Determination of the complete nucleotide sequence of the Sendai virus genome RNA and the predicted amino acid sequences of the F, HN and L proteins

Tatsuo Shioda, Kentaro Iwasaki<sup>1</sup> and Hiroshi Shibuta\*

Department of Viral Infection, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108 and <sup>1</sup>Department of Physiological Chemistry, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113, Japan

Received 14 January 1986; Accepted 24 January 1986

#### ABSTRACT

We previously determined the 3' proximal 5,824 nucleotides of the Sendai virus genome RNA (Nucleic Acids Res. 11, 7317-7330, 1983; Nucleic Acids Res. 12, 7965-7973, 1984), and present here the sequence of the remaining 5' proximal 9,559 nucleotides. Thus, this is the first paramyxovirus to have its genome organization elucidated. The set of complementary DNA clones used was prepared by the method of Okayama and Berg from polyadenylylated viral genome RNA. We sequenced the region containing the 5' proximal half of the F gene, and the subsequent HN and L genes, and predicted the complete amino acid sequence of the products of these genes. Sequence analyses confirmed that all the genes are flanked by consensus sequences and suggest that the viral mRNAs are capable of forming stem-and-loop structures. Comparison of the F and HN glycoproteins of Sendai virus with those of simian virus 5 strongly suggests that the cysteine residues are highly important for maintenance of the molecular structures of these glycoproteins.

#### INTRODUCTION

. Recently, the molecular cloning of complementary DNA (cDNA) to RNA has greatly contributed to the elucidation of the gene and genome structures of several RNA viruses. Analyses of the viral gene and genome structures provide essential information for the solving of several important problems such as the mechanisms of viral replication and transcription, the nature of each viral protein, biological events caused by viruses and genetical relationships among viruses.

Of paramyxoviruses, Sendai virus has been most extensively studied as a prototype, its genome being a single continuous RNA of about 15 kilobases long with negative polarity. Sendai virus contains six structural proteins, i.e. large (L), RNA polymerase (P), nucleocapsid (NP), hemagglutinin-neuraminidase (HN), fusion (F) and membrane (M) proteins. In addition, at least one non-structural viral protein designated as C (1) and a small transcript called leader RNA (2) have been identified in infected cells. In previous communications (3,4), we reported the sequence of the 5,824

# **Nucleic Acids Research**

nucleotides from the 3' end of the Sendai virus genome RNA and determined the complete primary structures of the first three genes and a part of the fourth gene. Our results as well as data presented by others established that the gene order is 3'-leader-NP-P+C-M-F- (3,4,5,6,7,8,9), and the subsequent gene order has been proposed to be -HN-L-5' (10,11). In order to fully elucidate the genome structure of Sendai virus, we continued to construct cDNA clones toward the 5' end, and present here the full sequence of the remaining region, that is, the 5' proximal 9,559 nucleotides, which contains the 5' proximal half of the F gene and the subsequent HN and L genes. Thus, we have determined the complete primary structure of the genome of Sendai virus strain Z, and have predicted the amino acid sequence of each gene product. We also compare the predicted structures of the F and HN glycoproteins of Sendai virus with those reported for another paramyxovirus, simian virus 5 (SV5) (12,13).

### MATERIALS AND METHODS

#### Preparation of viral RNA

Sendai virus strain Z was used. Viral 50S genome RNA was prepared from virions purified from infectious allantoic fluids of chicken eggs as described previously (3). Sendai virus mRNAs were prepared in a similar way to as described previously (3) using BHK-21 cells as the host cells, and RNA was extracted by the guanidium thiocyanate method (14). The mRNAs were selected from crude RNA by oligo(dT) cellulose column chromatography.

### Synthesis and cloning of the complementary DNA (cDNA)

Sendai virus genome RNA was polyadenylylated according to Inokuchi et al. (15) except that the amount of ATP:RNA adenyltransferase (poly A polymerase) was increased to 20-fold. Synthesis and molecular cloning of cDNA from the polyadenylylated genome RNA were performed according to Okayama and Berg using a pSV7186-derived vector-primer (16). <u>E. coli</u> K12 strain HB101 was transformed with the resulting recombinant plasmids by the standard method (17).

#### Sequence determination of cDNA

CDNAs were cleaved into fragments with appropriate restriction endonucleases, and after subcloning of the fragments into M13 phage (18), their nucleotide sequences were determined by the method of Sanger et al.(19).

