Structure of the Reovirus Cell-Attachment Protein: a Model for the Domain Organization of $\sigma 1$ [†]

MAX L. NIBERT,¹ TERENCE S. DERMODY,^{1,2} AND BERNARD N. FIELDS^{1,2,3*}

Department of Microbiology and Molecular Genetics, Harvard Medical School,¹ Department of Medicine, Brigham and Women's Hospital,² and Shipley Institute of Medicine,³ Boston, Massachusetts 02115

Received 11 July 1989/Accepted 4 December 1989

This report describes a model for the structure of the reovirus cell-attachment protein $\sigma 1$. S1 gene nucleotide sequences were determined for prototype strains of the three serotypes of mammalian reoviruses. Deduced amino acid sequences of the S1-encoded $\sigma 1$ proteins were then compared in order to identify conserved features of these sequences. Discrete regions in the amino-terminal two-thirds of $\sigma 1$ sequence share characteristics with the fibrous domains of other cellular and viral proteins. Most of the amino-terminal one-third of $\sigma 1$ sequence is predicted to form an α -helical coiled coil like that of myosin. The middle one-third of $\sigma 1$ sequence appears more heterogeneous; it is predicted to form a large region of β -sheet that is followed by a region which contains two short α -helical coiled coils separated by a smaller region of β -sheet. The two β -sheet regions are each proposed to form a cross- β sandwich like that suggested for the rod domain of the adenovirus fiber protein (N. M. Green, N. G. Wrigley, W. C. Russell, S. R. Martin, and A. D. McLachlan, *EMBO J.* 2:1357-1365, 1983). The remaining carboxy-terminal one-third of $\sigma 1$ sequence is predicted to form a structurally complex globular domain. A model is suggested in which the discrete regions of $\sigma 1$ sequence are ascribed to morphologic regions seen in computer-processed electron micrographic images of the protein (R. D. B. Fraser, D. B. Furlong, B. L. Trus, M. L. Nibert, B. N. Fields, and A. C. Steven, *J. Virol.* 64:2990-3000, 1990).

The $\sigma 1$ protein is the reovirus cell-attachment protein (20, 46) and is responsible for tissue- and cell-specific tropism (31, 39, 43, 47, 51). It is also the viral hemagglutinin (49). A serotype-specific humoral immune response is directed against the $\sigma 1$ protein (48), and three distinct serotypes of reovirus have been identified (36). The three serotypes are represented by prototype reovirus strains type 1 Lang (T1L), type 2 Jones (T2J), and type 3 Dearing (T3D) (33). The $\sigma 1$ protein (~50 kilodaltons [kDa]) is encoded by the reovirus S1 double-stranded RNA (dsRNA) gene segment, which also encodes the smaller $\sigma 1s$ protein (~14 kDa) in an overlapping reading frame (4).

Bassel-Duby et al. (4) predicted that the $\sigma 1$ protein has the morphology of a long fibrous tail topped with a globular head. This prediction was based on the observation that a large region in the amino-terminal half of the T3D $\sigma 1$ sequence contains a repeating pattern of apolar residues that might be expected to form a long α -helical coiled coil. These authors suggested that the $\sigma 1$ protein is attached to the reovirus particle by extension of its fibrous tail into the lumen of the $\lambda 2$ core spike. According to this model, the $\sigma 1$ head is found at the surfaces of virions where it can effect its function of binding to cell-surface receptors.

Furlong et al. (17) confirmed the tail-and-head morphology of $\sigma 1$ with electron microscopic studies of reovirus particles and isolated T2J $\sigma 1$ protein. These studies indicated that, in many cases, most of the $\sigma 1$ tail is not buried in the $\lambda 2$ spike but extends outward from viral particles so that the $\sigma 1$ head is projected 40 nm or more away from the particle surface. In these cases, reovirus has an appearance similar to that of adenovirus, with an extended fiber protein (44). These studies also indicated that the $\sigma 1$ tail is longer than predicted by the coiled-coil sequence region and suggested that addi-

tional sequences in $\sigma 1$ contribute to the formation of the fibrous tail.

This report describes our efforts to identify additional details of $\sigma 1$ protein structure. The S1 gene nucleotide sequences of reovirus strains T1L, T2J, and T3D were determined. The deduced $\sigma 1$ amino acid sequences were then compared in order to identify conserved features of these sequences. In conjunction with the work of Fraser et al. (16), in which the morphology of the $\sigma 1$ protein was defined in computer-processed electron micrographs, we were able to develop a model in which discrete regions of $\sigma 1$ sequence are ascribed to morphologic regions in the protein. These studies represent progress toward understanding the domain organization of the $\sigma 1$ protein. Knowledge of $\sigma 1$ structure should contribute to an understanding of its function as the cell-attachment protein, hemagglutinin, and serotype-specific protein of the mammalian reoviruses.

MATERIALS AND METHODS

Cells and virus. Spinner-adapted mouse L cells were grown in either suspension or monolayer cultures in Joklik modified Eagle minimal essential medium (Irving Scientific, Santa Ana, Calif.) that was supplemented to contain 5% fetal calf serum (Hyclone Laboratories, Logan, Utah), 2 mM glutamine, 1 U of penicillin per ml, and 1 μ g of streptomycin (Irving Scientific) per ml. Second- and third-passage L-cell lysate stocks of twice-plaque-purified T1L, T2J, and T3D reovirus were used to make purified virion preparations (17).

Purification of dsRNA. Purified virions (2 to 3 mg of viral protein as determined by optical density at 260 nm) were incubated in 1% sodium dodecyl sulfate (SDS)-50 mM Tris-1 mM EDTA (pH 8.0) in a total volume of 3.0 ml at 37°C for 15 min and then were extracted twice with phenol-chloroform (1:1) after the addition of 60 μ l of 5 M NaCl. The RNA was precipitated in 3 volumes of cold 100% ethanol after the addition of 1/10 volume 3 M ammonium acetate (pH 5.2). Viral RNA was subjected to electrophoresis in SDS-

* Corresponding author.

[†] This paper is dedicated to Wolfgang K. Joklik for his contributions to virology as a scientist and a mentor.

10% polyacrylamide gels for 400 mA h, gels were stained with ethidium bromide, and individual gene segments were excised. The RNA was electroeluted from gel slices for 200 mA h, precipitated twice in cold 100% ethanol, and suspended in 90% dimethyl sulfoxide (Aldrich Chemical, Milwaukee, Wis.)—1 mM Tris (pH 6.8) in a total volume of 10 μ L.

Sequencing dsRNA with primers. Oligonucleotide primers were made according to the published S1 nucleotide sequences for T1L (7), T2J (7), and T3D (4). Oligonucleotides for the T1L and T2J S1 genes were purified by C-18 Sep-Pak column chromatography (Waters Associates, Milford, Mass.), and those for the T3D S1 gene were purified by ion-exchange high-pressure liquid chromatography (Applied Biosystems, User Bull. 13:1–28, 1984). The dsRNA was denatured by incubation at 50°C for 45 min, and dideoxy sequencing reactions were performed, using oligonucleotide primers, reverse transcriptase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.), and [35 S]dATP (1,000 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) (6). Sequencing reactions were subjected to 8% polyacrylamide gel electrophoresis and autoradiography. Primers corresponded to sequences that occurred at approximately 200-nucleotide intervals on each strand of the dsRNA S1 genes. Regions of overlapping sequence were determined from adjacent primers so that the full-length sequences of each strand of the T1L, T2J, and T3D S1 genes were obtained.

Alignments of deduced amino acid sequences. Pairwise sequence alignments were performed with the alignment algorithm of Needleman and Wunsch (30) and the amino acid similarity scoring matrix of either Dayhoff et al. (10) or Risler et al. (34), as applied in the global alignment program GAP from the University of Wisconsin Genetics Computer Group (UWGCG) (11). Alignments scores were end weighted, and gap- and length-weight scores of between -1.00 and -0.10 , respectively, and -5.00 and -0.50 , respectively, were used. The final three-sequence alignment was derived by manually comparing small regions of sequence and then aligning these sequences to maximize first the number of conserved apolar residues (A, V, I, L, M, F, Y, W) and second the number of conserved residues of other characteristics.

Identity plots. Each possible comparison of the three $\sigma 1$ sequences (three two-sequence comparisons and the three-sequence comparison) was scored for alignment positions at which amino acid residues were absolutely conserved (match score = 1, nonmatch score = 0). Match scores were then summed over all consecutive 35-position (i to $i+34$) windows, and the sums were assigned to the middle ($i+17$) position in each window. Two cycles of additional smoothing were obtained by averaging the values for each three consecutive match sums (i , $i+1$, $i+2$) and assigning the average value to the position of the middle ($i+1$) sum. In the case of the three-sequence comparison, residues in all three sequences were required to be conserved in order for an alignment position to receive a match score.

