Sequences of the S1 genes of the three serotypes of reovirus

(cloning/homology/evolution/genetic relatedness/reading frames)

L. WILLIAM CASHDOLLAR*, RICHARD A. CHMELO, JON R. WIENER, AND WOLFGANG K. JOKLIK[†]

Department of Microbiology and Immunology, Duke University Medical Center, Durham, NC 27710

Contributed by Wolfgang K. Joklik, August 20, 1984

ABSTRACT The S1 genes of the three serotypes of reovirus have been cloned and sequenced. The S1 genes encode protein σ 1, the protein against which serotype-specific neutralizing antibodies are directed; it is also the reovirus hemagglutinin and cell-attachment protein and is a major determinant of host range/tissue specificity and of the nature of the interaction of reovirus with cells of the immune system. The S1 genes of serotypes 1, 2, and 3 are 1458, 1442, and 1416 nucleotides long, respectively. They possess untranslated regions 13, 13, and 12 nucleotides long at their 5' termini and 188, 229, and 36 nucleotides long at their 3' termini. They possess two open reading frames. The first starts with a "weak" initiation codon and extends for 418, 399, and 455 codons, respectively; this is the size expected for the σ 1 proteins. The other reading frame starts at a "strong" initiation codon about 70 residues down-stream from the 5' terminus but extends for only about 120 codons, being terminated by 3 in-phase termination codons in all three genes. The proteins encoded by these short open reading frames are quite basic. The serotype 1 and 2 SI genes are much more closely related to each other (28% homology) than to the serotype 3 S1 gene (5% and 9% homology, respectively). These figures are based on direct homology calculations, adjusted for 25% random coincidence. Serologic evidence and hydrophobicity profiles agree that the σ 1 proteins of serotypes 1 and 2 are much more closely related to each other (about 40% homology) than to that of serotype 3 (only about 20% homology). The fact that the serotype 1 and 2 SI genes are much more closely related to each other than to the serotype 3 SI gene is remarkable since for all other nine reovirus genes the serotype 1 and 3 genes are much more closely related to each other than to the serotype 2 gene. Mechanisms that may effect this remarkable evolutionary pattern are discussed.

There are three serotypes of reovirus. As judged by the ability of their double-stranded RNAs to hybridize with each other under standardized conditions, serotypes 1 and 3 are closely related (about 70%), whereas serotype 2 is related to serotypes 1 and 3 only to the extent of about 10% (1). The individual genes of the three serotypes vary in their relatedness from those that are most closely related, the 3 L genes. to the gene that has diverged to the greatest extent during evolution—namely, the SI gene, which encodes protein $\sigma 1$ (1). Protein σ 1 is not only the most type-specific of all reovirus proteins but is also functionally of great importance. It is the protein against which the major neutralizing antibodies are formed (2); it is the hemagglutinin (3); it is the viral cellattachment protein (4); it is the major determinant of the nature of the interaction of reoviruses with cells, including cells of the immune system (5-7); and it also appears to be the reovirus protein that switches off host cell DNA synthesis (8)

Remarkably, protein $\sigma 1$ is present in reovirus particles to the extent of only 24 molecules that are located pairwise at

12 positions on the reovirus surface where the core projections or spikes protrude through the outer capsid shell (4). Not only is protein σ 1 only a minor reovirus particle component but it is also synthesized in infected cells in only small amounts; its mRNA is one of the most inefficiently translated reovirus mRNAs [>50 times less efficiently than the most efficiently translated reovirus mRNA—namely, s4 RNA (unpublished results)].

We have devised a technique for cloning reovirus genes into pBR322 (9). We present here the sequences of the SI genes of all three reovirus serotypes, together with the deduced sequences of the σ 1 proteins that they encode.

MATERIALS AND METHODS

The method used for growing the Lang strain of reovirus serotype 1, the D5/Jones strain of serotype 2, and the Dearing strain of serotype 3 was that described by Smith *et al.* (10). The procedures used for extracting reovirus genomic RNA, transcribing it into full-length cDNA, and cloning the doublestranded cDNA into pBR322 have been described (9, 11). Total genomic RNA of each serotype was used as the template for preparing cDNA. Before cloning, the populations of single-stranded cDNA molecules were enriched for fulllength *S1* gene transcripts by electrophoresis in alkaline agarose gels (11). The cloned *S1* genes were sequenced by subcloning various restriction fragments into M13 mp8 and mp9 vectors (New England Biolabs) (12) and using the chainterminator method of Sanger *et al.* (13). All regions of all 3 genes were sequenced in both directions.

RESULTS

Sequences of the 3 S1 Genes. The sequences of the S1 genes of reovirus serotypes 1–3 are presented in Fig. 1. The 3 genes are 1458, 1442, and 1416 base pairs (bp) long, respectively. The first initiation codons start at positions 14, 14, and 13, respectively, and are followed by open reading frames 1254, 1197, and 1365 bp long, capable of encoding proteins 418, 399, and 455 amino acids long, respectively (Table 1). These are the sizes expected of the σ 1 proteins, based on electrophoretic migration rates in NaDodSO₄/polyacrylamide gels.

There is another initiation codon, in a different reading frame, that starts at residues 75, 66, and 71, respectively, in all 3 *S1* genes (Fig. 1). These are "strong" initiation codons (14), with purines in positions -3 and +4; by contrast, the upstream codons are all "weak" (Table 1). However, the σ 1 proteins are translated from the first set of initiation codons and not from the second, because the latter are followed by reading frames that are only 119, 125, and 120 codons long, respectively. Interestingly, these short reading frames are all terminated by 3 in-phase termination codons; they are clearly highly conserved.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: bp, base pair(s).

^{*}Present address: Department of Microbiology, Medical College of Wisconsin, Milwaukee, WI 53226.

[†]To whom reprint requests should be addressed.