# **Colony hybridization**

Bacterial colonies grown on nitrocellulose filters were lysed and fixed

according to Grunstein and Hogness (20). The nitrocellulose filters were used for hybridization (21) with the viral 50S RNA probe labeled with  $[\gamma - {}^{32}P]$  ATP (22) or the cDNA probe labeled with  $[\alpha - {}^{32}P]$  dGTP (23).

### Northern blot hybridization

Viral mRNA resolved in an agarose gel was transferred to a nitrocellulose filter as described previously (3). Hybridization of this filter with  $^{32}$ P-labeled cDNA was performed as described above.

# Enzymes and other materials

ATP:RNA adenyltransferase, <u>E</u>. <u>coli</u> DNA ligase, terminal deoxynucleotidyl transferase and ribonuclease H were purchased from P-L Biochemicals, Milwaukee, U.S.A.; avian myeloblastosis virus reverse transcriptase from Seikagaku Kogyo, Tokyo, Japan; M13 cloning and sequencing kits, nick-translation kits and all the radioactive compounds from Amersham Incorporation plc, Amersham, England; HAWP nitrocellulose membrane filters from Millipore, Bedford, U.S.A.; and bovine alkaline phosphatase, T4 polynucleotide kinase, <u>E</u>. <u>coli</u> DNA polymerase I, the Klenow fragment of DNA polymerase I, T4 DNA ligase and all the restriction endonucleases from Takara Shuzo, Kyoto, Japan.

### RESULTS

## Cloning and sequence determination of cDNA

We previously described the molecular cloning and sequencing of the Sendai virus genome cDNAs which cover the 3' proximal 5,824 nucleotides (3,4). In order to obtain cDNA clones which represent the remaining genome region, we tried to clone cDNAs which were prepared by reverse-transcription of partially digested and then <u>in vitro</u> polyadenylylated viral genome RNA. For this, 6 micrograms of Sendai virus genome RNA was incubated with 4 units of poly A polymerase, which is about 20-fold the amount used for the standard polyadenylation reaction. After the incubation, the 50S genome RNA was found to have been digested to an average size of about 28S (data not shown), which was most likely caused by the ribonuclease activity contaminating the poly A polymerase preparation. Subsequent synthesis and molecular cloning of cDNAs from the polyadenylylated RNA were performed according to Okayama and Berg (16).

Transformation of  $\underline{E}$ . <u>coli</u> HB101 with the recombinant plasmids yielded about 500 ampicillin-resistant colonies, from which 41 were selected on the basis of the intense hybridization signals with the Sendai virus genome RNA probe. Out of these clones, 27, which did not hybridize with probes of cDNA

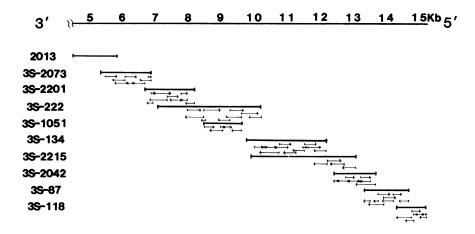


Fig. 1. Locations of previously obtained cDNA clone 2013 and newly prepared cDNA clones 35-2073, 35-2201, 35-222, 35-134, 35-1051, 35-2215, 35-2042, 35-87 and 35-118 in the genome RNA and the sequencing strategy. Arrows indicate the cDNA fragments sequenced and the direction of sequencing.

fragments corresponding to the NP, P+C or M genes (3,4), were selected by the colony hybridization technique. The mutual relationships among the 27 clones as well as their relationships to the previously reported clones were determined by colony hybridization tests using appropriate cDNA probes in combination with restriction map analyses, and it was found that these clones could be lined without any gap in order from the 3' proximity to the 5' proximity. We selected 9 clones for sequencing. The sequential relationships between these 9 clones and to the previously reported clone 2013 together with the sequencing strategy are shown in Fig. 1. Clone 3S-2073 showed an overlap of about 400 nucleotides with clone 2013, which had been shown to cover the 5' proximal 331 nucleotides of the M gene and the 3' proximal 1,013 nucleotides of the F gene (4). Clone 3S-118, which was identified as the most 5' proximal clone, had an identical sequence with the reported 5' terminal 70 nucleotides of the Sendai virus genome RNA (24) except for 3 substitutions, which was followed by 6 guanine residues and a PstI site derived from the linker DNA fragment. This indicates that 3S-118 included the very 5' end portion of the genome, and that we had obtained a set of cDNA clones that covered the entire genome RNA. By analyzing these 9 clones, we determined the sequence of the 9,559 nucleotides of the Sendai virus genome, which followed the 3' proximal 5,824 nucleotides that we had previously determined (3,4). Thus, we concluded that the total length of

the Sendai virus genome RNA is 15,383 nucleotides, which is in good agreement with the previous value estimated from the sedimentation coefficient obtained with a sucrose density gradient (25,26). The nucleotide sequence from position 4,781, the 5' end portion of the M gene, to 15,383, the 5' end of the genome RNA, is shown in Fig. 2.