Dot matrix analysis. Dot matrix analyses were performed by the technique of Maizel and Lenk (24) as implemented in the UWGCG programs COMPARE and DOTPLOT. The window-stringency comparison capability of COMPARE was used in all cases. The amino acid similarity matrix of Risler et al. (34) was found to yield significantly lower background in our analyses than that of Dayhoff et al. (10). A window length of 21 and a stringency of 31 were chosen to generate plots for presentation.

Secondary structure and hydrophobicity plots. Secondary structure predictions for $\sigma 1$ were performed by the technique of Argos and MohanaRao (1). Propensity values for

α -helix, β -strand, and β -turn (8) were assigned to each amino acid residue in the T1L, T2J, and T3D $\sigma 1$ proteins. The values for each type of secondary structure were smoothed along the length of an individual $\sigma 1$ protein by averaging the values for each three consecutive amino acid residues (i , $i+1$, $i+2$) and assigning the average value to the middle ($i+1$) residue in the triplet. The once-smoothed values were then subjected to a second cycle of smoothing by the same technique in order to generate twice-smoothed values. Because of the smoothing process, the amino-terminal and carboxy-terminal two residues of each $\sigma 1$ sequence were not assigned values by this technique. Consensus scores for α -helix, β -strand, and β -turn propensities at each position in the three-way alignment of $\sigma 1$ sequences were generated by first inserting gaps in each protein-length string of twice-smoothed propensity values and then averaging the values from the three sequences at each alignment position.

Hydrophobicity values (35) were also assigned to each amino acid residue in the individual T1L, T2J, and T3D $\sigma 1$ proteins. These values were twice smoothed over three consecutive positions by the same technique used to generate consensus secondary structure predictions. A consensus hydrophobicity score at each amino acid position in the three-sequence alignment of $\sigma 1$ was also generated by the same technique used for consensus secondary structure predictions.

RESULTS

Nucleotide sequences of reovirus T1L and T2J S1 genes. We applied the rapid sequencing technique of Bassel-Duby et al. (6) to S1 genomic dsRNA purified from virions of reovirus strains T1L, T2J, and T3D. The full-length T1L and T2J S1 coding strand nucleotide sequences and their encoded $\sigma 1$ and $\sigma 1s$ amino acid sequences are shown in Fig. 1. The T3D S1 sequence reported by Bassel-Duby et al. (4) was confirmed in this study (data not shown). The three prototype S1 genes have coding strand sequences containing small 5' and 3' noncoding regions of 13 and 37 nucleotides, respectively, in T1L, 13 and 38 nucleotides in T2J, and 12 and 36 nucleotides in T3D. The lengths of the encoded $\sigma 1$ proteins and their deduced molecular masses are as follows: T1L, 470 amino acids and 51,368 Da; T2J, 462 amino acids and 50,457 Da; and T3D, 455 amino acids and 49,185 Da. These relative masses are consistent with the relative migrations of the three $\sigma 1$ proteins in Tris-glycine-buffered SDS-polyacrylamide gels (33; data not shown).

Analysis of deduced $\sigma 1$ amino acid sequences. The deduced amino acid sequences of the T1L, T2J, and T3D $\sigma 1$ proteins were studied in an attempt to predict features of $\sigma 1$ protein structure. These studies are described in the sections that follow and consistently suggest that the three $\sigma 1$ sequences are divided into discrete regions which are likely to reflect a domain organization conserved in the three $\sigma 1$ proteins.

(i) Alignment of $\sigma 1$ amino acid sequences by the insertion of gaps. An alignment of the amino acid sequences of the three prototype $\sigma 1$ proteins is shown in Fig. 2. The consensus length for $\sigma 1$ in the three-sequence alignment is 478 alignment positions. Gap positions were inserted in the $\sigma 1$ sequences to optimize the alignment as follows: 8 in T1L, 16 in T2J, and 23 in T3D. The amino-terminal portion of each sequence contains one or two large gaps that are 3 to 14 alignment positions long. The locations of these gaps in our alignment approximately correspond to those in the alignment of Cashdollar et al. (7). A unique feature of our alignment is the cluster of small gaps (one or two alignment

A.	met asp ala ser leu ile thr glu ile arg lys ile val leu gln leu ser val ser ser asn gly ser gln ser lys glu ile glu glu ile	31	
	GCUAUUCGCGCCU <u>AUG</u> GAU GCA UCU CUC AUU ACA GAG AUA CGG AAA AUA GUA CUC CAA CUA UCU GUA UCA AGC <u>AUU</u> GGC UCC CAG UCA AAA GAA AUC GAG GAA AUC	11	106
	met ala pro ser gln lys lys ser arg lys ser		
	lys lys gln val gln val asn val asp asp ile arg ala ala asn ile lys leu asp gly leu gly arg gln ile ala asp ile ser asn ser ile ser thr ile	66	
	AAG AAA CAA GUC CAG GUC AAC GUU GAU GAU AUC AGG GCU GCC AAU AUU AAA CUC GAC GGA CUU GGA AGA CAG AUU GCU GAC AUC AGC AAU AGC AUC UCA ACC AUU	11	211
	arg asn lys ser arg ser thr leu met ile ser gly leu pro ile leu asn ser thr asp leu glu asp arg leu leu thr ser ala ile ala ser gln pro leu	46	
	glu ser arg leu gly glu met asp asn arg leu val gly ile ser ser gln val thr gln leu ser asn ser val ser gln asn thr gln ser ile ser ser leu	101	
	GAG UCA AGA UUG GGU GAG AUG GAU AAU CGA CUU GUG GGU AUC UCG AGU CAG GUC ACG CAA UUA UCU AAC UCA GUU AGC CAG AAC ACU CAG AGC AUA UCC UCA UUG	81	316
	ser gln asp trp val arg trp ile ile asp leu trp val ser arg val arg ser arg asn tyr leu thr gln leu ala arg thr leu arg ala tyr pro his trp		
	gly asp arg ile asn ala val glu pro arg val asp ser leu asp thr val thr ser asn leu thr gly arg thr ser thr leu glu ala asp val gly ser leu	136	
	GGU GAC AGA AUC AAU GCU GUC GAA CCA CGA GUU GAC AGU CUG GAU ACG GUC ACG UCU AAU CUC ACU GGA CGA ACA UCC ACU UUG GAG GCA GAU GUU GGA AGC UUA	116	421
	val thr glu ser met leu ser asn his glu leu thr val trp ile arg ser arg leu ile ser leu asp glu his pro leu trp arg gln met leu glu ala tyr		
	arg thr glu leu ala ala leu thr thr arg val thr thr glu val thr arg leu asp gly leu ile asn ser gly gln asn ser ile gly glu leu ser thr arg	171	
	CGG ACA GAA <u>CUA</u> GCG CUA ACA ACA CGG GUG ACA ACU GAG GUU ACA AGG UUA GAU GGU CUA AUC AAU AGU GGC CAG AAU UCG AUU GGU GAG CUA UCC ACA AGA	119	526
	gly gln asn		
	leu ser asn val glu thr ser met val thr thr ala gly arg gly leu gln lys asn gly asn thr leu asn val ile val gly asn gly met trp phe asn ser	206	
	CUA UCC AAU GUG GAG ACG UCU AUG GUG ACG ACG GCU GGA CGG GGA CUG CAG AAA AAC GGA AAC ACC UUG AAC GUC AUU GUA GGU AAU GGA AUG UGG UUU AAU AGU		631
	ser asn gln leu gln leu asp leu ser gly gln ser lys gly val gly phe val gly thr gly met val val lys ile asp thr asn tyr phe ala tyr asn ser	241	
	UCU AAU CAA UUG CAG CUC GAC CUU UCG GGG CAA UCA AAA GGG GUG GGA UUU GUC GGC ACA GGA AUG GUG GUU AAG AUU GAU ACU AAU UAU UUU GCU UAC AAU AGU		736
	asn gly glu ile thr leu val ser gln ile asn glu leu pro ser arg val ser thr leu glu ser ala lys ile asp ser val leu pro pro leu thr val arg	276	
	AAU GGA GAG AUU ACA UUG GUG AGU CAA AUC AAU GAA UUG CCA UCG CGC GUA UCA ACA CUG GAA UCA GCG AAA AUC GAU UCA GUU UUA CCU CCA UUA ACC GUA CGC		841
	glu ala ser gly val arg thr leu ser phe gly tyr asp thr ser asp phe thr ile ile asn ser val leu ser leu arg ser arg leu thr leu pro thr tyr	311	
	GAA GCG AGC GGC GUA CGU ACC CUG AGC UUU GGU UAU GAU ACG AGC GAU UUU ACA AUC AUC AAC UCC GUA CUG UCG UUA CGG UCA CGU UUG ACU CUU CCG ACA UAC		946
	arg tyr pro leu glu leu asp thr ala asn asn arg val gln val ala asp arg phe gly met arg thr gly thr trp thr gly gln leu gln tyr gln his pro	346	
	AGG UAC CCU CUG GAG CUC GAC ACA GCA AAU AAU AGA GUG CAG GUG GCA GAU CGU UUU GGC AUG CGC ACG GGU ACU UGG ACG GGA CAA UUG CAA UAU CAG CAC CCA		1051
	gln leu ser trp arg ala asn val thr leu asn leu met lys val asp asp trp leu val leu ser phe ser gln met thr thr asn ser ile met ala asp gly	381	
	CAA UUG AGU UCG AGA GCA AAU GUC ACU UUG AAU UUG AUG AAG GUG GAU GAU UGG UUG GUG UUG AGC UUU UCU CAG AUG ACG ACU AAC UCA AUA AUG GCA GAU GGG		1156
	lys phe val ile asn phe val ser gly leu ser ser gly trp gln thr gly asp thr glu pro ser ser thr ile asp pro leu ser thr thr phe ala ala val	416	
	AAA UUU GUG AUU AAU UUU GUG UCU GGG UUA UCU UCU GGA UGG CAG ACG GGG GAU ACU GAA CCA UCG UCA ACU AUU GAU CCA UUG UCU ACG ACA UUU GCC GCG GUC		1261
	gln phe leu asn asn gly gln arg ile asp ala phe arg ile met gly val ser glu trp thr asp gly glu leu glu ile lys asn tyr gly gly thr tyr thr	451	
	CAA UUU CUA AAU AAC GGU CAA CGC AUU GAU GCG UUU AGG AUC AUG GGA GUA UCG GAA UGG ACG GAU GGA GAA UUA GAG AUU AAG AAU UAU GGU GGC ACA UAC ACC		1366
	gly his thr gln val tyr trp ala pro trp thr ile met tyr pro cys asn val arg	470	
	GGU CAU ACU CAA GUA UAU UGG GCU CCG UGG ACG AUC AUG UAU CCA UGC AAU GUG AGG <u>UGA</u> AUCUAGCGCGAACCUCGGCACAAGGGGUCAAUCAUC		1463