25

1 DESTINGEDED TA DE AT GE DE TE ME DE ME AT AGE AT AT EST AL AL GE GAL AT EST AL CAL AGE AL TA DE GAL AT AL CE AL AL AL CE AL CAL AL CAL AL CE AL CAL AL																																
3 DETAILTION TOOL THE OLD THAT CHE CHA OTH CALL CALL THA COD OT CALL ALL CALL CALL CALL CALL CALL CALL																																91 91
<pre>9 GLA TC GAC GAA TT AGC AAC CTA ATT AAG GAC ATT CT GT GT GT TF ALG AC ATT GT CAT CAT AC TC AC COL GAC GAC GAC CAT CAT CAT CAT CAT CAT CAT CAT CAT C</pre>																																
3 CEC CTT GAA TCA ACC MA CA AT CAA CAC AT CAA ATA CAC TCA ATA CAC TCA ATA CAC TCA ATA CAT ATC CTO COD ATA CAC CAC CAC CAC ATA CAC CAA CAT CAA CAA																																
2 OF DEGA TTA TT TT TTA GEG GT TA GEG ATT GAL TEG GEG GT TA TT AAT AGE GEG ATT GAL TG AAT GEG ATT GAL AGE GAL GAL TT GE AAT GEG AAT GEG ATT GAL TT GEG AGE TT GEG GAL GAL TT GEG AAT GEG AAT GEG AAT GEG ATT GEG AGE ATT GEG ATT ATT ATT ATT ATT ATT ATT ATT ATT AT	3	GGG	CTT	GAA	TCA	AGG	GTC	TCG	GCG	CTC	GAG	AAG	ACG	TCT	CAA	ATA	CAC	TCT	GAT	ACT	ATC	CTC	CGG	ATC	ACC	CAG	GGA	CTC	GAT	GAT	GCA	
2 OF DEGA TTA TT TT TTA GEG GT TA GEG ATT GAL TEG GEG GT TA TT AAT AGE GEG ATT GAL TG AAT GEG ATT GAL AGE GAL GAL TT GE AAT GEG AAT GEG ATT GAL TT GEG AGE TT GEG GAL GAL TT GEG AAT GEG AAT GEG AAT GEG ATT GEG AGE ATT GEG ATT ATT ATT ATT ATT ATT ATT ATT ATT AT																																
A MC AMA GAA ATC ATC GCT GTT GAC CAA ACT CRC GAT GAC THE GTT GCA FCA GTC GAT GAT GTT GCA ATT GCA ATC CCAA GTT GCA ATC GAA ATT GAA ATC GAA ATC GAA ATT ATT																																
2 AT TOT GOT AGO GTA AGO GTA AGO GAL TOG GAT TOG CHT TAT TOG BAG OTG GUT GAL CAA TO TH GOL GAL GAG GAL ATT GOL CAB TOG GAL CAB COL GAL GAG ATT GOL CAB TOG GAL CAB COL GAL GAL GAL TO GAL AGA CAB CAB CAB TOG GAL GAL GAL GAL GAL AGA CAB CAB CAB CAB CAB CAB CAB CAB CAB CA																																
2 AT TOT GOT AGO GTA AGO GTA AGO GAL TOG GAT TOG CHT TAT TOG BAG OTG GUT GAL CAA TO TH GOL GAL GAG GAL ATT GOL CAB TOG GAL CAB COL GAL GAG ATT GOL CAB TOG GAL CAB COL GAL GAL GAL TO GAL AGA CAB CAB CAB TOG GAL GAL GAL GAL GAL AGA CAB CAB CAB CAB CAB CAB CAB CAB CAB CA																																
1 TET ATE GA GEC CTE CAA ACA GTT GE AAT GA CTT GA TEG ACT TEG AGT GTT ACC CAG TT GEF GTT GA GTA GTA CTT GA CAA GA GA CTT 350 1 AGU UTC AGT TAT CTC CAT GAT GA GAA ACT CCA CTT TIG GAG GA ACT GA GTT GA CTT ACA ACT GTT ACA ACA GAG GTT GAG CTA GAG CTT GAG CTA																																361
1 NG LEG AG, TET ANT CTE ACT TOG AGE TOG AGE ACT TEE GAS CON GAT CTT JGE ANG THE GAG AGE AGE AGE AGE AGE AGE AGE AT CE ATT TO GAS CON GAT CTE JGE ANG AGE ATT AGE																																
2 AGG GTA GTA ATT GTA TAT GTA GTA GTA GTA		101							•••				•••	•			•••		••									••		• • •	•••	
2 AGG GTA GTA ATT GTA TAT GTA GTA GTA GTA	1	ACG	GIC	ACG	тст	AAT	CTC	ACT	GGA	CGA	ACA	тсс	ACT	TTG	GAG	GCA	GAT	GTT	GG A	AGC	TTA	CGG	ACA	GAA	CTA	GCA	GCG	CTA	ACA	ACA	CGG	451
The Ack Act GAD GTT ACA AGE TTA GAT GGT CTA ATC AAT AGT GGC CAG ATT TG ATT GGT AGE CTA GAT AGA AGA GTA TGC AAT GTC GAG AGA GTA AGA GTA AGA AGA AAA GTA AAA GCA AGA AGA GTA ATA GAG AGA AGA GTA ATA GAG AGA GTC GAT ATA GGA GTC GAT TGC GTC AGA AGA GTC AAT ATG GTC GTC ATT AGA GTC GAG AGA GTC ATT AGA GTC AAAT GTC GAG AGA GTC AAT ATG GAG AGA GTC ATA ATG GAG AGA GTA ATG GAG AGA GTA ATG GAG ATA AGT GAG AGA GTC ATA ATG GAG GTC AGT AGA GTC GAG ATA ATG GTC AGA GTA AAT GAG GTC AGT AGA GTC AGA AGA GTC GAT ATA ATG GAG GTA ATG GAG ATT AGA TGC GTC AGT AGA GTC AGA AGA GTC GAT ATA ATG GAG GAA AGA GTC AGA AGA GTC GAG ATT AAT GAT GAG AGA AGA GTC AGA AGA GTC AGA AGA GTC GAG ATT AAT GAT GAG AGA AGA GTC AGA AGA GTC GGA AGA GTC GTC AGA AGA GTC AGA AGA GTC GTC GTC AGA AGA AGA GTT AGA GTC GTC GTC AGA AGA AGA GTT AGA GTC GTC GTC AGA AGA AGA GTT GTC GTC GTC AGA AGA AGA GTT GTC GTC AGA AGA AGA GTT GTC GTC AGA AGA AGA ATT AGA TGC GGA AGA GGG GTG AGA AGA GTT GTC GTC AGA AGA AGA GTT GTC GTC AGA AGA AGA GTT GTC GTA AGA AGA GTT GTC GTA GTC AGA AGA AGA GTT GTC GTA GTC AGA AGA AGA GTT GTC GTA GTA AGA ATT AGA GTC GGA AGA AGA GTT GTC GTA GTA GGA AGA ATT AGA GTC GG	2	ACA	GTC	A CG	GAT	AAT	CTC	СТТ	G AG	C.J.A	GCA	TCA	AGA	CTG	GAA	ACT	GAŃ	GTA	TCA	G CC	ATT	ACT	AAT	GAC	CTT	GGA	TCA	TTG	AAT	A CG	AGG	451
1 OFE AGG ACT CAA TIG AGG CAT CAC AT GAC GAC THE GAT GAT GAT GAT GAC ACT GAT GAG ACT GAT ACT GAG ACT GAT GAG GAT GCG TO GAG ACT GAT GAG GAT TA GAG GAT ACT GAT GAG GAT ACT GAG GAT GAT GAG GAT GAG GAT GAG GAT GAG GAG	3	GCA	GAG	CTA	CGC	GTT	GAT	CAC	GAC	AAT	CTC	GTT	GCG	AGA	GTG	GAT	ACT	GCA	GAA	CGT	AAC	ATT	GGA	TCA	TIG	ACC	ACT	GAG	CTA	TCA	ACT	450
1 OFE AGG ACT CAA TIG AGG CAT CAC AT GAC GAC THE GAT GAT GAT GAT GAC ACT GAT GAG ACT GAT ACT GAG ACT GAT GAG GAT GCG TO GAG ACT GAT GAG GAT TA GAG GAT ACT GAT GAG GAT ACT GAG GAT GAT GAG GAT GAG GAT GAG GAT GAG GAG					~ • •			400	***		COT				ACT.				T.CC				C'PA	TTC				TCC				5.4.1
A GET CAA GET CHA GET AGE AGE GET UDA GEU GAA CTE CAET AME GET ANT GTE GET AAC GET ATT GTA GET ANT GTE GAA ATT GTE GAA TAT AME GET ANT GTE GET AME GET ANT GET GAA CTE CAE ATT AME GET ANT GTE GET AME GET ATT GTE GAA CTE CHA TAT GTE GAA TAT CHE GAA ATT GTE CHA CAE TAT AME GET ANT GTE GET AME GET ATT GTE GAA ATT GTE GAA TAT AME GET ANT GTE GET ANT GAA TAT GTE GEA ATT AME GET ANT AME GET GAA TAT CHA GEG ANT AMT GTE GEA ATT GTE CAE TAT AME GET ANT GAA TAT GTE GAA TAT AMT GTE GAA ATT GTE GAA TAT AMT GTE AME ATT GTE GAA CAE ATT AMT GTE GAA TAT GTE GAA TAT GTE GAA ATT GTE GAA GTE AAA ATT GTE GAA GTE GAA ATT GTE GAA ATT GTE GAA GTE AAA ATT AAA T GTE GAA GTE GAA GTE GAA ATT AAA T GTE GAA GTE GAA ATT AAA T GT GTE GAA ATT AAA T GT AAT GAA GTE AAA ATT AAA T GT GTE GAA ATT GTE GAA GTE GAA ATT AAA T GT GTE GAA ATT AAA T GT GTE GAA ATT AAA AGE AT GTE GAA ATT AAA AGE AT GT GTE GAA ATT GGE GAG GGA ATT AGA T GAA AGT TAA AGA AGE T AAA C AAA AGA ATA AGA ATA GGET GAA AGT TAA GAA AGT TAA AGA AGT GTA AGA ATA AGA AG	2	GTG	ACG	ACT	GAA	TTG	AAC	GAT	GIC	CGC	CAA	ACT	ATT	GCT	GCG	ATA	GAC	ACG	CGT	CTC	ACG	ACA	CTG	GAG	ACC	GAT	GCC	GTG	ACG	TCG	GTT	
2 COT CAA GGC CIT CAG AAG ACT GGG ATT AAG GT ATT ALG GIT ATT GIG GIG ACG ACG CIT GAG GGG ATT ATT GIG GAG ATT AAT GAT GAG GGG ATT AAT ATT GIG GGG ATT AAT GAT CIT GIG CAG GGA ATT GAT GAG 3 CIC ICA ATC GGT ATT ACC GIT AGG ACG ATG GGG ATT AAT GAT GGA GIT GIG AGG ATT GIG GGG ATT AAT ATT TIT CIC AGG ATA GAT AAT ATT CIC AGG ATT AAT AAT CIC AGG ATG GGA ATT GAG AGG AG	3	CTG	ACG	TTA	CGA	GTA	ACA	τcc	ATA	CAA	G CG	GAT	TTC	GAA	TCT	AGG	ATA	FCC	ACG	TTA	GAG	CGC	ACG	G CC	GTC	ACT	AGC	G CG	GGA	G CT	ccc	540