## Analysis of the nucleotide sequence

Distribution of the translation termination codons in the plus strand (complementary strand to the genome) of the above region is shown in Fig.3, indicating the presence of three large open reading frames. Open reading frame op-4, which starts from nucleotide position 4,865 and encodes the F protein as reported previously (4), was found to terminate at nucleotide position 6,559, encoding a polypeptide of 565 amino acids with a molecular weight of 61,495. Downstream of op-4, two open reading frames, designated as op-5 (from 6,692 to 8,416) and op-6 (from 8,555 to 15,238), were observed. Op-5 corresponds to 575 amino acids, the calculated molecular weight of which is 63,408, and op-6 encodes a giant polypeptide of 2,228 amino acids with a calculated molecular weight of 252,864.

Previous studies (3,4,5,6,7,8) showed that the Sendai virus NP, P+C and M genes are flanked by consensus sequence R1 (UCCCACUUUC or UCCCAGUUUC) at the 3' end and by R2 (AUUCUUUUU) at the 5' end, and that the junctiuon between these genes is composed of R2-GAA-R1. This structure was also found for op-4, op-5 and op-6, i.e. they are preceded by UCCCUAUUUC, UCCCACUUUC and UCCCACUUAC, respectively, which were assigned for R1, and followed by AUUCUUUUUU (R2), and the junction between R1 and R2 was GAA or GGG. Thus, we assigned the sequences from R1 starting at position 4,812 to R2 ending at 6,632, from 6,636 to 8,523 and from 8,527 to 15,326 as the fourth, fifth and sixth genes, respectively, and GAA as the intergenic sequence with one exception between the fifth and sixth genes where it was -GGG-. It is of interest to note that the sequence, -GAA-, was again found after R2 of the last gene.

Since we have already assigned the first to fourth genes for NP, P+C, M and F, respectively, the fifth and sixth genes correspond to the HN and L genes. The coding capacity of op-5, 63,408, is very close to the reported value for the molecular weight of the unglycosylated form of the HN protein of 63K (27,28), while that of op-6, 252,864, is very similar to the estimated molecular weight of the L protein of about 200K (29). In order to confirm that op-6 really produces a large transcript, northern blot hybridization was performed between mRNAs from infected cells and a probe,