FIG. 1. The coding strand nucleotide sequences of the reovirus T1L (A) and T2J (B) S1 gene segments and the respective deduced σ_1 and σ_1 s amino acid sequences. The proposed start and stop codons for the translation of each σ_1 and σ_1 s protein are underlined. The deduced σ_1 amino acid sequences appear above and the deduced σ_1 s amino acid sequences appear below the nucleotide sequences. The S1 sequences differ from those previously published for T1L and T2J by Cashdollar et al. (7) in that they include an additional U between positions 1240 and 1241, an additional C between positions 1312 and 1313, and an additional CCC between positions 1436 and 1437 in T1L S1 (the latter insertion also being reported by Munemitsu et al. [28]); lack the previously reported G at nucleotide position 1121 and the previously reported G at nucleotide position 1258 in T2J S1; and include a small number of nucleotide substitutions that yield changes in the deduced amino acid sequences of both the T1L and the T2J σ_1 and σ_1 s proteins.

positions long) that occurs in a region spanning the middle of σ_1 (alignment positions 209 to 322). The remaining carboxy-terminal region of the aligned σ_1 sequences contains few gaps.

(ii) **Sequence identity among σ_1 proteins.** We examined the degree of identity shared by the T1L, T2J, and T3D σ_1 proteins. Each of the two-sequence comparisons, as well as the three-sequence comparison, was scored for the number of alignment positions at which amino acid residues are absolutely conserved (Table 1). The results indicate that, as previously noted by Cashdollar et al. (7), the T1L and T2J σ_1 proteins are more similar to one another than either is to T3D. The number of positions at which residues are conserved in all three sequences is small.

The positions at which residues are absolutely conserved in the different σ_1 proteins are not evenly distributed throughout the protein lengths. A comparison of T1L and T2J σ_1 (Fig. 3) identified two large regions of more-similar sequences (alignment positions 184 to 248 and 344 to 471). The other two-sequence comparisons and the three-sequence comparison, in contrast, did not identify any such large regions of similarity; sequences near alignment posi-

tion 400 received the highest identity scores in these comparisons (data not shown). These findings suggest that some regions of σ_1 have undergone less sequence change during the divergence of the σ_1 proteins of these strains than other regions.

(iii) **Dot matrix self-comparisons of σ_1 .** The deduced amino acid sequence of each σ_1 protein was compared with itself by the dot matrix technique (24). These analyses suggest that a large region of sequence in the amino-terminal half of each σ_1 protein is made up of repeats (Fig. 4). This sequence region extends from near the amino terminus to approximately amino acid 190 in T1L, amino acid 185 in T2J, and amino acid 180 in T3D. The region encompasses the heptad repeats proposed to form a long α -helical coiled coil in each protein (4, 17). The carboxy-terminal boundary of this region corresponds well with the carboxy-terminal extent of the heptad repeats indicated by Furlong et al. (17); however, its amino-terminal boundary suggests that the heptad repeats extend nearly to the amino terminus of σ_1 .

The dot matrix analyses also suggest that additional sequence repeats occur in a discrete region of each σ_1 protein that is carboxy terminal to the identified heptad repeats:

B	met ser asp leu val gln leu ile arg arg glu ile leu leu leu thr gly asn gly glu ser ala asn ser lys his glu ile glu glu ile	31	106
	GCUAUCCGACUC AUG UCG GAU CUA GUG CAG CUC AUA AGA AGG GAG AUC UUA CUG UUA ACU GGG AAU GGA GAA UCA GCC AAC UCG AAA CAC GAG AUC GAG GAA AUU		
	met glu asn gln pro thr arg asn thr arg ser arg lys leu	14	
	lys lys gln ile lys asp ile ser ala asp val asn arg ile ser asn ile val asp ser ile gln gly gln leu gly gly leu ser val arg val ser ala ile	66	211
	AAG AAA CAA AUU AAA CAC AUC UCU GCU GAU GUC AAC AGG AUC AGU AAC AUC GUU GAU UCA AUC CAA GGA CAA CUG GGU GGA UUA UCU GUA CGC GUG UCA GCC AUU		
	arg asn lys leu lys thr ser leu leu met ser thr gly ser val thr ser leu ile gln ser lys asp asn trp val asp tyr leu tyr ala cys gln pro leu	49	
	glu ser gly val ser glu asn gly asn arg ile asp arg leu glu arg asp val ser gly ile ser ala ser val ser gly ile asp ser arg leu ser glu leu	101	316
	GAA UCG GGA GUU AGU GAG AAC GGC AAU CGA AUU GAU AGA CUC GAG CGA GAU GUC UCC GGC AUA UCG GCU AGC GUU AGC GGA AUC GAU UCG CGU UUA UCC GAG CUG		
	asn arg glu leu val arg thr ala ile glu leu ile asp ser ser glu met ser pro ala tyr arg leu ala leu ala glu ser ile arg val tyr pro ser trp	84	
	gly asp arg val asn val ala glu gln arg ile gly gln leu asp thr val thr asp asn leu leu glu arg ala ser arg leu glu thr glu val ser ala ile	136	421
	GGU GAC CGA GUC AAU GUU GCA GAA CAG CGA AUU GGC CAG UUG GAU ACA GUC ACG GAU AAU CUC CUU GAG CGA GCA UCA AGA CUG GAA ACU GAA GUA UCA GCC AUU		
	val thr glu ser met leu gln asn ser glu leu ala ser trp ile gln ser arg ile ile ser leu ser glu his gln asp trp lys leu lys tyr gln pro leu	119	
	thr asn asp leu gly ser leu asn thr arg val thr thr glu leu asn asp val arg gln thr ile ala ala ile asp thr arg leu thr thr leu glu thr asp	171	526
	ACU AAU CAC CUU GGA UCA UUG AAU ACG AGG GUG ACG ACU GAA UUG AAC GAU GUC CGC CAA ACU AUU GCU GCG AUA GAC ACG CGU CUC ACG ACA CUG GAG ACC GAU		
	leu met thr leu asp his	125	
	ala val thr ser val gly gln gly leu gln lys thr gly asn ser ile lys val ile val gly thr gly met trp phe asp arg asn asn val leu gln leu phe	206	631
	GCC GUG ACG UCG GUU GGU CAA GGG CUU CAG AAG ACU GGG AAC UCG AUU AAG GUU AUU GUG GGU ACG GGG AUG UGG UUC GAC CGC AAU AAU GUU CUG CAG UUA UUC		
	val ser asn gln gln lys gly leu gly phe ile asp asn gly met val val lys ile asp thr gln tyr phe ser phe asp ser asn gly asn ile thr leu asn	241	736
	GUA UCG AAC CAG CAG AAA GGG UUG GGA UUC AUA GAC AAU GGA AUG GUA GUG AAA AUA GAU ACC CAG UAU UUC AGC UUC GAU AGC AAU GGC AAC AUA ACU CUG AAC		
	asn asn ile ser gly leu pro ala arg thr gly ser leu glu ala ser arg ile asp val val ala pro pro leu val ile gln ser thr gly ser thr arg leu	276	841
	AAC AAC AUA AGU GGU CUG CCG GCG CGA ACA GGU UCC CUC GAG GCA UCU CGU AUC GAU GUG GUA GCG CCA CCG CUU GUG AUA CAG UCU ACU GGU AGC ACU CCG CUA		
	leu arg leu met tyr glu ala val asp phe val val thr asn asn val leu thr leu arg asn arg ser val thr pro thr phe lys phe pro leu glu leu asn	311	946
	CUG CGU CUC AUG UAC GAG GCU GUG GAC UUC GUG GUU ACU AAC AAC GUU CUC ACA CUG AGA AAU CGA UCG GUC ACG CCA ACA UUC AAG UUU CCU CUG GAG UUG AAU		
	ser ala asp asn ser val ser ile his arg asn tyr arg ile arg leu gly gln trp ser gly gln leu glu tyr his thr pro ser leu arg trp asn ala pro	346	1051
	AGU GCU GAU AAC UCA GUG AGC AUU CAU AGA AAU UAC CGC AUU AGA CUU GGG CAA UGG UCA GGU CAA UUG GAA UAU CAC ACG CCG AGU UUG CGU UGG AAU GCU CCC		
	val thr val asn leu met arg val asp asp trp leu ile leu ser phe thr arg phe ser thr ser gly ile leu ala ser gly lys phe val leu asn phe val	381	1156
	GUC ACG GUU AAU UUG AUG CGA GUA GAC GAU UGG CUC AUU UUG AGU UUU ACU CGG UUU UCG ACG AGC GGC AUC UUA GCG UCA GGA AAG UUU GUA UUG AAC UUC GUA		
	thr gly leu ser pro gly trp ala thr gly ser thr glu pro ser thr thr thr asn pro leu ser thr thr phe ala ala ile gln phe ile asn gly ser ser	416	1261
	ACU GGU UUG UCU CCA GGG UGG GCG ACU GGG AGU ACC GAG CCC UCG ACA ACU ACU AAC CCA CUG UCA ACG ACG UUU GCU GCA AUU CAG UUC AUC AAU GGG UCA UCU		
	arg val asp ala phe arg ile leu gly val ala glu trp asn ala gly glu leu glu ile thr asn tyr gly gly thr tyr thr ala his thr asn val asp trp	451	1366
	CGC GUA GAC GCC UUU AGA AUC UUG GGA GUC GCA GAG UGG AAU GCC GGG GAA CUA GAG AUC ACG AAU UAU GGC GGA ACA UAU ACA GCG CAU ACC AAU GUC GAC UGG		
	ala pro met thr ile met tyr pro cys leu gly	462	1440
	GCG CCG AUG ACC AUU AUG UAC CCA UGU CUG GGC UGA GGAUCCGGGUGUCCACUCGCGACAGUGGCGACUACUC		