3 CFC FCA ATC CGT FAA TAC CGF ATG ACC ATG GCA TTA AAT GAT GCA CTC: ACG TTG TCA GCG ATA ATT GT CGC ATC CGA ATT GTC CGA ATT GTC GCA ATG GCA ATG GCA ATT GTC GTC GAA ATG GTA GTC ATA ATT TAT TAT TAT TAT TAT TAT GTA GTC AAT TAT CAT ATA ATT ATT TAT TAT GTA GTC AAT ATT ATA ATT ATT ATG GTA GTA ATT ATA ATT ATT																																
1 TET AAT CAA TIG CAG GTE GAC CIT ICG GGG CAA TCA AAA GGG GTE GGA TIT GTE GGC ACA GGA ATG GTE GT TAG ATT GAT ACT ATT TIT AT ACT GAT GAT ATT ACT AC																																
2 TTA TOG ANG CAG CAG ANA GGG TTU GGA TTU ATA CAC ANT GUA ATG GTA GTG ANA TTU GAA ATC CAG ATT GUT ANT ANT AND AND AGT CAG CAG ATT AND ANT CAG ATT CAG ATT AGG ATT CGA ATT GUT ANT ANT AGG TTU CAG ATT GUT ATT ANT AGG TTU CAG ATT GUT ANT ANT AGG TTU CAG ATT GUT ATT ANT AGG TTU CAG ATT GUT ATT ANT AGG TTU CAG ATT GUT ATT	5				••••										•		••••															
3 AGG GGT CTG AAT AFT CAA AAT GGT GGA CTT CAG TIT GGA TIT AAT ACT GAT CAA ITG CAG ATA GTT AAT AAT AAT CTG ACT CTG AAG ACG 720 1 FIT GCT TAG AAT AGT AAT GGA GAT ACA TIN GTG AGT CAA ATG AAT GAAT GAA TUU GCA TOG GGG GGA CAA GCA CTG GTA GGG GGA ACA GGT CGG GGA CAA GGT CGC TTG GAG GTG GTA CGA CTG GTA GGG GAA ATG ATT AGT GUT CTG CGG GGG GGA ACA GGT CGC CTG GAG GCA CTG TG GAA CGA GGT GTG GTA GGG GAA ATG ATA AGT GUT CTG CGG GGG GGA ACA GGT CGC CTG GAG GTG GGA GGA	1	TCT	AAT	CAA	TTG	CAG	стс	GAC	CTT	TCG	GGG	CAA	TCA	***	GGG	GTC	GGA	TTT	GTC	GGC	ACA	GGA	ATG	GTG	GTT	AAG	ATT	GAT	ACT	AAT	TAT	721
1 FIT GET TAC ANT ACT ANT GOA GAG ATT ACA TTU GTU AGT CAA ATG AAT GAA TTU GCA TOG CUC GTA TUA ACA CTU GAA TCA GUU AAA ATG 2 AAC ATA ACT CTU AAC AAC ATA ACT GUT CTU CUC GUT GUU GUU GUU GUU GUU GUU GUU GUU AAA ATG 3 ACT GTU TTT GAT TET ATC AAC GAA CATA ACT GUT CUC GUU GUU AAG GTT AC GTU GUU TAT GAT ACG ACG GAT TTT ACA ATG ATG 3 ACT GTU TTT GAT TET ACT ACC GTA CGG GAA ACT GUU GUU AGU GUU GUU GUU GUU GUU GUU GUU																																
2 AAC ATA ACT CTG AAC AAC ATA ACT GUT CTG CGG GGG CGA ACA GGG TCC CTG GAG GCA TCT GT GT GTG GCA GTG GTA GT ATC ATC GTG GCA CGC CTT SII ACT GTG TTT GAT TCT ATC AAC TCA AGG ATA GGG GCA ACT GAG CAA ACA GGT TCC CTG GAG GCA GTG ACT GCC TTG AGA TTA AAC AGT ACC ACT ACT 1 GAT TCA GTT TTA CAT CCT CCA TTA ACC GTA GGC GAA GGC AGG GGC GTA GGT ACC TC3 AGC TTT GTT ACT ACT ACC ACC ACG CTG AGA TTA AAC AGT ACG 2 GTA ACA CTT TA CCT CCA ATA ACC GTA GGC GAA GGC AGG GGC GTA GGT AGC TC3 GGC GTT GTG GTA ACT ACT ACT ACT ACT ACA ACC ACA CTG AGA 3 ACC ATA CAC TCT ACT GT ACT ACT GGT TG ACT ACT GC GT GTG ATG ACT ACG AGG GGT GTG GTA GGT ACT GTG GTT ACT ACT ACT ACT ACT ACT ACT ACT ACT A	3	ACG	GGT	CTG	AAI	ATT	CAA	~~1	651	GGA	CII	CAG		CG A	m	~~1	ACI	GAL	LAA	110	CAG	A 1 A	611	~~1	AAL	AAC	116	ACI	CIU	AAG	ALG	720
2 AAC ATA ACT CTG AAC AAC ATA ACT GUT CTG CCG GCG CGA ACA GCG TCC CTG GAG GCA CTG CTT GTG TAT GAT GTG GCA CCG CTT SII ACT GTG TTT GAT TCT ATC AAC TCA AGG ATA GGG GCA ACT GAG GAA ACA GGT TCC CTG GAG GCA CTG GTA GCA CCG TTG AGA TTA AAC AGT ACC ACT GG 1 GAT TCA GTT TTA CAT CCT CCA TTA ACC GTA GGC GAA GGG AGG GGC GTA GCT ACC CTG AGC GTA GTTA CTA CTG ACT ACC ACC ACT ACC ACT GG 2 GTA CAC TCT ACT GT ACT ACT GCA CTA CGC CTA CTG GCT CTG ATG TAC CAC CTG AGC CTG AGC TTA GTA ACT ACT ACT ACT ACT ACT ACT ACT ACT A	,		COT	T • C								TT	67 0	107	~	170		~ • •	TT		Tee	000		P (* 4		CT.	C • •	TC 4				211
1 GAT TGA GTT TTA CCT CGA TTA ACC GTA CGC GAG GGC AGC GGC GTA CGT ACC CT3 AGC TTT GGT TAT GAT ACG AGC GAT TTT ACA ATC ATC 2 GTG ATA CAG TCT ACT GGT ACC AGC GGT CT4 CGT CTC ATY TAC GAG GGT GT3 GAC TTC GTG GTT ACT AACA ACC GTT CT CACA ATT AT 3 ACC AAC GTG CTG GAT ATG CTA ATA GAC AGT TCA CAC CTT GAA ATT AAT TAT TCT AST GGA CG CTA ACT GTT AGA TGG ACT CCC CCC ATT TG 1 AAC TCC GTA CTG CTG GAT ATG CTA ATA GAC AGT TCA CAC TTT GAA ATT AAT ATT CAT AST GGA CGA GAC ACA CGA AAT AAT AGA ATG CGC CTA AGC 1 AAC TCC GTA CTG CTG GAT ATG CTA ATA GAC AGT TCA CAC TTT GAA ATT AAT ATT CAT GGA CGC GAC GAC ACA CGA AAT AAT AGA ATG CGC CTG AGA GTG CGC 1 AAC TCC GTA CTG CTG CTG ATA CGC TAC AGG TTC AGA CTT GGA ACT AGA TAG AGT GGA CTA ACT AGA ATA AGA ATT AGC GGC ATT AG 1 AAC TCC GTA CGG CAA CTG CGC CACA TTC AGG ACT TGG ACG GGA CAA TTA AGG TTA AGG CTC CGA CAC GGA ATA GAA ATA AGA ATT AGC GGC ATT AG 1 AGC ACG AGT CGT TTT GGC ATG GGC AGG GGT ATT TGG AGG GGA CAA TTG GAA TTAT CAG CGA CAA TTG TGG AATA GGA ATT GTC CGC ATT 3 AGC AAT GGT CAG GT CAA GGT GAT ATG GAA CTG GGA AGG GGA CAA TTG GAA TAT CAG CGA CAA TGG CAG AGG ATG GTC AGC GAT ATT GGA GGC ATT AGC GGC GT AAT TGG GAG GGA AAT TAT CAC AGG GTA ATT TTA AGG TTT AGG CAG AGT AGG ATT GGC AGT GTT GGA ATG GGC ATT GGA AGG GTA ATT TGG AGG GTA ATT TGG AGG GTA ATT TGG AGG GT AAT TGG AGG GTA ATT TGG AGG GTA AGC GGA ATT TTG ATG GAG GTA ATA TAT GGA GG TTA GGA ATT GGA AGG GTA ATA TTG AGG GGA GAA TTG GGA GTA ATT TGG AGG GTA AGC GGA GAA TTT TG ATG AGG AGT AGA GGA ATT AGA ATA AT	2	AAC	ATA	ACT	CTG	AAC	AAC	AAC	ATA	AGT	GGT	CTG	CCG	GCG	CGA	ACA	GGT	TCC	CTC	GAG	GCA	101	CGT	A'IC	GAT	G TG	GTA	G CG	CCA	CCG	CTT	811
2 GTG ATA CAG TET ACT GGT ACC ACT GGG CTA CTG GGT CTC ATF TAC GAG GET GTG GAC TTC GTG GT ACT AAC AAC GTT CTC ACA GTG ACA GTG GAC AGG GTA ATG GAG GTG GTG GAC ATG GTA ATA GAG AGT TGA AGA GTT GAG AGT GTG AGA GTG GAG AGT GTG AGA GTG GTG	3	ACT	GTG	TTT	GAT	TCT	ATC	AAC	TCA	AGG	A T A	GGC	GCA	ACT	GAG	CAA	AGT	TAC	GTG	GCG	TCG	GCA	GTG	ACT	ccc	TTG	AG A	ŤΤ A	AAC	AG T	AG C	810
2 GTG ATA CAG TET ACT GGT ACC ACT GGG CTA CTG GGT CTC ATF TAC GAG GET GTG GAC TTC GTG GT ACT AAC AAC GTT CTC ACA GTG ACA GTG GAC AGG GTA ATG GAG GTG GTG GAC ATG GTA ATA GAG AGT TGA AGA GTT GAG AGT GTG AGA GTG GAG AGT GTG AGA GTG GTG																																