	10	20	30	40	50	60	70	80	90	100	
	R2	<b>P1</b>							-CCAGUGAGGA		4800
				UCACGAACCA							4900
op-4										UUUAGUGACU	5000
				AUCAUGACUC							5100
				CUCCCUACGG							5200
				CCAUGAUAGC							5300
				UUAGCUACUG							5400
				CUAGUUUGGG							5500
				AGCUUAAAGC							5600
	AAUGACUCUA	AUACUGGUGU	UAGUCCUGUC	CCGUCAGAUU	GUAGAGACUA	CAGUAAAUAU	GUCUUGUCUA	GUUUCCUUGC	CACUAUCUAC	ACCUAGAUCU	5700
	CUCUAUGUAC	CAGUGGGACA	GACACUUCUA	GGGAUAAGAA	AGACUUCAGG	GUCCACACGA	GUAUGUGUUC	CGUAGUAGAU	AAAGAAUGUU	GUAUCUGCCC	5800
	CUCCUUACCA	UACACUGACA	GGGGUCGGUA	UAUGAGUCAG	CACGAAGAAA	GAAUCCCCCA	CGUCUGUAUU	GGCUAACACA	ACUCAGGUCU	AACUGGAUAU	5900
	AUACGGGGUC	CCUAGGGCGU	GUUGACUAUG	GACUGUCGGU	CGUUUUCACA	UAGGACCCCC	UGUGUUGUUC	CACAGGACAG	UGUUUUCAAC	ACCUGUCGGA	6000
	AUAGGGGUUC	AAACGAAAAC	ACUUACCCCC	GCAACAACGA	UUGACGUAUC	GUAGGUGUAC	AUGGACGCCC	UGUCCGGCUU	CUGGUUAGUC	AGUCCUAGCG	6100
	AGAUUUCCAC	AUCAUAAGGA	UUGGGUACUG	UUGACACCAG	AAUAUCCACA	GUUACCCCAU	CUUAACAUAC	GAUUGGCCUC	UCCCGUGCUA	CGGUGAACCC	6200
	CCCAGGUCUU	GAACUGUCAG	CCAGGACGUU	AACGAUAGUC	UGGGCAACUA	UAAAGAGAGU	UGGAACGACU	ACGAUGCUUA	AAGAACGUUC	UGAGAUUCCG	6300
	ACUCGAACUC	UUUCGUGCCU	UUUAGGAGAG	CCUCCAUCCA	UCUACCAUGU	UGAGUUCUCU	CUGACACUAA	UGCUAGUAUC	AUCAAUACCA	GCAUUAUAAC	6400
	CACCAGUAAU	AUCACUAGUA	GUAGCACGAA	AUAUCUGAGU	CUUCCAGUUA	CGAUUACCCA	UUAGGUCUAC	UGGCAUAUGG	CUCCCUGUGU	AUGUGUAAUC	6500
	UCGGCUUCUA	GUCUGUAUAC	AUGUGUUUGC	CACCCAAACU	ACGUUACCGA	CUCUUUUCU	CUAGUGCUGG	UAAUAGUCUA	CAGAACAUUU	CGUCCGUAUC	6600
	AUAGGCAACU	CUAGACAUAU	AUUAUUCUUU	UUGAAUCCCA	CUUUCACUCC	AGCGCGCCAU	GAAAUCGAAA	GUGGAGUUUG	UUCGUGUCUA	UACCUACCA	6700
	CUAUCCCCGU	UUGCACUGAG	CAUGACCAGA	UGAAGAGGAU	CACCAUCGUG	GUGUUUUGGU	CGUAGUĆCAA	CCCUCUCCAG	UUCAUUUCGG	CUGUGUACCA	6800
	ACGACUAAGA	GAGUAAGUGG	GUCACCCGAA	ACAGUUAACG	GUGUCACUAG	UAGACAUAGU	AUUAAAGACG	AUCUGUUCCC	AUAUCAUACU	UUCUCAUGAG	6900
op-5	UUACUGACAU	CUCCGUAACU	UGUACUCGUC	GUCCCUCCAC	UUUCUCAGUG	AAUGGUCAGA	UUAUUCCGUU	CUCCAAUAUC	GUUCCCGACA	GUUGUAAGUC	7000