FIG. 1—Continued.

approximately amino acids 245 to 305 in T1L, amino acids 240 to 300 in T2J, and amino acids 230 to 290 in T3D (Fig. 4). The sequences in this region are related to the repeats in the amino-terminal region; however, the degree of relatedness between the two regions varies in the different $\sigma 1$ proteins, being greatest in T3D and in an initial portion of the carboxy-terminal repeat region of T1L.

The carboxy-terminal repeat region of each $\sigma 1$ protein is separated from the amino-terminal repeat region by an intervening region that is almost devoid of internal sequence similarity. This intervening region corresponds approximately to amino acids 190 to 245 in T1L, amino acids 185 to 240 in T2J, and amino acids 180 to 230 in T3D. The remaining carboxy-terminal one-third of each $\sigma 1$ protein shows little similarity to the other repeat regions. Thus, dot matrix self-comparisons suggest that each $\sigma 1$ protein is divided into at least four discrete regions.

(iv) **Secondary and supersecondary structures suggested by patterns of apolar residues in discrete regions of $\sigma 1$.** Repeating apolar residues are an important determinant of protein secondary and supersecondary structure (22, 27). Recent work has confirmed the critical importance of apolar residues in determining protein structure and stability (23). We examined the aligned $\sigma 1$ sequences for positions at which any of a set of apolar residues (A, F, I, L, M, V, W, and Y) occur in all three sequences (Fig. 2). This technique favors the identification of repeats of apolar residues by eliminating the noise of apolar residues that occur sporadically in individual $\sigma 1$ sequences.

(a) **Apolar repeats, stutter and skip sequences, and acid-**

base pairs in the proposed long coiled-coil region of $\sigma 1$. When the sequences for a large amino-terminal portion of the prototype $\sigma 1$ proteins were written in heptads (Fig. 5) (27), a pattern of repeating apolar residues (noted by closed circles in Fig. 2) was identified in which apolar residues regularly occupy the *a* and *d* positions. This pattern was described previously in the T3D (4) and T2J (17) $\sigma 1$ sequences. By providing hydrophobic interactions between two α -helices (Fig. 6), these apolar residues could contribute most of the thermodynamic force needed to stabilize a coiled-coil dimer formed by this region of $\sigma 1$ (27).

Each $\sigma 1$ protein contains discontinuities in the apolar repeat pattern of the heptad region (Fig. 5): a stutter sequence like that described in the intermediate filament proteins (40) and a skip sequence like that described in myosin (26). When the $\sigma 1$ amino acid sequence is written as a progression of heptads, the stutter is seen as an abrupt switch of apolar residues from the *a* and *d* to the *e* and *a* positions. The skip, which follows the stutter in this region of $\sigma 1$, is seen as a switch of apolar residues from the *e* and *a* to the *f* and *b* positions. The stutter sequence is equivalent to a four-residue insertion or a three-residue deletion in the progression of heptads; the skip sequence is equivalent to a one-residue insertion or a six-residue deletion. The consequences of stutter and skip sequences for coiled-coil structure are suggested in Fig. 6.

The stutter sequence in T3D $\sigma 1$ was not previously noted by Bassel-Duby et al. (4), perhaps because only 1.5 in-phase heptads follow the stutter in this protein. In T1L and T2J $\sigma 1$, however, 4.5 and 3.5 in-phase heptads, respectively, follow

FIG. 2. Alignment of the deduced $\sigma 1$ amino acid sequences of T1L (1L), T2J (2J), and T3D (3D). The single-letter amino acid code is used. Alignment positions are numbered beneath the sequences; amino acid positions at the end of each line of individual sequence are numbered on the right. Gaps in the aligned sequences are indicated (.). Symbols: (●), positions of conserved apolar residues that conform to a heptad repeat pattern; (○), positions of conserved apolar residues that conform to an alternating apolar-polar repeat pattern; (*), positions of conserved apolar residues that do not conform to a repeat pattern. Closed and open circles with lines beneath indicate positions in the repeats at which apolar residues are not conserved in all three sequences. Acid-base pairs are connected by underlines in regions of sequence characterized by heptad repeats.

the stutter (Fig. 5). The skip sequence is observed at the carboxy-terminal boundary of the long heptad repeats in each $\sigma 1$ protein and further increases the amount of sequence that may be considered to be included in this region by 0.5 heptads. Including the heptads that follow the stutter and skip sequences, therefore, the proposed boundaries of the long heptad repeat region in each $\sigma 1$ protein are as follows: T1L, amino acids 27 to 183 (22.5 heptads, ~44 α -helical turns); T2J, amino acids 28 to 176 (21.5 heptads, ~42 α -helical turns); and T3D, amino acids 28 to 173 (21 heptads, ~41 α -helical turns).

TABLE 1. Sequence identity among $\sigma 1$ proteins of reovirus prototype strains

Proteins compared	% Identity among:				
	Whole $\sigma 1$ proteins	$\sigma 1$ protein region ^a			
		T(i+ii)	T(iii)	T(iv)	H
T1L:T2J	49	38	60	41	65
T1L:T3D	26	25	24	24	28
T2J:T3D	27	28	22	20	31
T1L:T2J:T3D	17	14	19	11	24

^a $\sigma 1$ sequence regions ascribed to protein regions as summarized in the legend to Fig. 11.

In the alignment of $\sigma 1$ amino acid sequences (Fig. 2) (7), T1L has a 14-residue and T2J has a 7-residue insertion relative to T3D in the region of the stutter sequence. These insertions contribute to the difference in the number of heptads that follow the stutter in the three $\sigma 1$ proteins (Fig. 5). The variation of sequence length in multiples of heptads is consistent with the proposed α -helical coiled-coil structure

FIG. 3. Identity scores at each position along the aligned T1L and T2J $\sigma 1$ sequences. The 50% identity level is marked with a horizontal line.