3 AGG AAG GTG GTG GTG GAT ATG CTA ATA GAC AGT TCA ACA CTT GAA ATT AAT TCI AGT GGA CAG CTA ACT GTT AGA TCG ACA TCC CCG AAT TTG 930 1 AAC TCC GTA CTG TCG TTA CGG TCA AGG TTG GACT GTG CTT CCG ACA TAC AGG TAC GCG CTC GTG GAG GTC GAC ACT GTA AAT TAC GTG ATT AGA GTG CAG GTG GT 2 AAT CCG ATA GCT CAG CCA ACG TTA AGC GTT CCT CTG GAG TTG AAT AGT GCT CTG AAC TCA GTA ACT CAG AAT AAT AGA GTG CCG ATT AGG AGT GTG ATA GCG ATT GGA ATT AGG GTT CCT TTT 3 AGG TAT CCG ATA GCT GTG AGG GGT ACT TGG AAG AGG AGG AGA ATG GCA AAT TAT AGG TTT AGG CAG AGG ATG GGA AAT GTC ACT TTG 1081 3 AGG TAT CCG ATA GCT GAG CGC AGG GGT ACT TGG AAG GGA GAA TTG GAA TAT CAG CAC CGA CAA ITG AGA AGT AGG AAT GTC ACT TGG 11081 3 TCT GGC AAT TGG CAA TGG GAA TAT CAC ACG CCG AAC TCC GAA ATT GTA GAT GTA CAT ATA AGG GTT AGA ATT AGA ATT GGA GGA AAT TTG AG AGG ATT TGG ATT AGA AGT TCG AGT AGG GGT AAT TGG AGG GGA AAC TCC GAC ATT TTT ATT GTA GAT GAT GGA GTA GAG CAT ATA CAG ATT TGG ATT AGA ATT TGA AGG AGT AGG CGG GT AAT TGG AGG GGA AAT TTG GA GTA GAG GTA AGT GGG TTT AAG AGG GTG AAC TCC GAC ATT TTT ATT GTA GAT GAA ATT AGA ATT GGA GGG GAA ATT TGG ATT AGA AGT CCG AGA ATT GGA GGG GAA ATT GGA GGG GGA AAC TCG GAG GAC ACT GGA GAG CAC CCG GGA GAA CT CG GGG GAA ATT GGA GGA GAA CTT GTT AGG GTT ATA GGA GGA GAA CTT GGA GGA GAA																																
2 AAT CGA TCG GTC ACG CCA ACA TTC AAG TTT CCT CTG GAC TTG AAT ACT GCT GAT AAC TCA GTG AGG ATT CAT AGA AAT TAC GGC ATT ACA 3 AGG TAT CCG ATA GCT GAT GTT AGG GGC GUT ATC GGA ATG AGT CCA AAT TAT AGG TTT AGG CAG AGC ATG TCG ATA GGA AAT GTC TCC TAT 991 1 GCA CAT CUT TTT GGC ATG CGC ACG GGT ACT TGG ACG GGA CAA TTG CAA ATT TAT AGG TTT AGG CAC CAA FTG AGG ATA GGA AAT GTC ACT TTG 2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC ACG CCG AGT TTG CGT TCG CAA TAT CAG CAC GAT GGT AAT TTG ATG CGA GAA GAC GAT GTC ACT TTG 2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC ACG CCG AGT TTG CGT TCG CAC GAT GAT CAC ACG GTC AAT ATT GAT CAT CAT ATA TTG ATG CAT GAA GAC GAT 1081 3 TCT GGT AGT GGG CTG AAT TGG AGG GTA CAC GTG TAC TCC GAC ATT TTT ATT GTA GAT GAT TAC ATA CAT ATA TGT CTT CCA GAT GAC GAT 1081 1 AAT TTG ATG AAG GTG GAT GAT TGG TTG GTG TTG AGC TTT TCT CAG ATG ACG ACT AAC TAC CAA ATA ATG GGA GAA TTT GTG ATT AAT 2 TCG CTC ATT TTG AGT TTA CT CCG TTT TGG AGG GAG CAC CCG GGC GAT CTT AGC GTC AGG AAA GTT TGT ATT GAA CTT CGT AAC TGG TTT GCT CCC 1 1170 1 TTT GTG TCT GGA TAG CGG GGA GAT CTA TGG TGG AGC GGG GAT ACT GAA CCA ACG AGG AAAT TAT CAA CGA TTG CAA GGA CAC TGG GAG CAC TGG CGG CGC TCC 1 1170 1 TTT GTG TCT GGG TTA TCT TCT GGA AGG CAG ACG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TTG CGG CGG TCC 1 1170 1 TTT GTG TCT GGG TTA TCT TCT GGA AGG CAG ACG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TTG CGG CGG TCC 1 1170 1 TTT CAT AAT GAC GTG GAC TAC CGA ACC GAG ACT GAA CTA GGA GTA GGA ATTAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTAAGGACTAAGGACTAAGGACTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGAGATTAAGAGGATCAAGGAACTAAGGGACTAACCAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGGAATGCAC																																
2 AAT CGA TCG GTC ACG CCA ACA TTC AAG TTT CCT CTG GAC TTG AAT ACT GCT GAT AAC TCA GTG AGG ATT CAT AGA AAT TAC GGC ATT ACA 3 AGG TAT CCG ATA GCT GAT GTT AGG GGC GUT ATC GGA ATG AGT CCA AAT TAT AGG TTT AGG CAG AGC ATG TCG ATA GGA AAT GTC TCC TAT 991 1 GCA CAT CUT TTT GGC ATG CGC ACG GGT ACT TGG ACG GGA CAA TTG CAA ATT TAT AGG TTT AGG CAC CAA FTG AGG ATA GGA AAT GTC ACT TTG 2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC ACG CCG AGT TTG CGT TCG CAA TAT CAG CAC GAT GGT AAT TTG ATG CGA GAA GAC GAT GTC ACT TTG 2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC ACG CCG AGT TTG CGT TCG CAC GAT GAT CAC ACG GTC AAT ATT GAT CAT CAT ATA TTG ATG CAT GAA GAC GAT 1081 3 TCT GGT AGT GGG CTG AAT TGG AGG GTA CAC GTG TAC TCC GAC ATT TTT ATT GTA GAT GAT TAC ATA CAT ATA TGT CTT CCA GAT GAC GAT 1081 1 AAT TTG ATG AAG GTG GAT GAT TGG TTG GTG TTG AGC TTT TCT CAG ATG ACG ACT AAC TAC CAA ATA ATG GGA GAA TTT GTG ATT AAT 2 TCG CTC ATT TTG AGT TTA CT CCG TTT TGG AGG GAG CAC CCG GGC GAT CTT AGC GTC AGG AAA GTT TGT ATT GAA CTT CGT AAC TGG TTT GCT CCC 1 1170 1 TTT GTG TCT GGA TAG CGG GGA GAT CTA TGG TGG AGC GGG GAT ACT GAA CCA ACG AGG AAAT TAT CAA CGA TTG CAA GGA CAC TGG GAG CAC TGG CGG CGC TCC 1 1170 1 TTT GTG TCT GGG TTA TCT TCT GGA AGG CAG ACG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TTG CGG CGG TCC 1 1170 1 TTT GTG TCT GGG TTA TCT TCT GGA AGG CAG ACG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TTG CGG CGG TCC 1 1170 1 TTT CAT AAT GAC GTG GAC TAC CGA ACC GAG ACT GAA CTA GGA GTA GGA ATTAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTAAGGACTAAGGACTAAGGACTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGACTTAAGGAGATTAAGAGGATCAAGGAACTAAGGGACTAACCAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGAATGCACGGAATGCAC																																
3 AGG TAT CCG ATA GCT GAT GTT AGC GGC GGT ATC GGA ATG GAT AGT CCA AAT TAT AGG TIT AGG CAG AGC AGG AGG ATA GGA ATT GTC TCC TAT 990 1 GCA CAT CGT TIT GGC ATG CGC ACG GGT ACT TGG ACG GGA CAA TTG CAA ATT GA CAG CGA CAA TTG CAA CTG CGA CAA TGG CAA TGG CAA TGG CAA GG CGA ACT TGG AAG GGA CAA TTG CAA TAT CAG CAG GGG AAAT TGG CAG GGA GAA TAT CAC ACG CGG AGT TGG CAG GGT GGA ATT GGA GGA GAA TGG CAA GTG GAG GGA GAA TGG CAA GGA GAG ATT GGA GGA GAA TGG CAA GGA GGA GAG GGA GAA TGG CGA AGG GGA AAT TGG AGG GTA AAG GGG GTA AAT TGG AGG GTA AAT TGG AGG GTA AAC TCC GAC ATT TTT ATT GTA GAT GAT TAC ATA ATA TGT CTT CCA GGT TTT GAC 1081 1 AAT TTG GGT GGG CG AAT TGG AGG GTA ACT GG AGG CGA ACT TCC GAC ATT TTT ATT GTA GAT GAT TAC ATA ATA TGT CTT CCA GGT TTT GAC 1081 1 AAT TTG AGG AGG GTG AAT TGG TGG GTG AGG CTG AGG CGC GAC TTT TAT TTT ATT GTA GAT GAT TAC ATA ATA TGT CTT CCA GGT TTT GAC 1081 1 AAT TTG AGG GTG GAT GAT TGG TTG AGC TTT GG AGG CTT AGG ATT ACC ACT AAC TAC ATA ATA GGG AGA ATTT GTA TGG ATT AGG ATT AGG GTT AGG ATT AGG ATT AGG ATT AGG AGG																																
2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC AGG CGG AGT TTG CGT TGG CAT GGT GGC AGT GGC AGT GGT GAA TTG GT GGG CAA TAT TGG AGG GAA TAT CAC AGG CGG AGT TTG CGT TGG CGC GTC AGG GTT AAT TTG ATG GGG GTA GAC GAT TGC AGG AGT AGC GAC TAT TTT ATT GTA GAT GAT GAT AAT ATA TGT CTT CCA GCT TTT GAC 1081 3 TCT GGT AGT GGG CTG AAT TGG AGG GTA CAG GTG TAC AGG GCG AGT TTT ATT GTA GAT GAT GAT AAT ATA TGT CTT CCA GCT TTT GAC 1081 1 AAT TTG ATG AAG GTG GAT GAT TGG TTG AGG GTG TAC AGG GTG TTT CT CAG ATG ACG ACT AAC TAC ATA ATG GCA GAT GGG AAA TTT GTA TAT AAT 1171 2 TGG CTC ATT TTG AAG GTG GAT GAT TGG TTG AGC GTT TAG AGG GGC GAT CTT AGC ACG AAC TAC TGA AGT AAT GAA CTT CT AAC GGA ACT GGA TTG TTA ACT GGA GGA GAT CTA TGG TTT AAC AGG GTG GGC AAC TGG CAG ACT GGA GAT CTA TGG TTG AAC TAT GGT GGA GAT AT GG GTG GGG GAT ACT AAC TAC TAA ACG GGA GTA CTA AGG ACT GGA ACT TAG GAG CAC TGG GGC GAC TAG GAG ACT GGA ACT AAC TAC TAA GGG TGG AGC ACT GGA GAG AAC TAC TAA GGA GAA CTA CTAA GGG TGG AGC ACT GGA GGA ACT GAA ACT ATA GGG TGG GGC GGT GGC GCT CAG TAA TAT AGA AAT CTG 1261 1 TTT GTG TCT GGG TAA TCT TCT GGA TGG CAG GGG GGG GAF ACT GAA CGA TGG TAG AGC ACT GAA GCA TTG GAG CAC TGG AGC ACT GAA CAG AAC TAC TAA GGG TGG GGC GGC GGC GGC GGC GGC GGC GGC																																