5	UCGAGACACG	UUUGGCCUUA	GGGUCAGAAC	AACUUGUUUU	UGUCGUCCCU	ACAGUAGGUC	UACUAACUAU	UCAGCACGUC	GUCUGUUCUC	GAGUGAGUCG	7100
	UGACACUCUC	AUGCUAGCGU	CAGGUGGUAC	GGCUACCUUA	ACGGGGUGAA	CUCGGUGUAU	CAAAGACCUC	UACGGGACAG	CCUCUUGGCA	UAGAAUCGAG	7200
	UCUAGGACUU	UAGAGUAACG	ACGGACCAGG	CUCGAACAAU	AGACCAAGAU	GUUGCUAGAG	ACCUACACAA	UCCGAGGGAA	GUGAGAGUUA	ACCGCUCCGU	7300
	UAGAUACGGA	UAAGUAGUUU	AGAGUAAUGU	GUUCCAACAC	GACUGUAUCC	CUUUAGUAUA	GUCCAGGACG	UCGAUCCCAU	GUAUAGUGAG	UUAAGUCUAU	7400
	ACAAGGGACU	AGAAUUGGGG	CAUCACAGGG	UGUGAAUACU	GUAGUUGCUG	UUAGCCUUUA	GUACGAGACA	CCACCGUUGG	CCCUGAUCCC	CAAUAGUCGA	7500
	AACGAGGUAC	GGCUGACAUC	UGCUUUCUUG	GCUGAUGAGA	UCACUACCAU	AACUCCUAGA	CCAGGAACUA	CAGGACCUAG	AGUUUCCCUC	UUGAUUCAGA	7600
	GUGGCCAUAG	CGUUGUCGCU	CCAUCUAGAA	CUAGUGGGCA	AGAGACGUGA	UAUGGGGUCA	CAUCCGUUGC	CGUAACGUUG	UCUUCCGAGU	AACUAUAAAAG	7700
	AACCCAUACC	ACCUGAUUGG	UGGGGAGACG	UCCCACUAUG	UUUUACAUCC	UGGGUUCCUA	CGGUUGUCCA	CAGCGUUCUG	UGUACGUUAC	UCCGAGACUU	7800
				GUCGCACUAG							7900
				AUUUUUAAUUUAA							8000
	CUCAUGAACU	ACAGUCGGUG	GGAAACUGAU	AGUUGACCUG	UGGAGUACUU	CGGAACAGAU	CUGGUCCUUU	AUUUCUCACG	UUAACCAUGU	UAUUCACAGG	8100
				ACGAAUAGGU							8200
	UUGGGUUGUU	AGUACAUAAG	AUUGUGAUGA	UUGUAAUAUU	UAUACAAUUC	CUAUUUCCUA	CAAGUUAAUC	UCCGACGUAU	AUGGUGCUGU	AGCACAUAGU	8300
				AGUAGCUCUA							8400
				AUCGUCCGAA							8500
										UAUGAGAUAG	8600
1				CCCCCUUCUA							8700
				GCCUCCUAAC							8800
9	AAUCUGGCUA	UGUGCAAACU	UGGCAUGGGU	UGGAUGAGAG	UCCUUAAUGA	AUCCGAACUA	UAUGGUCUCU	AUACACUGUU	UUAGGCUAGG	CAGAAGCGCC	8900
ġ				CACCCAAGGU							9000
				CUAUUGACUA							9100
				CCUGGGGAAC							9200
	UAGAACAGUA	CUAUGACUUG	UUCAACUGUA	ACUGUCCCAU	AUAGGAUUGG	GGACUCGACC	AGAACUACAU	AACACUACAA	CAUCUUCCUU	CCACCUUAUA	9300
				GUAACCCUAU							9400
				GGGGAUAGUG							9500
				CAUCUCUGCA							9600
				GAAGAAAUCC							9700
				GAUAUGCUCA						,	9800
				CAGAUCUUGA							9900
				AUAUCUUGGU							10000
	CGUACCCUGA	GACAUAUGGG	CCUAUCAUUA	GACAUGAUAU	UUCGGGGUCU	CAGACUUCUC	UGGGCCGCCG	AAUAACUUCA	CAAGUAUUUA	CUACUCUUAA	10100