FIG. 4. Dot matrix self-comparisons of the T1L, T2J, and T3D $\sigma 1$ amino acid sequences. Amino acid positions in the compared sequences are indicated on the right and upper axes. A diagonal line of identity is apparent in the center of each graph. Lines occurring parallel to the central diagonal indicate related sequences that occur in different parts of each protein.

of this region of $\sigma 1$. Furthermore, the conservation of the stutter sequence, despite insertions in T1L and T2J relative to T3D $\sigma 1$, suggests that it is important for $\sigma 1$ protein structure in this region.

The reovirus $\sigma 1$ proteins contain several acid-base pairs (connected by underlines in Fig. 2 and indicated by connected circles in Fig. 5) in the $(n)g$ and $(n+1)e$ positions of their long heptad repeat regions (where n and $n+1$ indicate any two successive heptads). Residues in these positions in two adjacent α -helices in a coiled coil are closely positioned (Fig. 6), and acid-base pairs in these positions might contrib-

ute ionic interactions to the stabilization of a coiled coil (27). Following a stutter sequence in a continuous progression of heptad repeats, acid-base pairs that are appropriately positioned for interhelical interaction in a coiled coil occur in the b and d , rather than in the e and g , positions (Fig. 6). $b-d$ acid-base pairs occur following the stutter sequence in all three $\sigma 1$ proteins (Fig. 2 and 5). Thus, acid-base pairs in the long heptad repeat regions of the three $\sigma 1$ proteins could provide ionic interactions to a coiled-coil dimer at an average frequency of one ionic interaction per every three α -helical turns. The large number of $e-g$ and $b-d$ acid-base pairs in the heptad repeat regions of the three $\sigma 1$ proteins is consistent with the conclusion that this region of $\sigma 1$ sequence, including sequence to the carboxy-terminal side of the stutter sequence, forms an α -helical coiled coil.

(b) Region of proposed cross- β structure with similarity to adenovirus fiber protein in the middle of $\sigma 1$. A distinct pattern of repeating apolar residues was identified in the three $\sigma 1$ proteins in a region immediately carboxy terminal to the long heptad repeats (alignment positions 188 to 255). Conserved apolar residues in this region (noted by open circles in Fig. 2) occur in small clusters that are separated from one another by short lengths of polar residues. Each cluster contains either two or three conserved apolar residues occurring at every other amino acid position; the residues between apolar residues in these clusters are mostly polar. Eight of these clusters occur in this region, the first seven of which alternate between containing two and three apolar residues each (2-3-2-3-2-3-2-2). Alternating polar and apolar residues is a pattern characteristic of β -strands in many proteins (1, 22). The small clusters of apolar residues in $\sigma 1$ are therefore suggestive of short β -strands. In addition, the short polar regions that separate these clusters contain residues that are commonly found in β -turns (8). These observations support the prediction that, in contrast to the preceding proposed α -helical coiled-coil region, sequences between alignment positions 188 and 255 in $\sigma 1$ form a series of alternating short β -strands and β -turns.

In this region of sequence, no gaps were inserted to maintain an optimal alignment between the T1L and T2J $\sigma 1$ sequences; however, three single-position gaps were inserted in the T3D sequence to maintain an optimal alignment between it and the other two sequences (Fig. 2). A fourth single-position gap was inserted immediately following this region of sequence in both T1L and T2J in order to maintain their alignments with T3D. These gaps do not disrupt the pattern of conserved apolar residues in the three sequences; the gaps were placed in the lengths of polar residues that separate the apolar clusters. Thus, according to the above prediction, gaps in this region are accommodated as differences in the lengths of sequences forming β -turns and do not affect the lengths of the intervening β -strands.

A model for the alternating β -strand/ β -turn structure of the three $\sigma 1$ proteins in this region is shown in Fig. 7 and is equivalent to the "cross- β " protein conformation discussed by Geddes et al. (18). A feature of this model is a hydrophobic face formed by apolar residues potentially projecting from one side of the extended β -sheet. Fifty-five of 57 residues in these positions in the three $\sigma 1$ proteins are apolar; only lysine 189 in T1L $\sigma 1$ and lysine 182 in T2J $\sigma 1$ are not consistent with this pattern. The amphipathicity of the proposed β -sheet in each $\sigma 1$ protein suggests that hydrophobic interactions between $\sigma 1$ subunits in a cross- β sandwich conformation may stabilize a $\sigma 1$ dimer formed by this region of sequence.

Green et al. (19) proposed a similar cross- β sandwich

FIG. 5. An amino-terminal region in the T1L, T2J, and T3D $\sigma 1$ amino acid (AA) sequences in which conserved apolar residues conform to a heptad repeat pattern. The sequences of each protein are written in heptads. *a*, *b*, *c*, *d*, *e*, *f*, and *g*. The seven amino acid positions in each heptad. Apolar residues (A, F, I, L, M, V, W, and Y) are capitalized throughout the sequences. Repeat positions predominantly occupied by apolar residues are circled. A stutter in the heptad progression occurs where paired apolar residues switch from the *a* and *d* to the *e* and *a* positions. A skip in the heptad progression occurs at the carboxy-terminal end of this sequence where paired apolar residues switch from the *e* and *a* to the *f* and *b* positions. Acid-base pairs in the *e* and *g* and *b* and *d* positions of consecutive heptads (pre- and poststutter, respectively) are circled and connected with lines.

structure for the entire length of the rod domain of the adenovirus type 2 fiber protein. The deduced amino acid sequences of fiber proteins from two other adenovirus strains are consistent with this proposal (9, 38). We compared adenovirus fiber proteins and reovirus $\sigma 1$ proteins for regions of sequence similarity. Figure 8 illustrates a similar region that was found in a portion of the rod domain of the adenovirus type 3 fiber protein and a portion of the proposed cross- β region of reovirus T1L $\sigma 1$. Polar and apolar clusters corresponding to proposed β -turns and β -strands, respectively, are conserved in these two sequences.

(c) Short α -helical coiled coils and a smaller cross- β region proposed in the middle of $\sigma 1$. A short region of $\sigma 1$ sequence immediately carboxy terminal to the proposed cross- β region (alignment positions 256 to 273) has the features of an α -helical coiled coil. Apolar residues in this region (noted by closed circles in Fig. 2) occur in a heptad repeat pattern defining three heptads. This region incorporates a skip

sequence as apolar residues in the most carboxy-terminal heptad switch from the *a* and *d* to the *b* and *e* positions. In addition, an *e*-*g* acid-base pair (arginine at alignment position 262 and glutamate at alignment position 267) is conserved in these heptads. These features support the conclusion that this region of sequence forms an α -helical coiled coil. Immediately carboxy terminal to the three heptads are found a proline-proline pair in the T1L and T2J $\sigma 1$ proteins and a threonine-proline pair in the T3D $\sigma 1$ (alignment positions 276 and 277); these residues may serve to demarcate the end of this short coiled coil.

Carboxy terminal to the proposed short coiled coil, an apolar repeat pattern suggesting alternating β -strands and β -turns recurs (alignment positions 277 to 312). This region includes four clusters of either two or three conserved apolar residues alternating with polar residues (2-3-2-2; conserved apolar residues noted by open circles in Fig. 2). Gaps that optimally align the three $\sigma 1$ sequences in this region were

FIG. 6. (A) Schematic coiled coil. The dimensions of an ideal α -helix (~ 3.6 residues per turn) dictate that apolar residues in positions *a* and *d* of a heptad forming an α -helix occur on the same side of a cylinder defined by the helix. This positioning of apolar residues gives the helix a hydrophobic face. Interaction between the hydrophobic faces of two adjacent α -helices in the coiled coil (suggested by dotted lines) partially removes the apolar residues from the aqueous environment and is expected to provide the major force stabilizing a dimer of α -helices. Since the residues at other heptad positions (*b*, *c*, *e*, *f*, and *g*) project into the surrounding solution, they are usually polar. Residues in the *e* and *g* positions are properly positioned to provide additional interactions between the two α -helices, such as ionic interactions between acidic and basic residues in these positions. (B and C) Effects of stutter (B) and skip (C) sequences on coiled-coil structures. Switching apolar residues in a progression of heptads from positions *a* and *d* to positions *e* and *a* (a stutter) or to positions *b* and *e* (a skip) necessitates a reorientation of the α -helices in the coiled coil if the helices are to continue to interact via their hydrophobic faces. Stutter and skip sequences serve to decrease the superhelical twist of the coiled coil by $\sim 50^\circ$ and $\sim 100^\circ$, respectively. After a stutter sequence, polar residues (e.g., acid-base pairs) in positions *b* and *d*, rather than in positions *e* and *g*, are properly positioned to provide additional interactions between α -helices.

inserted only in proposed β -turns. Thus, the model of a cross- β sandwich, with hydrophobic interactions occurring between amphipathic β -sheets to stabilize a $\sigma 1$ dimer, applies to this region as well as to the more amino-terminal cross- β region. Alignment positions 317 to 319 represent another conserved apolar cluster that may contribute an additional β -strand to the proposed second cross- β region; however, this cluster is bounded on both sides by conserved proline residues (alignment positions 315 and 320) and may adopt a more open conformation.