2 CTT GGG CAA TGG TCA GGT CAA TTG GAA TAT CAC AGG CGG AGT TTG CGT TGG CAT GGT GGC AGT GGC AGT GGT GAA TTG GT GGG CAA TAT TGG AGG GAA TAT CAC AGG CGG AGT TTG CGT TGG CGC GTC AGG GTT AAT TTG ATG GGG GTA GAC GAT TGC AGG AGT AGC GAC TAT TTT ATT GTA GAT GAT GAT AAT ATA TGT CTT CCA GCT TTT GAC 1081 3 TCT GGT AGT GGG CTG AAT TGG AGG GTA CAG GTG TAC AGG GCG AGT TTT ATT GTA GAT GAT GAT AAT ATA TGT CTT CCA GCT TTT GAC 1081 1 AAT TTG ATG AAG GTG GAT GAT TGG TTG AGG GTG TAC AGG GTG TTT CT CAG ATG ACG ACT AAC TAC ATA ATG GCA GAT GGG AAA TTT GTA TAT AAT 1171 2 TGG CTC ATT TTG AAG GTG GAT GAT TGG TTG AGC GTT TAG AGG GGC GAT CTT AGC ACG AAC TAC TGA AGT AAT GAA CTT CT AAC GGA ACT GGA TTG TTA ACT GGA GGA GAT CTA TGG TTT AAC AGG GTG GGC AAC TGG CAG ACT GGA GAT CTA TGG TTG AAC TAT GGT GGA GAT AT GG GTG GGG GAT ACT AAC TAC TAA ACG GGA GTA CTA AGG ACT GGA ACT TAG GAG CAC TGG GGC GAC TAG GAG ACT GGA ACT AAC TAC TAA GGG TGG AGC ACT GGA GAG AAC TAC TAA GGA GAA CTA CTAA GGG TGG AGC ACT GGA GGA ACT GAA ACT ATA GGG TGG GGC GGT GGC GCT CAG TAA TAT AGA AAT CTG 1261 1 TTT GTG TCT GGG TAA TCT TCT GGA TGG CAG GGG GGG GAF ACT GAA CGA TGG TAG AGC ACT GAA GCA TTG GAG CAC TGG AGC ACT GAA CAG AAC TAC TAA GGG TGG GGC GGC GGC GGC GGC GGC GGC GGC																																
3 TET GGT AGT GGG CTG AAT TGG AGG GTA CAG GTG AAC TEC GAC ATT TTT ATT GTA GAT GAT TAC ATA CAT ATA TGT CTT CCA GCT TTT GAC 1030 1 AAT TTG ATG AAG GTG GAT GAT TGG TTU GTG TTG AGC TTT TCT GAG ATG ACG ACT AAC TCA ATA ATG GCA GAT GGG AAA TTT GTG ATT AAT 2 TGG CTC ATT TTG AGT TTT ACT CGG TTT ACG CGG TTT GTG AGC GGC GAT CTT AGC GC AGG AAA GTT TGT ATT GAA CTC GT AAC TGG TTT GTG ATT AAT 3 GCT TTC TCT ATA GCT GGG GGA GAT CTA TCG TTG AAC TAT CG TTG AAC TAT GTA ACT GAT AAC AGT ACA GGA GAC ACT GAG CGC GCT 1170 1 TTT GTG TCT GGG TTA TCT TCT GGA TGG CAG AGG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TGC CGC GGT TC 2 AGG CTG GGC GAC TGG GAG TAC CGA GGC AGG AGG GGG GAT ACT GAA CCA TCG TCA ACT ATT GAT CAT TGA CTA CGA CAT TTG COG CGG TCC 2 AGG CTG GGC GAC TGG GAG TAC CGA GGC CTU GAC AAC TAC TAT GAA CCA TCG TCA ACT ATT GAT CAT GAG CAT TTG CAG CAT TTG CGA GGA CAG AAC TAC TAC TAA GGG TTG TGA CGG GGT GGT GGT GGG CGT CAG TAT ATG AG AAT CTG 1 TTT CAT AAT GAC GTG GTC ACA TAT GGA GCA CAG ACT GTA GGT ATA GGG TTG TCG TCG GGT GGT GGT GGC CCT CAG TAT ATG AGA AAT CTG 1 AAT TTC TAAA TAACGGTCAACGCATTGATGCCTTAGGATCATGGGAGTATGGAATGGAACGCATGGACATTAACGAATTAAGAATTATGGTGGCACATAACACGGGCATGACCATATTGTAGCCATGT 1 TAAAACGGTCGAGGGAGTGGAATGCCGGGGAACTAGGAATCATGGGGGAACATATACAGGGGATGGAACATTAACGGACTGGGCGCCGATGCCAATATAGGAGCCATTATGTAGCCATTGT 1 TAAAACGGTCGGGGTCGGAGAGTGGAATGCAAGGAATCAGGGAATCATGGGGGAGCAATTACGAATTAGGAGTGAAGCAATTAGGAGCCGATGGCACCATATAGTGGCCGCGCGATGCCAATGTGAACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACCAATGTGGACGCATGGACAATATACGGGCGCAACCAATATACGGCGCAACACTGA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCTCCGTGGAGGAATGCAAGGGAATCCAAGGAATCAAGGGGGAACCAAGGGGGTCAATCAA																																
1 AAT TTG ATG AAG GTG GAT GAT TGG TTU GTG TTG AGG TTT TGT GAG ATG AGG ACT AAC TCA ATA ATG GCA GAT GGG AAA TTT GTG ATT AAT 1171 2 TGG CTC ATT TTG AGT TTT ACT CGG TTT TGG AGG GGC GAC CTT AGC GGC GAC ATG TA AGC GGA GAT GTT AAC TGG ACT GAC GGG TTT GTG ATG GAC GGG TTT GTG ATG GAC GGG TTG GAC GGG GGC GAC ATG TTA GAC GGA GAC ACT GGA GAC ACT GAC GGC GGC GGC TTG TTA CCA GGA GAC ACT GAC GGC GGC GGC TCC 1171 3 GCT TTC TCT ATA GCT GGA TGG CAG AGG GGG GAT ATT GT AGC GGA CTG TTA CCA GGA GTA CTT ACA GGA GAC ACT GAC GAC GGC GGC TCC 1261 1 TTT GTG TCT GGG TTA TCT TCT GGA TGG CAG ACG GGG GAT ACT GAA CT GAT CG TCG TCG ACT ACT ACG GGA GAC ACT GGA GCC CTU GAC AAC TAC TAC GAA CTA CTG TTG GCGACATTGAGGACGTATGGAATGGA																																
2 TGG CTC ATT TTG AGT TTT AGE CGG TTT TGG AGG AGG GGG GGC GGC GGC GGC GGC AGG AGA GTT TGT ATT GAA CTT GTT AGC TGG TTG GGC GGT TTG TTA GGG TTG TTA AGG GGG GGT GGG TTG TT																																
3 GGT TTC TGT ATA GCT GAC GGT GGA GAT CTA TCG TTG AAC TTT GTT ACC GGA TTG TTA CCA CCG TTA CTT ACA GGA GAC ACT GAG CCC GCT 1170 1 TTT GTC TCT GGG TTA TCT TCT GGA TGG CAG ACG GGG GAF ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CTA CGA CAT TNG CCG CGG TCC 1261 2 AGG CTG GGC GAC TGG GAC TAC CGA GCC CTC GAC AAC TAC TAA CGA ACT ACT GAA CGA ACT ATT GAT CAA TTA CGA CAT TNG CGA CAT TNG CGG CGG TCC 1261 3 TTT CAT AAT GAC GTG GTC ACA TAT GGA GCA CAG ACT GTA GCT ATA GGG TTG TCG TCG GGT GCT CAG TAT ATG AGT AAC AAT CTG 1260 1 AAT TTC TAA AT AACGGTCAACGCATTGATGCGTTAAGGATCAACGGAATGGAATGGAATGGAACGCATGGAGAATTAAGAATTATGGTGGCACATACACGGGCAACCAATATACGAGGGCACATATACGAGGCGAAGCGAATGGAACGCATGGAACGCATGGAGCGATGGAATCAGGAGCGATGGAATCATGGGGGCGCGATGGCACATATACGGCGCGATGGCACATATACGGGCGCATGGACATTATGGTGGCCGATGGCACATATATGGTGG CCC ATG 1377 2 TTAGAATCTTCGGAGTCGCAGGGAGTGGAATGCAAGGAATCAGGAATGAGGAATGGGAAGATAACAGGGGTCAATCAGGAGTGGGACGATATACGGCGCATGGACGATGGAACGGATGGAATGTGGACGGATGGAATCAGGACGCATGGGCGCGATGGCACGATGGACGCGATGGCACGATGGGAGCGATGGAAGGGGTCAATCAGGAGTGGACGTATACGGGCGCGATGGCACATATACGGCGCGATGGCGCGATGGCGCGATGGACGTGGCGCGATGGCCGATGGCCGATGGCCGATGGCCGATGGCCGATGGCCGATGGCGCGATGGCGCGATGGACGATGGAGGGGTCAATCACCAATGTCGGCGCGATGGCGCGATGGCGCGATGGCGCGATGGCGCGATGGCGCGATGGCGCGATGGCGCGATGGACGCGATGGACGGATGGACGGATGGAGGGGTCAATCATGGGGGGCGAACTAGGAGGGGTCAATCATGGGGGGCGAACTAGGAGGGGTCAATCATGGGGGGCGAACTAGGAGGGGTCAATCACGAATGTGGAGGGGCGGATGGACGGATGGACGGATGGGGGTGAATCAGGAGGGGTCAATCATGG GAG TAC CTG GTT GAG GGG GGT GGC TCA ATT ACG GAC TCA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCTCCGGGGTGGCACAGTGATGCAAGGGGGAATCTAGGGGGGAACTCAGGGGGCCAATCATCATCC 1458 2 CTGGGCTGAGGATCCGGGTGCTCCACGGCGAATGGCGGAATGGGGGGGAATCGGGCGCAATCAGCATGATCATC 1458	ı	AAT	ттg	ATG		GT(5 GA1	GA1	t TGC	. TT G	GTO	; TTG	AGO	TTT	тст	CAG	ATG	ACG	. ACT		TCA	ATA	ATG	GCA	GAT	GGG	***	ттт	GTG	ATT	AAT	1171