		10	20	30	40	50	60	70	80	90	100	
	1			UUAAUACACC			CUCCUCCUCA		CAUGUCAGAG		ICUACINICOU I	10200
UNICCUUC: MUNICUUC UNUEURGAU CAMUUUKUM ACUALINGA MACKAGU COCKIGGGU CUULAUMA MAURCAUUU UNUEURANINGA 1956   UNICUUUU UCUUUKUU CUUUUKUU CUUUKUKA AKEECECKUU UCUUUKUKA GUCUUKUU MAURCAUUU MAURCAUUUUKUUKUU MAURCAU CUUUKUKUU MAURCAUUU MAURCAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
GUALUSCUU CUCUURGUU CCGURCUUU UNUEURGA ACCCCCCA MACCUCCUU UCUCUCUCA ACCUCUA ACCUCUCU ACCUUCUCU ACCUUCUCUU ACCUUCUCU ACCUUCUCUU ACCUUCUCUU ACCUUCUUU ACCUUCUUU ACCUUUUU ACUUCUUU ACUUUUUU AUAA CCUUU AUAA ACUUUUUU AUAA CCUUU AUAA ACUUUUUU AUAA CAUUUUUU AUAA CCUUU AUAA ACUUUUUU AUAA CAUUUUUU AUAA CAUUUUUU AUAA CAUUUUUU AUAA CAUUUUUU AUAA CAUUUUUUU AUAA CAUUUUUU AUAA CAUUUUUU AUAA CAUUUUUU AUAA CAUUUUUUU AUAA CAUUUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUUU AUAACUUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUU AUAACUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUUU AUAACUUUUU AUAACUUUU AUAACUUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUU AUAACUUUUUU AUAACUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUU AUAACUUUUUU AUAACUUUUUUUU												
UBJCUECCE AUNCHIEGE AURCHIEGE MARCHEREN 10660   ACQUINCULU MARANCEAM GUUCUSAM AURCHIEGE AURCHIEGE 10760   BUGGEDING MARANCEAM GUUCUSAM GUUCUSAM AURCHIEGE AURCHIEGE 10760   BUGGEDING MARANCEAM GUUCUSAM GUUCUSAM CUULUARUSA	1											
ACCURGENCU AUMANCEGAN GUICUNGAME AMUUGACU ACUMAGUE GUAACUUC ACUNGUNU UKENACUUC MAGUAGUE GUICACUUC 10700   UGGECUCU AUMANCEGAN GUICUNGUE GUICAUGUI GUIAUUUUUUU CUI AUMAGUUU CECCECCEUL UCUUCAUU ACUNCECAUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
USECULACU ACUURUMA GUCUNAUA GUCUNAUA GUCUNAUA GUNUNACUA UNUUCAUN ACUURUMAUA ANAGUUCUA GUNUNCUA ANAGUUCUA GUNUNUUCUU UNUUNAUUA ANAUAUUA UNUUUUAUA ANAUAUUA UNUUUUUU UNUUNAUUA GUNUNAUUA ANAUAUUAU UNUUUUUUUU UNUUNAUUUU UNUUNAUUA GUNUNUUU UNUUNAUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUU UNUUNAUUUU UNUUNAUUUUU UNUUNAUUUU GAAACAANA GAAUNUUUU UNAUAUUUUU NAUAUUUUU GAAUNAANA UNAUUUUUU AANAANAA UNAUUUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNAUUUUUU UNUUNAUUUUU UNUUNAUUUU UNUUNAUUUUU UNUUNAUUUUU UNUUNUUUU UNUUNUUUU UNUUNUUUU UNUUNU												
UNARAGUNG UAACUUNAG UGACUUNCUC UUUUUUUUA CAALUCCU UUUUUUU AINAAGCAC ACAAUUCUU GAAUUCGAA CUUUUCUU 10990   UGUCAAUU AAUUCUUCU UUUUUUUUU AAUUUUU UUUUUUUU												
CAUGGACAUE GAUGUCUAMU UUUUUUUUUU CAUGUCUCU UUUUUUUUU CAUGUCUUU AUCCUUUUUA AUCCUUUUUA CUUUUUUUUUU CUUUUUUUUU CUUUUUUUU CUUUUUUUUU CUUUUUUUUU CUUUUUUUUU CUUUUUUUUU CUUUUUUUUU CUUUUUUUUUU CUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU											1	