A region of $\sigma 1$ sequence between alignment positions 323 and 333 may also form a short α -helical coiled coil. Apolar residues (indicated by closed circles in Fig. 2) form a heptad repeat pattern defining two heptads. An *e-g* acid-base pair is found in the T1L $\sigma 1$, but no comparable acid-base pair occurs in either the T2J or the T3D $\sigma 1$ protein. A model for the proposed structure of the three $\sigma 1$ proteins between alignment positions 256 and 336 is shown in Fig. 9.

(d) Sequence patterns in the remaining amino-terminal and carboxy-terminal regions of $\sigma 1$. The most amino-terminal region of the three $\sigma 1$ proteins (alignment positions 5 to 16) also has features of an α -helical coiled coil. Apolar residues in this region (noted by closed circles in Fig. 2) occur in a heptad repeat pattern defining two heptads, but *e-g* acid-base pairs are not found. This proposed short amino-terminal coiled-coil region is separated from the long coiled-coil region by 11 alignment positions that are occupied primarily by polar residues.

There is no consistent pattern of conserved apolar residues carboxy terminal to $\sigma 1$ alignment position 334. Clusters containing from two to four apolar residues occurring at

every other amino acid position (suggesting β -strand conformation) are observed frequently in this region of sequence; however, these clusters do not conform to a regular repeat pattern like that described for the proposed cross- β regions. The lack of a pattern of conserved apolar residues supports the conclusion that this region of $\sigma 1$ forms a complex globular domain.

(v) Averaged secondary structure and hydrophobicity scores. We investigated whether secondary structure prediction techniques such as that of Chou and Fasman (8) would contribute to our characterization of the domain organization of $\sigma 1$. The accuracy of such secondary structure predictions can be improved by averaging scores across a set of multiple related sequences (1, 2). Therefore, predictions of $\sigma 1$ secondary structure were made by first defining Chou-Fasman scores (α -helix, β -strand, and β -turn) for each amino acid position in the individual $\sigma 1$ sequences (data not shown) and then averaging the scores for each position in the three-sequence alignment (Fig. 10). An identical procedure was performed by using the values of Rose et al. (35) to obtain an average hydrophobicity score at each alignment position (Fig. 10).

The results of these analyses again suggest that the $\sigma 1$ sequence is divided into regions having distinct characteristics. Alignment positions 1 to 20 represent a region of sequence that is characterized by high α -helix, β -strand, and hydrophobicity scores and low β -turn scores. This region corresponds to the proposed short α -helical region at the amino terminus of $\sigma 1$. A small region of sequence near alignment position 25, in contrast, has high β -turn scores and low α -helix, β -strand, and hydrophobicity scores. It may

FIG. 7. A region in the middle of the T1L, T2J, and T3D $\sigma 1$ amino acid sequences in which conserved apolar residues conform to an alternating apolar-polar repeat pattern. The sequences of each protein are written in such a way as to suggest a region of extended β -sheet secondary structure that is formed from a series of alternating short β -strands and β -turns. Apolar residues are capitalized throughout the sequences. Repeat positions mostly occupied by apolar residues in the proposed β -strands are circled. :, Hydrogen bonds proposed to occur between consecutive β -strands.

correspond to a short hinge region near the amino terminus. Alignment positions 30 to 190 have α -helix scores that remain consistently above a score of 0.90, in contrast to the rest of the sequence. This region is also hydrophilic and has low β -strand and β -turn scores. It corresponds to the proposed long α -helical coiled-coil region. Alignment positions 190 to 250 and 270 to 310 have periodically high β -strand and β -turn scores that fluctuate out of phase with each other. These regions also have low α -helix scores and contain intervals of greater hydrophobicity than the surrounding sequences. They correspond to the two proposed cross- β -sheet regions in the middle of $\sigma 1$. Small regions near alignment positions 260 and 320 are each marked by an increase in the α -helix score and low β -strand, β -turn, and hydrophobicity scores. These small regions correspond to the two proposed short α -helical coiled-coil regions in the middle of $\sigma 1$. Alignment positions 330 to 478 have irregularly fluctuating scores of all types and correspond to the pro-

posed carboxy-terminal globular domain. Thus, averaged secondary structure and hydrophobicity scores may be used to support our conclusions concerning the domain organization of $\sigma 1$.

DISCUSSION

Model for the domain organization of $\sigma 1$. In this report, we identify a number of regions with distinct characteristics that occur in the $\sigma 1$ amino acid sequence (summarized in Fig. 11A to G). In many cases, these characteristics suggest that a predominant secondary (α -helix or β -sheet) or supersecondary (α -helical coiled coil or cross- β sandwich) structure is formed by each region. Thus, these sequence regions might be expected to represent independent structural units, or domains, in the $\sigma 1$ protein.

The observation that the $\sigma 1$ sequence is divided into discrete regions does not directly address how the structural

FIG. 8. A region of similarity between the reovirus T1L $\sigma 1$ protein and the adenovirus type 3 fiber protein. Identical amino acid residues (*) and closely related residues (10) (:) are indicated. Positions of apolar residues proposed to occur in β -strands in the two proteins are shown (O) above the sequences.

FIG. 9. A region in the middle of the T1L, T2J, and T3D $\sigma 1$ amino acid sequences in which conserved apolar residues conform to alternating repeat patterns: heptad repeats followed by alternating polar-apolar repeats followed by heptad repeats. The sequences of each protein are written and labeled as described in the legends to Fig. 5 and 7.

domains are spatially arranged in the $\sigma 1$ protein. Considering the tail-and-head morphology of $\sigma 1$ (3, 16, 17), we now suggest a model in which the structural domains are arranged end-to-end along the length of the extended $\sigma 1$ protein. According to this model, the amino-terminal sequences of $\sigma 1$ are located near the tail end of the protein and the carboxy-terminal sequences are located in the head. We propose that the regions of $\sigma 1$ sequence correlate with morphologic regions, T(i), T(ii), T(iii), T(iv), and H, seen in computer-processed electron micrographic images of T2J $\sigma 1$ (Fig. 11H and I) (16). Our findings suggest that these structural features are shared by $\sigma 1$ proteins from all three reovirus serotypes.

Furlong et al. (17) made observations suggesting that the $\sigma 1$ protein can be associated with viral particles in both an extended and a retracted state. Therefore, $\sigma 1$ appears capable of assuming two distinct conformations. The current study ascribes sequence regions to morphologic regions seen in images of extended $\sigma 1$ but does not address the structure of retracted $\sigma 1$.

Correlation of sequence and morphologic regions. We predict that a small region of sequence at the amino terminus of $\sigma 1$ (alignment positions 5 to 16) forms a short α -helical coiled coil and correlates with morphologic region T(i). This region is also hydrophobic, which probably accounts for the tail-

to-tail rosettes that form in preparations of purified $\sigma 1$ (3, 16). The region is bounded at its carboxy terminus by another small region (alignment positions 17 to 27) having high β -turn and Ω -loop (21; data not shown) scores. A turn/loop in this position may provide a hinge between the proposed short amino-terminal coiled coil and the succeeding long coiled coil. We postulate that the amino-terminal coiled coil and the associated turn/loop include sequences that anchor the extended $\sigma 1$ protein to viral particles (17).

This study extends an examination of the proposed long coiled-coil region (alignment positions 28 to 187) to $\sigma 1$ proteins from prototype strains of all three reovirus serotypes. Features not previously described by Bassel-Duby et al. (4) are identified, notably stutter and skip sequences and acid-base pairs in the heptad repeats. We predict that the proposed long coiled-coil region forms morphologic region T(ii).

A stutter sequence occurs in the coiled-coil rod domains of the three fibrinogen proteins (13), the bacteriophage T7 tail fiber protein (41), and all intermediate filament proteins (32, 40); skip sequences occur in myosin (26). Both stutter and skip sequences have been proposed to be accommodated in a coiled coil by a rotation of the α -helical strands that reduces the twist of the superhelix (26, 32). The function of

FIG. 10. Averaged secondary structure and hydrophobicity scores for the aligned $\sigma 1$ amino acid sequences. Secondary structure propensity scores (α -helix, β -strand, and β -turn) (8) and hydrophobicity scores (35) were averaged over the T1L, T2J, and T3D $\sigma 1$ sequences. The neutral preference value for secondary structures (1.00) and the mean hydrophobicity value (1.00) are marked with horizontal lines in the respective graphs. X-axis tick marks in the upper graphs correspond to those labeled as to alignment position in the bottom graph. Bars in the α -helix graph indicate where heptad repeats of apolar residues occur in the aligned $\sigma 1$ sequences; bars in the β -strand/ β -turn graph indicate where alternating apolar-polar repeats occur in the aligned $\sigma 1$ sequences.

this effect is unknown but in some cases, including $\sigma 1$, may be important for the packing of coiled-coil dimers in higher-order oligomers (5, 26, 32). The location of a skip sequence at the carboxy-terminal end of the long $\sigma 1$ coiled coil also suggests that it may be important for the transition to a distinct (cross- β) structure formed by the next region of sequence.