1 TTT GTG TCT GGG TTA TCT TCT GGA TGG CAG ACG GGG GAF ACT GAA CCA TCG TCA ACT ATT GAT CCA TTG CCA CAT TIG CCG CGG TCC 1251 2 AGG GTG GGC GAC TGG GAG TAC CGA GCC CTC GAC AAC TAC TAA GTA GTA GCA CCA TCG TCAACGACGTTGCTGCAATCCAATGGTGCAATCGCAGTCTCGCGTAAACGACGCCT 1276 3 TTT CAT AAT GAC GTG GTC ACA TAT GGA GCA CAG ACT GTA GTA ACT ATA GGG TTG TCG TCG GGT GGT GCG CCT CAG TAT ATG AGT AAG AAT CTG 1260 1 AAT TTC TAA AATGGGTCGAACGCATTGATGCGTTAGGATCATGGGAGTATGGAATGGACGGATGGAGAATTAGAGATTAAGAGATTATGGTGGCACATACACCGGTCATACTCAAGG 1377 2 TTAGAATCTTGGGAGTCGAAGGGAATGGCAGGGAATCAGGAATCAGGAGGGGAACATATACAGCGCATACCAATGTGGAGCGGCGCGATGGACGATAGACATATGTGGCGGGAACTAGAATCACGCGGAACGAATCAGGAGTGGACGAATGCGACGGGCGCGATGGACGATGGGCGCCGATGGCGCCGATGGCGCGCGATGGCGCGAATGAGGATCAGGAATCAGGGGGTGGAATCAAGGAGTCAACGAC AAT ACG GAC TGA AAC AGT AAG TGG CCT GCC ATG 1377 3 TGG GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1350 1 TATATTGGGGCTCCGTGGAGCACATGATGCGAATGGAAGTGAAGGGGAATCAGGCGAATCAGCGAATCATC 1458 2 CTGGGCTGCAGCACTCACGCGCAACAGTGGCGCAATCAGCGCGAACGCGGACCAAAGGGGTCAATCCATC 1442																																
2 AGG GTG GGC GAC TGG GAG TAC CGA GCC CTU GAC AAC TAC TAA CTAA CCACTUTCAACGACGTTTGCTGCAATTCAGTTCATCATGATGGCGCACTCCUCUTAGACGCCCCT 1276 3 TTT CAT AAT GAC GTG GTC ACA TAT GGA GCA CAG ACT GTA GCT ATA GGG TTG TCU TCG GGT GGT GGC CCT CAG TAT ATG ACT AAG AAT CTG 1260 1 AAT TTC TAA AATGGGTCGACGGCATTGATGCGTTTAGGATCATGGGAGGAATGGAAGGGATGGAGGATGGAGGATTAGGAATTAGGGTGCACATAACACGCGGTCATACTCAAG 2 TTAGAATCTTCGGGGTCGCAGGGGGAATGCCCGGGGAACTAGGAGGAATCAGGGGGGGAGGAATTAGAGGATTAAGGAATTATGGTGGCGCACATACGCGGTCATACTCAAG 3 TGC GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCTCCGTGGGAGCACATGGTATCCATGGAGGTGAATCTAGGGCGAATCAGGGGCCAATCATCATCA 2 TTAGGAGTCGCGGGGCGCACATGGTAGCATGGAGGTGAATCTAGGGGGGAACTAAGAGGGGTCAATCACGGCGCACATGGGCGCCGATGGCGCCGATGACCATTATGTACCCATGT 3 TGC GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCCTCCGTGGGACGACTCATGTATCCATGGAAGGGGAATCTAGGCGCGAATCGGCGCAATCGGCGCCAATCATCA 2 CTGGGCTGAGGATCCCGGGCTCCCACTCGCGCAACTGGGCGCGAATCGGCGCAATCGGCGCAATCGGCGCGAATCACCA 1 TATATTGGGCCTCCGTGGACGACTCATGTATGCGAGGGGAATCTAGGCGCGAATCGGCGCAATCGGCGCAATCATC 1 TATATTGGGCCTCCGTGGGCGCCCACTGGCGCAATGGCGCGCGAATCGGCGCAATCGGCGCAATCGGCGCAATCATC 1 TATATTGGGCTTCCGGCGTCCCCACTGCGCAACTGGCGCGCGAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCGAATCGGCGCAATCGGCGCGAATCGGCGCAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGCGGCGCGGCGGCGGCGGCGGGTCAATCGCGCGCG		601	110	101		1 60	GAU		GGI	GAS	,	1 100	110	AAC		611	ACC	. GU /	116	, 114			, 11,		AUA	GGA	GAU	AUI	GAG		GUI	11/0
2 AGG GTG GGC GAC TGG GAG TAC CGA GCC CTU GAC AAC TAC TAA CTAA CCACTUTCAACGACGTTTGCTGCAATTCAGTTCATCATGATGGCGCACTCCUCUTAGACGCCCCT 1276 3 TTT CAT AAT GAC GTG GTC ACA TAT GGA GCA CAG ACT GTA GCT ATA GGG TTG TCU TCG GGT GGT GGC CCT CAG TAT ATG ACT AAG AAT CTG 1260 1 AAT TTC TAA AATGGGTCGACGGCATTGATGCGTTTAGGATCATGGGAGGAATGGAAGGGATGGAGGATGGAGGATTAGGAATTAGGGTGCACATAACACGCGGTCATACTCAAG 2 TTAGAATCTTCGGGGTCGCAGGGGGAATGCCCGGGGAACTAGGAGGAATCAGGGGGGGAGGAATTAGAGGATTAAGGAATTATGGTGGCGCACATACGCGGTCATACTCAAG 3 TGC GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCTCCGTGGGAGCACATGGTATCCATGGAGGTGAATCTAGGGCGAATCAGGGGCCAATCATCATCA 2 TTAGGAGTCGCGGGGCGCACATGGTAGCATGGAGGTGAATCTAGGGGGGAACTAAGAGGGGTCAATCACGGCGCACATGGGCGCCGATGGCGCCGATGACCATTATGTACCCATGT 3 TGC GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1 TATATTGGGCCTCCGTGGGACGACTCATGTATCCATGGAAGGGGAATCTAGGCGCGAATCGGCGCAATCGGCGCCAATCATCA 2 CTGGGCTGAGGATCCCGGGCTCCCACTCGCGCAACTGGGCGCGAATCGGCGCAATCGGCGCAATCGGCGCGAATCACCA 1 TATATTGGGCCTCCGTGGACGACTCATGTATGCGAGGGGAATCTAGGCGCGAATCGGCGCAATCGGCGCAATCATC 1 TATATTGGGCCTCCGTGGGCGCCCACTGGCGCAATGGCGCGCGAATCGGCGCAATCGGCGCAATCGGCGCAATCATC 1 TATATTGGGCTTCCGGCGTCCCCACTGCGCAACTGGCGCGCGAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCAATCGGCGCGAATCGGCGCAATCGGCGCGAATCGGCGCAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGAATCGGCGCGCGGCGCGGCGGCGGCGGCGGGTCAATCGCGCGCG	,	TTT	GTC	TCT	000	; TT	. TC1		r 664	TCO	C.40		600	G A 7	ACT	GAA		TCO				. G		TTT	С Т 4	CG.+	C AT	TTC	000		TCC	1261
1 AAT TTC TAA ATAACGGTCAACGCATTGATGCGTTTAGGATCATGGGAGTATGGAATGGACGGATGGAGAATTAGAGATTAAGAATTATGGTGGCACATACACCGGTCATACTCAAG 1377 2 TTAGAATCTTGGGGTCGCAGGGGGGGGGGACTAGGAGTAGGAGTATGGGGGGAACTATAACAGGGGCATAGCAATGTGGGCGCGATGGGCGGGGGGGG	2	AGG	GTG	GGG	GAG	C TG	G GAG	S TAG	C CG /	A GCO	C C T 4	GAC	:	TAC	TAA	000	CACTG	TCAA	CGAC	GTTI	GCT	CAAT	TCAG	TTCA	TCAA	TGGG	TGCA	TCTC	GCGI	AGAC	GCCT	1276
2 TTAGAATCTTGGGAGTGGGAGTGGAATGGC;GGGAATGAGAGATCAGGAATCAGGGGGAACATATAGAGGGGATGGGGGGCGAATGGGCGCGATGAGCATTATGTACCCATGT 1395 3 TGG GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1350 1 TATATTGGGCTCCGTGGACGATCATGTATCCATGGAAGTGAAGTGAAGGGGAATCAGGGGACAATGGGGCCAATCATC 1458 2 CTGGGCTGAGGATCCGCGGCGCACTCACGCAAGTGGCGAATGGCGCAATCGCGCAATCAGGGGTCAATCATC 1458	3	TTT	CAT		GAC	GT(G GTC	C ACI	A TAI	GG <i>I</i>	GC	CAG	ACT	GTA	GCT	ATA	GGG	TTG	TCG	TCG	GG1	GG1	r GC0	сст	CAG	TAT	ATG	AGT	AAG	***	CTG	1260
2 TTAGAATCTTGGGAGTGGGAGTGGAATGGC;GGGAATGAGAGATCAGGAATCAGGGGGAACATATAGAGGGGATGGGGGGCGAATGGGCGCGATGAGCATTATGTACCCATGT 1395 3 TGG GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1350 1 TATATTGGGCTCCGTGGACGATCATGTATCCATGGAAGTGAAGTGAAGGGGAATCAGGGGACAATGGGGCCAATCATC 1458 2 CTGGGCTGAGGATCCGCGGCGCACTCACGCAAGTGGCGAATGGCGCAATCGCGCAATCAGGGGTCAATCATC 1458													·																			
3 TGG GTG GAG CAG TGG CAG GAT GGA GTA CTT CGG TTA CGT GTT GAG GGG GGT GGC TCA ATT ACG CAC TCA AAC AGT AAG TGG CCT GCC ATG 1350 1 TATATTGGGCTCCGTGGACGATCATGTATCCATGCAATGTGACGTGAATCTAGCGCGCAATCGGCACAAGGGGTCAATCATC 1458 2 CTGGGCTGAGGATCCGGGTGCTCCACTCGGCACAGTGGCGACTCATC 1442	1	. AAT : TTA	TTC GAAT	CTTO	GGAC	AACG GTCG	GTCA/ CAGA(ACGCI GTGGI	ATTG/ AATG(NT G.20 CC 7 G (GAA(NGGA1 CTAG <i>i</i>	GATO	GGAG ACGI	ATCA	GAA1	CGGAC CGGAA	GGA1	GGAG TAC/	SAATI Agcga	CAGAC CATAC	SATT/ CCAA	NAGA/ Egtco	ATTAT	GGTG GGCG	G CAC	TGAC	ACCG CATT	GTCA ATG1	TACT TACCO	CAAG	
2 CTGGGCTGAGGATCCGGGTGCTCCACTCGGCACAGTGGCGACTCATC 1442																																
2 CTGGGCTGAGGATCCGGGTGCTCCACTCGGCACAGTGGCGACTCATC 1442																																
	1	TAT CTG	ATTG GCCT	GGC1	CCG:	TGGA CGGG	CGAT (TGCT)	CATG: CCAC	TATC	CATG(CAAT(Stgg/	GACI	GTG/	ATCI			TCGG	CACI	AGGO	GGTC/	ATC	ATC	14	\$58								
																	GCACI	GGGG	CAT	TCA	rc	14	16									