UGUCGAUUU UAACUGCUC UGUMGUAM CAUCGUCUA CAMACASUA UCAUUUUCCU AUAUSAUKU ACCUCUUA AAUSGUGUC CGUAGUCUA CAGUCUUU GAACAGUCU UGUMGACUU GUAUGUUG GUUUUUCCAUL 11200   UAMAGACAM AUSULCCGUA GARCUCUU UGACAGUCUU GAACAGUCU UCAACCUCU UAUGUAUUU AGUUCUUM GAUUGUUUUCCU 11300   AUAUGCUAU MARUCCCUA GUUCUUA CCACUCUU GAACAGUCU UCAACUUCCU UAUGUAUUU AGUUCUUM GAUUGUUU GAUUCUCUU   ACAMUCUUM UAUCCUUG UCCGAUCGUAC GUUACAAC UUAGUACUU UCAACUUC UAUUUUM AGUUCUUM GUUCUAUU   AUAUCUUUM UAACUUUU CUAUUUU CCCAA GUUUGUAU AAGUAUUU UUAAGUUU UCAACUUC UUUUUUUM AGUUCUUM GUUCUUCUU   AUAUCUUUU UAUACUUAU GUUCUUACUU CUACGAA GUUAGAAUU UAACAAUG GAUCUUUUU UAUUUUUM AGUUCUUM GUUCUUCUU GUUUUCUUU GUUUUUUUU   AUAUUUUUUU UAUCUUAUU AUAAUAUUUUU UUAAGAAUU AAUUUCUUUU UUUUUUUU GUUUUUUUU AGUUCUUUU UUUUUUUUUU											1	
GARCUGGUUC ACACAUMAGA CEMAGCUUG UGACAUCUA CUUUUGUCUA GACGAACMAG CUUUUNAGAU UGAUUUGUAUA GCUUUURACA 11200   AUMAGAGAU MUGAAUUC CAUGUAACGA AACMAUUUU GGACAUCGU CEACAGUU AUGUACUCU AUGUAUUUU GUUUGUAUA GUUUURAAU 11300   AUMAGAGUU MUGAAUUC CAUGUAACGA AACMAUUUU GGACAUGUA CUNAGUUUU CUAACHUCU UAUACAUCUCU AUGUAUUAUA GUUUUAUA CUUACUCUA AUGUAUUUA CUAACHUCU UAUACAUCUCU UAUAUUUUUU 11300   AUMCUCUUUU MUGAAUUU CAUGUAGAUUUUU ACUUUGUAA AUGUAUUUU CUAAGUUUU CUAACHUCU UUAACHUCUUUUA GUUUUCUUUU AUUUUUUUU 11600   AUMACAUGU UUUAAGAUUUUUUA CGAUUUUUU AUAGUUUU CUAAGUUUU CUUAACHUCUUUUUUUUUUUUUUUUUUUUUUUUUU												
AUMAGAGAU MUMCHGAU CACAUAUUCU GGCAAUCCU AUGUCAUUU MUMCHGAUU CACUUCUUA CCUUCUAAU CUUCUAAU CUUCUAAU CUUCUAUUU CUUCUAUUU CUUCUAUUU CUUUUCUAU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUU CUUUUUUUUU CUUUUUUUUU CUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
AUGUCUAGU UNUGAMAUUC CALUCUUMA CCGACUCUA ACGUCACMA UNAGGUCGUU UACMACCUC UNAGUUGUS UNCAGAUGUA AUCUAGUUCUAN UGAUCUANAU GUCCAGUUCAN 11400   ACAMICUUMU UNACCUEGG GGCUURAGG UUGGANGCA CUGAGAUGU UCUAGUUUU UCUCUCUAN UUGAUCUANAU GUCCAGUUUUUUUUU 11500   MUGUCUAGU UNACCUCAG GGCUURAGG CUGGAUGUCA CUGAGUUUU CUAGUUAU NACCACUUG GAGGOGUAA AGUUGUANAU UUGAUGUUUA UUGAUGUUN 11600   MUGUCUAGU UUGAGUUGA AGGUUUAGG CCUUNAGUA GUAGACACA GAGANGUUA AUCUCUCUC UCUCUCUA GAUUUGUUUA AUUUUUUU 11900   UUGAAUAGU CUUCUGUUU ACCUUNAUU CCUCUUUUUU AUAUUCUCUCU AUAUUUCUUU 11900   UUGAAUAUAU CUUUGUUUU ACCUUNAUU CAUNUUUU AUAUUUCUCU AUAUUUCUU CUAUUUUU AAUUUCUUUU 11900   UUGAGUUUU GUGGUIAAUAU AUUUUUUUUU GUUUUUGAGU AUUUCUGUUU AAAUUUCUU CUUUUUUUU UUGAAUAUA AUUUCUUUUU UUUUUUUUUU												
ACAMUCUUUA UAACCUCUGG GECGUCAUGG UCGGAUGGA CUMAGUUUU CUAAGUAUU UCAACUAGA AMUCUUUGA AMUCUUUCUUA UMUKUUUM 11500   UUMGUUUUG GECCAUGGA UUCAAAGAU UUACCCCAA UUUGGAAL AACCAUUG GAGGGUAA GAGUCCAUU UUKURUUCUA UMUKUUUM 11600   AUBACAGAU UUAGAUGGA UUACACCCAA UUUGAGAA GACACACACAC CACAAUUUG CUUCUUCUU GAGUUGACUCAU UUKUUUUGAUUU GAUUCUUUGUUU GAGUUCUCUU GAGUUGACUCAU UUKUUUUGUUU GAUUCUUUU CUUCUUUUU GAGUUUUUCUU GAUUUUUUUUUU											1	
