal half of the tail fiber of bacteriophage T4. *J. Mol. Biol.* 132:101-131.
9. Fraser, R. D. B., B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki. 1967. Effect of glycyl residues on the stability of the α -helix. *Biopolymers* 5:251-257.
10. Fraser, R. D. B., B. S. Harrap, T. P. MacRae, F. H. C. Stewart, and E. Suzuki. 1978. Structure of the α -keratin microfibril. *J. Mol. Biol.* 108:435-452.
11. Fraser, R. D. B., and T. P. MacRae. 1963. Conformation of proteins and related synthetic polypeptides. Academic Press, Inc., New York.
12. Furlong, D. B., M. L. Nibert, and B. N. Fields. 1988. Sigma 1 protein of mammalian reoviruses extends from the surfaces of viral particles. *J. Virol.* 62:246-256.
13. Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for providing the secondary structure of globular proteins. *J. Mol. Biol.* 120:97-120.
14. Geisler, N., E. Kaufmann, and K. Weber. 1985. Antiparallel orientation of the two double-stranded coiled-coils in the tetrameric protofilament unit of intermediate filaments. *J. Mol. Biol.* 182:173-177.
15. Green, N. M., N. G. Wrigley, W. C. Russell, S. R. Martin, and

- A. D. McLachlan. 1983. Evidence for a repeating cross- β sheet structure in the adenovirus fiber. *EMBO J.* **2**:1357–1365.
16. Huber, R. 1979. Conformational flexibility in some protein molecules and its functional significance. *Trends Biochem. Sci.* **4**:271–276.
17. Huber, R., and W. S. Bennet. 1983. Functional significance of flexibility in proteins. *Biopolymers* **22**:261–279.
18. Karplus, P. A., and G. E. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* **72**:212–213.
19. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**:105–132.
20. Manjula, B. N., B. L. Trus, and V. A. Fischetti. 1985. Presence of two distinct regions in the coiled-coil structure of the streptococcal Pep M5 protein: relationship to mammalian coiled-coil proteins and implications to its biological properties. *Proc. Natl. Acad. Sci. USA* **82**:1064–1068.
21. McLachlan, A. D., and J. Karn. 1983. Periodic features in the amino-acid sequence of nematode myosin rod. *J. Mol. Biol.* **164**:605–626.
22. Nagata, L., S. A. Masri, D. C. W. Mah, and P. W. K. Lee. 1984. Molecular cloning and sequencing of the reovirus (serotype 3) S1 gene which encodes for the viral cell attachment protein $\sigma 1$. *Nucleic Acids Res.* **12**:8699–8710.
23. Nagata, L., S. A. Masri, R. T. Pon, and P. W. K. Lee. 1987. Analysis of functional domains on reovirus cell attachment protein $\sigma 1$ using cloned S1 gene deletion mutants. *Virology* **160**:162–168.
24. Nemethy, G., S. J. Leach, and H. A. Scheraga. 1966. The influence of amino-acid side-chains on the free energy of helix-coil transitions. *J. Phys. Chem.* **70**:998–1004.
25. Nibert, M. L., T. S. Dermody, and B. N. Fields. 1990. Structure of the reovirus cell-attachment protein: a model for the domain organization of $\sigma 1$. *J. Virol.* **64**:2976–2989.
26. Parry, D. A. D. 1975. Double helix of tropomyosin. *Nature (London)* **256**:346–347.
27. Parry, D. A. D., and R. D. B. Fraser. 1985. Intermediate filament structure. 1. Analysis of IF protein sequence data. *Int. J. Biol. Macromol.* **7**:203–213.
28. Scheraga, H. A. G., G. Nemethy, and I. Z. Steinberg. 1962. The contribution of hydrogen bonds to the thermal stability of protein conformations. *J. Biol. Chem.* **237**:2506–2508.
29. Smith, R. E., H. J. Zweerink, and W. K. Joklik. 1969. Polypeptide components of virions, top components, and cores of reovirus type 3. *Virology* **39**:791–810.
30. Steven, A. C., R. Stall, P. M. Steinert, and B. L. Trus. 1986. Computational straightening of images of curved macromolecular helices by cubic spline interpolation facilitates structural analysis by Fourier methods, p. 31–33. *In* J. Metuzals (ed.), *Electron microscopy and Alzheimer's disease*. San Francisco Press, San Francisco.
31. Steven, A. C., B. L. Trus, J. V. Maizel, M. Unser, D. A. D. Parry, J. S. Wall, J. F. Hainfeld, and F. W. Studier. 1988. Molecular substructure of a viral receptor-recognition protein: the gp17 tail-fiber of bacteriophage T7. *J. Mol. Biol.* **200**:351–365.
32. Takahashi, S., and T. Ooi. 1988. Three- α -helical coiled coil as a proposed model of a thin rod segment of bacteriophage T3 tail fibers. *Biochem. Biophys. Res. Comm.* **150**:1244–1250.
33. Trus, B. L., and A. C. Steven. 1981. Digital image processing of electron micrographs—the PIC system. *Ultramicroscopy* **6**:383–386.
34. Unser, M., B. L. Trus, and A. C. Steven. 1989. Normalization procedures and factorial representations for classification of correlation-aligned images: a comparative study. *Ultramicroscopy* **30**:299–310.
35. Valentine, R. C., and N. M. Green. 1967. Electron microscopy of an antibody-hapten complex. *J. Mol. Biol.* **27**:615–617.
36. Van Heel, M., and J. Frank. 1981. Use of multivariate statistics in analyzing the images of biological macromolecules. *Ultramicroscopy* **6**:187–194.
37. Varghese, J. N., W. G. Laver, and P. M. Coleman. 1983. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature (London)* **303**:35–40.
38. Willey, R. L., D. H. Smith, L. A. Lasky, T. S. Theodore, P. L. Earl, B. Moss, D. J. Capon, and M. A. Martin. 1988. In vitro mutagenesis identifies a region within the envelope gene of the human immunodeficiency virus that is critical for infectivity. *J. Virol.* **62**:139–147.
39. Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature (London)* **289**:366–373.
40. Yeung, M. S., M. J. Gill, S. S. Alibhal, M. S. Shahrabadi, and P. W. K. Lee. 1987. Purification and characterization of the reovirus cell attachment protein $\sigma 1$. *Virology* **156**:377–385.