FIG. 1. Sequences of the serotype 1-3 SI genes of reovirus. The sequences are those of the plus (mRNA) strands.

Relatedness of the 3 S1 Genes. The relatedness of the 3 S1 genes was determined by assessment of base/base identities after locating possible deletions and insertions in each gene and matching cognate sequences. The results are presented

in Table 2. The S1 genes of serotypes 1 and 2 are much more closely related to each other than to that of serotype 3; the absolute homologies of the S1 gene of serotype 3 with those of serotypes 1 and 2 exceed the 25% random matching level

Table 1. Organi	zation of t	ne SI	genes of t	he three	reovirus	serotypes
-----------------	-------------	-------	------------	----------	----------	-----------

Sero- type	Length of 5'- untranslated sequences, bp	Sequences around first initiation codon	Length of long reading frame, bp	Position of second initiation codon	Sequences around second initiation codon	Length of short reading frame, bp	Length of 3'- untranslated sequences, bp
1	13	CCT ATG G	1254	75*	GCA ATG G	357	188
2	13	CTC ATG T	1197	66	GGA ATG G	375	229
3.	12	CGG ATG G	1365	71	ATA ATG G	360	36

*This is the third initiation codon. The second initiation codon is followed by a termination codon 10 codons downstream.

Table 2. Relatedness of the 3 S1 genes of reovirus

		% homology		
Serotypes	Long reading frame	Conserved portion of 3'-terminal untranslated region	σ 1 proteins	
1:2	28	46	36	
1:3	5	21	16	
2:3	9	27	14	

All values are corrected for random coincidences (25% for nucleic acid and 5% for protein sequences, respectively).

by only 5% and 9%, respectively. The homologies of the long and short open reading frames are similar. Interestingly, there is a rather highly conserved 96-bp sequence near the 3' terminus of the plus strand of all 3 genes (Fig. 2 and Table 2); for this sequence the direct homology values, uncorrected for random coincidences, range from almost 50% to >70%. The significance of this high degree of relatedness is not clear, especially since only in the serotype 3 gene is this sequence in the coding region.

Amino Acid Sequences of the Three σ 1 Proteins. The amino acid sequences of the three σ 1 proteins are presented in Fig. 3. There is significant homology between all these proteins provided that account is taken of several deletions/insertions. The sequences of the serotype 1 and 2 proteins are remarkably congruent; there is only one deletion or insertion of 7 amino acids starting at position 150 and two minor deletions/insertions near the COOH terminus. The sequence of the serotype 3 protein is only slightly less congruent; relative to the sequence of the serotype 1 protein, it contains an insertion of 4 amino acids at positions 33–36, a deletion of 14 amino acids starting at residue 153, and a deletion of 7 amino acids starting at position 316.

When cognate sequences are brought into apposition in this manner, the serotype 1 and 2 σ 1 proteins are 41% homologous with each other and 21% and 19%, respectively, with the σ 1 protein of serotype 3. Forty-three amino acids (slightly more than 10%) are shared by the σ 1 proteins of all three serotypes.

The amino acid compositions of the three σ 1 proteins are unremarkable. They contain about the same number of hydrophobic and uncharged polar amino acids and of basic and acidic amino acids. The serotype 1 σ 1 protein lacks cysteine, and the others contain only one molecule of cysteine. Thus, the pairs of σ 1 protein molecules that exist in reovirus particles are unlikely to be -SS-linked dimers.

The hydrophobicity profiles of the three σ l proteins are shown in Fig. 4. The serotype 1 and 2 σ l protein profiles are seen to be rather similar, but that for the serotype 3 protein is quite different, except in the terminal regions. It is conceivable that the domain for the conserved σ l function—namely, ability to bind to cells—is located in one or other of the terminal regions.

Molecular Properties of the Proteins Encoded by the Short Reading Frames. The short reading frames of the SI genes of serotypes 1, 2, and 3 are capable of encoding proteins that are 119, 125, and 120 amino acids long, respectively. These proteins are basic: they contain 29–35 lysine and arginine residues but only 10–13 aspartic and glutamic acid residues. They contain 46–49 nonpolar and 39–44 uncharged polar amino acids. The proteins of serotypes 1 and 2, 1 and 3, and 2 and 3 share 35%, 17%, and 27% homology, respectively.

DISCUSSION

The SI gene is the reovirus gene that has diverged most markedly during evolution (1). The serotype 1 and 2 SI genes are still related, after taking account of sequence deletions and insertions, to the extent of 53%, but the serotype 3 SI gene is related to the other two to the extent of only 30% and 34%, respectively. After adjusting for a background of 25% for random matching, these values reduce to absolute homology values of 28%, 5%, and 9%, respectively.

These gene homology relationships are reflected in those of the three σ 1 proteins. The serotype 1 and 2 σ 1 proteins share 41% of their amino acid residues. Many of the amino acid substitutions are conservative and the hydrophobicity profiles of the two proteins are very similar. The serotype 3 σ 1 protein, on the other hand, is very different. It shares no more than 20% of its amino acids with either the serotype 1 or the serotype 2 σ 1 protein, and its hydrophobicity profile is significantly different from that of the other two.