Ionic interactions between acid-base pairs in the *e* and *g* (or *b* and *d*) positions of α -helical chains in a coiled coil have been proposed to reinforce the hydrophobic interactions that occur between apolar residues in the *a* and *d* (or *e* and *a*) positions (27). The acid-base pairs in $\sigma 1$ may also provide intercoil (interdimer) interactions in the $\sigma 1$ tetramer (M. L. Nibert, unpublished data).

A region in the middle of $\sigma 1$ sequence (alignment positions 188 to 255) is not characterized by apolar residues in a

heptad repeat pattern and hence is not likely to have an α -helical coiled-coil structure. Instead, this region contains a distinct sequence pattern of small clusters of alternating apolar and polar residues that are separated by small clusters of polar residues. This sequence pattern, which is similar to that found in the rod domain sequences of adenovirus fiber proteins, suggests that the secondary structure of this region includes a series of short β -strands separated by β -turns. Chou-Fasman secondary structure predictions are consistent with the proposal that these sequences form a series of alternating β -strands and β -turns. This region is proposed to form morphologic region T(iii).

Green et al. (19) suggested that rod domain sequences in the adenovirus fiber protein assume a cross- β conformation in which consecutive β -strands associate in an antiparallel fashion to form an extended β -sheet. These authors also

FIG. 11. The $\sigma 1$ protein is divided into discrete regions. (A) The positions of gaps in the $\sigma 1$ three-sequence alignment. (B) Regions of high identity score between T1L and T2J $\sigma 1$ proteins. (C) Regions of internal $\sigma 1$ sequence similarity by dot matrix analysis. (D) Regions exhibiting a heptad repeat pattern of apolar residues. (E) Regions exhibiting an alternating apolar-polar repeat pattern. (F) The positions of conserved proline residues. (G) The positions of conserved aromatic residues. (H) Morphologic regions of $\sigma 1$ identified by electron microscopy (16) as correlated with current sequence analyses. (I) A model of $\sigma 1$ structure. Alignment positions are noted at the bottom and are in scale with each of the panels above.

proposed that the β -sheets formed by two fiber protein polypeptide chains interact in a cross- β sandwich arrangement, in which apolar residues in each β -strand provide intersheet hydrophobic interactions. Extended cross- β structures have also been proposed for β -keratins (37), the insect silk proteins (18), and the distal tail fiber of bacteriophage T4 (14). We propose that β -sheet regions in the $\sigma 1$ tail have a cross- β sandwich structure like that predicted for the adenovirus fiber protein; however, an accurate description of the orientation of β -strands in these regions requires higher-resolution studies.

The most carboxy-terminal sequences proposed to contribute to the $\sigma 1$ tail are expected to form a distinct structural region that we term the neck and that we predict to correlate with morphologic region T(iv). Sequences in this region are proposed to form two short coiled coils (alignment positions 256 to 276 and 323 to 333) separated by a second, smaller cross- β region (alignment positions 277 to 322). The more amino terminal of these short coiled coils incorporates a conserved *e-g* acid-base pair and a conserved skip sequence. The skip sequence is like that found at the carboxy-terminal end of the long heptad repeats in that it immediately precedes a proposed cross- β region. Chou-Fasman secondary structure predictions suggest that both short heptad regions are α -helical. Conserved proline residues border the proposed cross- β region on either side (Fig. 11F) and lend support to the division of the neck into subregions.

We describe this region as the $\sigma 1$ neck for two reasons. First, it is adjacent to the $\sigma 1$ head. Second, it corresponds to the morphologic region in which kinking is most commonly observed in isolated $\sigma 1$ (16) and might therefore contain regions of greater flexibility than the rest of the protein. This flexibility may be important for the capacity of $\sigma 1$ to assume a less extended conformation when packaged in virions (17; D. B. Furlong, unpublished data). Alternatively, flexibility of the neck may be important for permitting changes in the orientation of the $\sigma 1$ head that facilitate receptor binding.

The carboxy-terminal one-third of $\sigma 1$ sequence (alignment positions 334 to 478) is proposed to form morphologic region H, the $\sigma 1$ head. A consistent repeat pattern of conserved apolar residues is not observed in this region of sequence; however, clusters of alternating apolar and polar residues in this region are likely to form a number of β -strands. Consideration of hydrophobicity and secondary structure predictions argue that the head is divided into two distinct portions. The amino-terminal portion (alignment positions 334 to 375) is more hydrophobic and is predicted to be formed from β -strands that are separated by short β -turns. The carboxy-terminal portion (alignment positions 376 to 478) is more hydrophilic and is predicted to be formed from α -helices and β -strands that are mostly separated by large β -turns or Ω -loops (data not shown). The head probably folds so that a large proportion of its surface is formed by the large carboxy-terminal turn/loops (21). The head region is also characterized by a large number of conserved aromatic residues which other regions of $\sigma 1$ do not contain (Fig. 11G).

Several studies have suggested that sequences within the proposed $\sigma 1$ head region are important for the properties of $\sigma 1$ as the serotype-specific protein and cell-attachment protein of reovirus (6, 29, 45, 50, 52). Current data suggest that different reovirus serotypes bind to either different (12, 15, 25, 42, 46) or the same (20, 45) cell-surface receptors. Thus, either especially variable or especially conserved sequences in the $\sigma 1$ head might be expected to represent receptor-binding epitopes. The head is the part of $\sigma 1$ that is most conserved between strains of different serotypes (Table 1). In particular, alignment positions 391 to 409 represent the longest conserved region found in $\sigma 1$ and coincide with the largest of the turn/loops predicted in the $\sigma 1$ head. This region is likely to be exposed at the protein surface and may form a receptor-binding epitope common to all three $\sigma 1$ proteins.

Oligomers of $\sigma 1$. Previous biochemical studies (3, 5) indicate that the $\sigma 1$ protein occurs as an oligomeric complex containing four $\sigma 1$ polypeptide chains, a homotetramer. The studies of Fraser et al. (16) further suggest that the $\sigma 1$ tetramer is formed from two $\sigma 1$ dimers that bond side by side. We postulate that repeating apolar residues in both α -helical coiled-coil and cross- β -sheet regions proposed in the $\sigma 1$ tail are likely to provide a large number of hydrophobic interactions between subunits within each dimer (intradimer interactions). Interactions that occur between $\sigma 1$ dimers (interdimer interactions), in contrast, are likely to include mostly hydrogen and ionic bonds (5, 16; Nibert, unpublished).

Multiple sequences and analysis techniques. This report describes a model for the domain organization of the reovirus $\sigma 1$ protein. The use of multiple $\sigma 1$ amino acid sequences and multiple techniques to analyze these sequences was probably the major determinant of our ability to derive this model. Another likely determinant was the amount of variability in the $\sigma 1$ sequences that were used; enough variability occurs in these sequences to have allowed sequence noise to be eliminated but not enough to have obscured the sequence

patterns proposed to be important for $\sigma 1$ structure. A combination of careful analysis of multiple related amino acid sequences and high-resolution electron microscopy has proved to be a powerful technique for studying the structure of the reovirus cell-attachment protein. Our findings should be very useful for correlating the functions of $\sigma 1$ with structural characteristics of the protein. We have recently applied the model to a study of $\sigma 1$ function as the reovirus hemagglutinin (T. S. Dermody, M. L. Nibert, R. Bassel-Duby, and B. N. Fields, submitted for publication).

ACKNOWLEDGMENTS

We express our appreciation to Elaine Freimont and Melissa Porter for expert technical assistance and to Marcia Kazmierczak for help with preparing the manuscript. We also thank Bruce Fraser, Alasdair Steven, Kevin Coombs, Deirdre Furlong, Leslie Schiff, Barbara Sherry, Ken Tyler, and Skip Virgin for essential discussions and reviews of the manuscript.

This work was supported by Public Health Service research grant 5R37 AI13178-13 from the National Institute of Allergy and Infectious Diseases. Support to B.N.F. was also provided by the Shipley Institute of Medicine. T.S.D. was the recipient of Physician-Scientist Award 5K11 AI00865-01 from the National Institute of Allergy and Infectious Diseases. M.L.N. was supported by National Research Service Award 5T32 GM07753-10 from the National Institute of General Medical Sciences.