The reading frames from which the σ 1 proteins are translated are clearly the long open reading frames that start at residue 13 or 14. The fact that the initiation codons at the beginning of these reading frames are weak is in accord with the fact that the S1 mRNAs are translated poorly and that only small amounts of the σ 1 proteins are formed in infected cells (16). By contrast, the initiation codons that open up the short open reading frame are strong, and ribosome binding protection (17) and f-Met dipeptide formation experiments (18) indicate that they are functional. It should be possible to identify these proteins in infected cells, possibly by the use of antisera directed against certain regions of these putative proteins.

A curious feature of the σ^1 protein of serotype 3 is its length; it is the largest of the three σ 1 proteins, but in every NaDodSO₄/polyacrylamide system that has been used—in particular, in the phosphate-based system of Zweerink et al. (19) and in the Tris/glycine systems of Laemmli (20) and Maizel (21)—it migrates more rapidly than the σ 1 proteins of serotypes 1 and 2. In fact, in the Maizel system the serotype 3 σ 1 protein migrates faster than the serotype 3 σ 2 protein, the molecular weight of which is known to be about 38,000 (9); in the other two systems its molecular weight, compared with that of a variety of marker proteins, is about 42,000 (10, 19). However, its actual molecular weight is almost 48,000. It has been suggested that the σ 1 protein may be associated with the plasma membrane (22); but the σ 1 protein in mature reovirus particles does not appear to be a cleavage product of a precursor because the size of the free form of protein $\sigma 1$ in infected cells is the same as that of the form that exists in mature virus particles (19). Thus, the reason for the size discrepancy remains to be discovered.

The various monoclonal antibodies that have been raised against the σ 1 protein of reovirus serotype 3 differ in their relative ability to neutralize infectivity on the one hand and inhibit hemagglutination (that is, cell binding) on the other (23, 24). This suggests that the σ 1 proteins comprise at least two domains—one containing the dominant epitope and the

Serotype

1	AATGGACGGATGGAGAATTAGGAGATTAAGAATTATGGTGGCGCACATACACCGGTCATACTCAAGTATATTGGGCTCCGTGGACGATCATGTATCCATG	1411
2	.GAT.CCGCC.CCCAT.A.CGCA.TCG.CGATCTC	1394
3	.GCAT.C.TCG.T.ACGTGT.G.GGTGGC.CA.TTACGCT.AA.CAGTA.GCG.CATCG.TTCCGC.	1367

FIG. 2. Sequences of conserved regions near the 3' termini of the plus strands of the 3 SI genes. Identity is indicated by periods. The sequences are 96 nucleotides long.



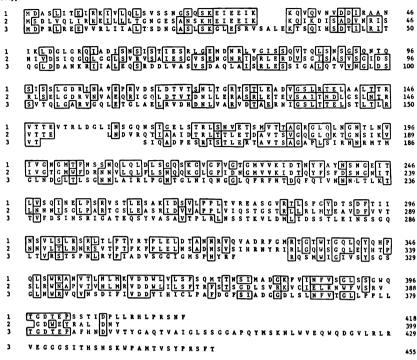


FIG. 3. Sequences of the σ 1 proteins of reovirus serotypes 1–3.

other containing the region with affinity for cell receptors. Further, the amino acid residues that are shared between the three σ 1 proteins tend to be clustered in the terminal regions (see Fig. 3); this is indicative of conservation of function and also suggests the existence of domains. Availability of the sequences of the three σ 1 proteins opens up several approaches for detailed studies concerning their functions. A remarkable aspect of the 3 SI genes is the nature of their genetic relatedness pattern. The total genomes of serotypes 1 and 3 are far more closely related to each other than to that of serotype 2 (1). All individual genes show the same pattern: all individual genes of serotypes 1 and 3 are much more closely related to each other than to those of serotype 2. The only exceptions are the 3 SI genes for which the sequence

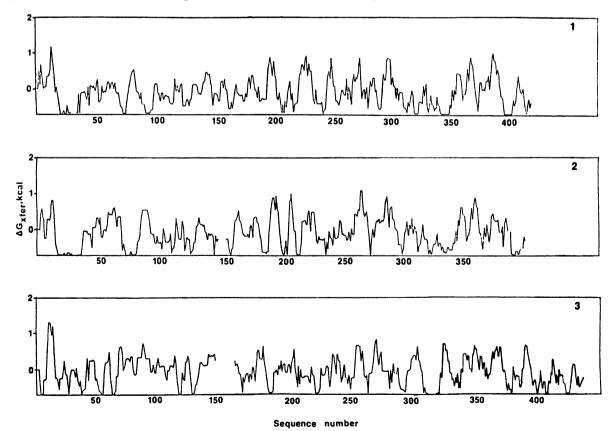


FIG. 4. Hydrophobicity profiles (15) of the σ l proteins of reovirus serotypes 1–3.

28 Biochemistry: Cashdollar et al.

analysis presented here demonstrates that the genes of serotypes 1 and 2 are much more closely related to each other than to that of serotype 3. This agrees with the antigenic evidence that antibodies against serotype 3 protein σ 1 are absolutely specific, whereas those against serotype 1 and 2 σ 1 proteins are partly cross-reactive (25). Thus, 9 of the 10 reovirus genes show a serotype 1:3 relatedness pattern, whereas the SI gene exhibits serotype 1:2 relatedness. Perhaps the ready occurrence of gene reassortment is responsible for this pattern of evolution, for one possible explanation for it is that at some time during evolution the SI genes of serotypes 2 and 3 became associated with each other's (that is, the heterologous) gene pools. It is conceivable that the evolution of an ancestral reovirus did indeed proceed along three independent major pathways, yielding three independent gene pools (1), with the SI gene diverging most rapidly because of selective immunologic pressures on the one hand and tolerance of extensive structural diversification on the other (because protein σ does not appear to be a true capsomer component, in contrast to the other two reovirus outer shell components, μ 1C and σ 3). Reassortants would be formed readily but be selected against because heterologous capsids would be less stable than homologous ones. At some relatively recent stage of evolution, exchange of serotype 2 and 3 SI genes might then have been favored by some selective mechanism like improved stability of heterologous σ 1-capsid combinations caused by a structural modification, or the characteristics of the available immunologic pressures, perhaps modified by the acquisition of new host species.

We gratefully acknowledge the expert technical assistance of Patricia McClellan-Green and Bruce Macdonald. This work was supported by Research Grant 1 RO1 08909 and Program Project Grant 2 PO1 CA 30246 from the National Institutes of Health.

- 1. Gaillard, R. K. & Joklik, W. K. (1982) Virology 123, 152-164.
- Weiner, H. L. & Fields, B. N. (1977) J. Exp. Med. 146, 1305– 1310.

ž

- Weiner, H. L., Ramig, R. F., Mustoe, T. A. & Fields, B. N. (1978) Virology 86, 581-584.
- Lee, P. W. K., Hayes, E. C. & Joklik, W. K. (1981) Virology 108, 156–163.
- 5. Weiner, H. L., Drayna, D., Averill, D. R., Jr., & Fields, B. N. (1977) Proc. Natl. Acad. Sci. USA 74, 5744-5748.
- Fontana, A. & Weiner, H. L. (1980) J. Immunol. 125, 2660– 2664.
- Weiner, H. L., Greene, M. I. & Fields, B. N. (1980) J. Immunol. 125, 278–282.
- 8. Sharpe, A. H. & Fields, B. N. (1981) J. Virol. 38, 389-392.
- Cashdollar, L. W., Esparza, J., Hudson, G. R., Chmelo, R., Lee, P. W. K. & Joklik, W. K. (1982) Proc. Natl. Acad. Sci. USA 79, 7644–7648.
- Smith, R. E., Zweerink, H. J. & Joklik, W. K. (1969) Virology 39, 791–810.
- Cashdollar, L. W., Chmelo, R., Esparza, J., Hudson, G. R. & Joklik, W. K. (1984) Virology 133, 191–196.
- 12. Messing, J., Crea, R. & Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.
- 13. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 14. Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
- 15. Rose, G. D. & Roy, S. (1980) Proc. Natl. Acad. Sci. USA 77, 4643-4647.
- 16. Zweerink, H. J. & Joklik, W. K. (1970) Virology 41, 501-518.
- 17. Kozak, M. (1982) J. Mol. Biol. 156, 807-820.
- Cenatiempo, Y., Twardowski, T., Shoeman, R., Ernst, H., Brot, N., Weissbach, H. & Shatkin, A. J. (1984) Proc. Natl. Acad. Sci. USA 81, 1084–1088.
- Zweerink, H. J., McDowell, M. J. & Joklik, W. K. (1971) Virology 45, 716–723.
- 20. Laemmli, U. K. (1970) Nature (London) 227, 680-685.
- 21. Maizel, J. V. (1971) Methods Virol. 5, 177-246.
- 22. Kauffman, R. S., Lee, S. & Finberg, R. (1983) Virology 131, 265-273.
- 23. Hayes, E. C., Lee, P. W. K., Miller, S. E. & Joklik, W. K. (1981) Virology 108, 147-155.
- 24. Spriggs, D. R., Kaye, K. & Fields, B. N. (1983) Virology 127, 220-224.
- 25. Gaillard, R. K. & Joklik, W. K. (1980) Virology 107, 533-536.