LITERATURE CITED

- Argos, P., and J. K. MohanaRao. 1986. Prediction of protein structure. *Methods Enzymol.* **140**:185-207.
- Argos, P., J. Schwarz, and J. Schwarz. 1976. An assessment of protein secondary structure prediction methods based on amino acid sequence. *Biochim. Biophys. Acta* **439**:261-273.
- 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.
- Bassel-Duby, R., A. Jayasuriya, D. Chatterjee, N. Sonenberg, J. V. Maizel, Jr., and B. N. Fields. 1985. Sequence of reovirus hemagglutinin predicts a coiled-coil structure. *Nature (London)* **315**:421-423.
- 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.
- Bassel-Duby, R., D. R. Spriggs, K. L. Tyler, and B. N. Fields. 1986. Identification of attenuating mutations on the reovirus type 3 S1 double-stranded RNA segment with a rapid sequencing technique. *J. Virol.* **60**:64-67.
- Cashdollar, L. W., R. A. Chmelo, J. R. Wiener, and W. K. Joklik. 1985. The sequence of the S1 genes of the three serotypes of reovirus. *Proc. Natl. Acad. Sci. USA* **82**:24-28.
- Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* **47**:251-276.
- Chroboczek, J., and B. Jacrot. 1987. The sequence of adenovirus fiber: similarities and differences between serotypes 2 and 5. *Virology* **161**:549-554.
- Dayhoff, M. O., W. C. Barker, and L. T. Hunt. 1983. Establishing homologies in protein sequences. *Methods Enzymol.* **91**:524-545.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Dichter, M. A., and H. L. Weiner. 1984. Infection of neuronal cell cultures with reovirus mimics in vitro patterns of neurotropism. *Ann. Neurol.* **16**:603-610.
- Doolittle, R. F., D. M. Goldbaum, and L. R. Doolittle. 1978. Designation of sequences involved in the "coiled-coil" interdomainal connections in fibrinogen: construction of an atomic scale model. *J. Mol. Biol.* **120**:311-325.
- Earnshaw, W. C., E. B. Goldberg, and R. A. Crowther. 1979. The distal half of the tail fibre of bacteriophage T4: rigidly linked domain and cross- β structure. *J. Mol. Biol.* **132**:101-131.
- Epstein, R. L., M. L. Powers, R. B. Rogart, and H. L. Weiner. 1984. Binding of 125 I-labeled reovirus to cell surface receptors. *Virology* **133**:46-55.
- Fraser, R. D. B., D. B. Furlong, B. L. Trus, M. L. Nibert, B. N. Fields, and A. C. Steven. 1990. Molecular structure of the cell-attachment protein of reovirus: correlation of computer-processed electron micrographs with sequence-based predictions. *J. Virol.* **64**:2990-3000.
- 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.
- Geddes, A. J., K. D. Parker, E. D. T. Atkins, and E. Beighton. 1968. "Cross- β " conformation in proteins. *J. Mol. Biol.* **32**:343-358.
- 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.
- Lee, P. W. K., E. C. Hayes, and W. K. Joklik. 1981. Protein $\sigma 1$ is the reovirus cell attachment protein. *Virology* **108**:156-163.
- Leszczynski, J. F., and G. D. Rose. 1986. Loops in globular proteins: a novel category of secondary structure. *Science* **234**:849-855.
- Lim, V. I. 1974. Algorithms for prediction of α -helical and β -structural regions in globular proteins. *J. Mol. Biol.* **88**:873-894.
- Lim, W. A., and R. T. Sauer. 1989. Alternative packing arrangements in the hydrophobic core of λ repressor. *Nature (London)* **339**:31-36.
- Maizel, J. V., Jr., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA* **78**:7665-7669.
- Maratos-Flier, E., C. R. Kahn, D. R. Spriggs, and B. N. Fields. 1983. Specific plasma membrane receptors for reovirus on rat pituitary cells in culture. *J. Clin. Invest.* **72**:617-621.
- McLachlan, A. D., and J. Karn. 1983. Periodic features in the amino-acid sequence of nematode myosin rod. *J. Mol. Biol.* **164**:605-626.
- McLachlan, A. D., and M. Stewart. 1975. Tropomyosin coiled-coil interactions: evidence for an unstaggered structure. *J. Mol. Biol.* **98**:293-304.
- Munemitsu, S. M., J. A. Atwater, and C. E. Samuel. 1986. Biosynthesis of reovirus-specified polypeptides: molecular cDNA cloning and nucleotide sequence of the reovirus serotype 1 Lang strain bicistronic s1 mRNA which encodes the minor capsid polypeptide $\sigma 1a$ and the nonstructural polypeptide $\sigma 1bNS$. *Biochem. Biophys. Res. Commun.* **140**:508-514.
- 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 deletion mutants. *Virology* **160**:162-168.
- Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**:443-453.
- Onodera, T., A. Toniolo, U. R. Ray, A. B. Jensen, R. A. Knazek, and A. L. Notkins. 1981. Virus-induced diabetes mellitus. XX. Polyendocrinopathy and autoimmunity. *J. Exp. Med.* **153**:1457-1473.
- Parry, D. A. D., and R. D. B. Fraser. 1985. Intermediate filament structure. I. Analysis of IF protein sequence data. *Int. J. Biol. Macromol.* **7**:203-213.
- Ramig, R. F., R. Cross, and B. N. Fields. 1977. Genome RNAs and polypeptides of reovirus serotypes 1, 2, and 3. *J. Virol.* **22**:726-733.
- Risler, J. L., M. O. Delorme, H. Delacroix, and A. Henaut. 1988. Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *J. Mol. Biol.* **204**:1019-1029.
- Rose, G. D., A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* **229**:834-838.
- Rosen, L. 1960. Serologic grouping of reovirus by hemagglutination inhibition. *Am. J. Hyg.* **71**:242-249.

37. **Rudall, K. M.** 1952. The proteins of the mammalian epidermis. *Adv. Protein Chem.* **7**:253–290.
38. **Signas, C., G. Akusjarvi, and U. Pettersson.** 1985. Adenovirus type 3 fiber polypeptide gene: implications for the structure of the fiber protein. *J. Virol.* **53**:672–678.
39. **Spriggs, D. R., R. T. Bronson, and B. N. Fields.** 1983. Hemagglutinin variants of reovirus type 3 have altered central nervous system tropism. *Science* **220**:505–507.
40. **Steinert, P. M., A. C. Steven, and D. R. Roop.** 1985. The molecular biology of intermediate filaments. *Cell* **42**:411–419.
41. **Steven, A. C., B. L. Trus, J. V. Maizel, M. Unser, D. A. D. Perry, 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.
42. **Tardieu, M., and H. L. Weiner.** 1982. Viral receptors on isolated murine and human ependymal cells. *Science* **215**:419–421.
43. **Tyler, K. L., D. A. McPhee, and B. N. Fields.** 1986. Distinct pathways of viral spread determined by reovirus S1 gene segment. *Science* **233**:770–774.
44. **Valentine, R. C., and H. G. Pereira.** 1965. Antigens and structure of adenovirus. *J. Mol. Biol.* **13**:13–20.
45. **Verdin, E. M., G. L. King, and E. Maratos-Flier.** 1989. Characterization of a common high-affinity receptor for reovirus serotypes 1 and 3 on endothelial cells. *J. Virol.* **63**:1318–1325.
46. **Weiner, H. L., K. A. Ault, and B. N. Fields.** 1980. Interaction of reovirus with cell surface receptors. I. Murine and human lymphocytes have a receptor for the hemagglutinin of reovirus type 3. *J. Immunol.* **124**:2143–2148.
47. **Weiner, H. L., D. Drayna, D. R. Averill, Jr., and B. N. Fields.** 1977. Molecular basis of reovirus virulence: role of the S1 gene. *Proc. Natl. Acad. Sci. USA* **74**:5744–5748.
48. **Weiner, H. L., and B. N. Fields.** 1977. Neutralization of reovirus: the gene responsible for the neutralization antigen. *J. Exp. Med.* **146**:1305–1310.
49. **Weiner, H. L., R. F. Ramig, T. A. Mustoe, and B. N. Fields.** 1978. Identification of the gene coding for the hemagglutinin of reovirus. *Virology* **86**:581–584.
50. **Williams, W. V., H. R. Guy, D. H. Rubin, F. Robey, J. N. Myers, T. Kieber-Emmons, D. B. Weiner, and M. L. Greene.** 1988. Sequences of the cell-attachment sites of reovirus type 3 and its anti-idiotypic/antireceptor antibody: modeling of their three-dimensional structures. *Proc. Natl. Acad. Sci. USA* **85**:6488–6492.
51. **Wolf, J. L., R. S. Kauffman, R. Finberg, R. Dambrauskas, B. N. Fields, and J. S. Trier.** 1983. Determinants of reovirus interaction with the intestinal M cells and absorptive cells of murine intestine. *Gastroenterology* **85**:291–300.
52. **Yeung, M. C., D. Lim, R. Duncan, M. S. Shahrabadi, L. W. Cashdollar, and P. W. K. Lee.** 1989. The cell attachment proteins of type 1 and type 3 reovirus are differentially susceptible to trypsin and chymotrypsin. *Virology* **170**:62–70.