UNAGUICUUG GECCACUGA AUCAANAGAU CUNACECGAA GUCUGGGAAU ANGCACAUUG GAGGGECUNA GAGUUCCUAU UUGUGCUAU AUUUUCUUAU 11600   NAGGACAUC UNAGACAG GUCUUNAGGA GUUAGGAU MGCACACCA GAAAGUUGE UUGUCUCAU GUUCUCUUA GUUCUUAC GUCCUUAGA ACUUNCUUG 11900   UNAGUIGCU UURAGUUGA CAGUUCUU CUAGUUCUU GAGCAUCUCU UNAGGAUUG CUCCGUUACE CUCCUCUUAU GUUCUUACUCU 11900   UUCAGAGUU CUUUCUUUGU UUCUUUU UUCUUCUUA UUAGUUCUU AACUUCAUUC GUCCGUUA AUUUUUUCU CAGUUUUGU UUUUUUU CUUCUUUU GUUCUUUU UAUUUCUUUU GAUUUUUG AUUUUCUUUU AUUUUUUU CAUUUUUGUUUU AACUUAUUU CAUUUUGUUUU GAUUUUUGU UUUUUUUU CAUUUUUGUUUUG												
ABIGACGAUC UNAGACAGAC GUCUUNAGG CUUNAGMA UGACAGACA GAGANGUGC UCUALICAC UCUUCUCCUA GAGUUGACC GAGACAGAGA 11700   AUACCUGCCU CUUCAGUAG ACGCUUCUA CCCAULAUL UNAGACCAL UNAGANUGA ACCUCAULCE UCUCGUUAA UCUCUAUAA CUUCAGUACA CUUNCUUGA ACUUCAULA CUUCAGUAA CUUCCUUGA GUCCUACA ACUUNCUGA 11900   UUCAGACALE CUUUCUGUUGU AGCUUNUAC UCUUNUAA CUUACAGA AGUCACUCC AUCAGAACAG UNAUUCUUAA CUUCUGUUGU CUUUNAACCU AGUUCUUGU UUUNAACCU AGUUCUUGU UUUNAACCU AGUUCUUGU UUUNAACCU AGUUCUUGU UUUNAACCU AGUUCUUGU GUCUCUGUU GUCUAUACCU AGUUCUUGU GUCUCUGUU GUCUAUACCU ACAGUUCCU AUGUAUCACU AGUUCUUAA AUUUCUUAG GGAUNAAMAC 12200   CUUGAGGUG ACUULUUCU GAGUUUGGG UUUUUNAACCA AGUCAUUGU GACCUUGUG CUUAGAGCCU UUUGUAUCU GUCUUCUUGU CUUCUUGU CUUCUUGU UUUUAAUUU GAUUCUUUU GACUUUUGU GACUUUUGU GUCUUUUU GUCUUUUU GUCUUUUU GUCUUUUU GUCUUUUU GUCUUUUU AUUUAAUUUUMA AUUUCUUACA CUUCAGGAAG UUUUUUUAG CUUCUAAGAA UCUCUUMA AUUUCUAACCU GUUUUUUU GUUUUUU GUCUUUUUU GUCUUUUUU GUCUUUUUU GUUUUUU GUUUUUU GUUUUUU GUUUUUUU GUUUUUU												
AUACCUGGCC UUUCAGUAGG ACGGUUCCA CCGAGUUCCA CCGAGUUCCA CCGAGUUCCA CCGAGUUCCA CUAGUUCCU CUUUCCUCU CUUUCCUCUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
UUCAGAGAUC ACUCUCGGUC GCAAUCCUUU CCUCUCAUA GAUNACCCU UAAUCCUCC GAACAGUUA UACUAGUA CGUCAUCCU UGUAACCU AGUGGACUG 11900   UUCAGAGUC CUUUGGCCAC UUUUUGUUGU AGUUUAAUCU CAUNACACA AGUCACUCC AUGACACCC AGUUCCUC ACAGUUCUA AACGUCCAA 12000   AAUGCCUCU GGGUNABUAC CCGAUCUUUG AGUUNACUC AAUAACACA AGUCACUCA UAANUAGUU CCACAGUUCUA AACGUCCUA AACGUCUA AACGUCCUA AACGUCUUA AACGUCCUA AACGUCUUA AACGUCCUA AACGUUUUG CAUCAGUUUG CAUCAGUUCU AUAACUCC AAUGUUUUG CAUCAGUUUG AUUACACUC AGUUGAUACC CAUUNUGUU AUUACAUUUUU AUAACUCUA AUUUCUUUA GAUUCGAUA ACUUCUUU AUAACUUCA AUUAAUUU AUUUUUUUU												
CUUGAGAGUE CUUUUGGECAE UUUUUGGUUGU AGCUUUUAEU CAUAUACAEA AGUEARCUEG AUGEACAEGE AGAUUCEGUE UUUUAEACEU AGGUGAGUE 12000   AUGECECUEU GGUUAGUAE CEGAUEUUUG UGGUEUGGA AUGUECGAA AUCUECCAA AUCUECCUU AUAAUAACUU EAAGUEUEE AEACUUUUA AAGUECCAA 12100 12100   CUUCUCUEU GGGUUAGUA ECGAUEUUUG UGGACCECAU EAUUAUUUUA GAUUCUUUG GAAUUGUUUE CAUAGACEG AUUUCUUUA GAUUCGAEA 12100 12100   CUUAUUCUGU AACUUUU AUACAECUU CUUCUGEGA GAUUAUCUUU AUUUUUUU AUUCUUUA GAUUGACUEG AUUCCUUUA GAUUCGAAC UUUCUUAA CUUUUGAACA CUUUUUAAACUUUU AUUCUUUUGA AUUUGUUUE GAUUUGACCU GAUUUGAACA AUUUGAUCU 12100   CUUAUUUUUU GUACUUCUU CUUCUUUUUU GUUUUUUU AUUUUUUU AUUUUUUUU AUUUUUUUU												
ANJGECUEU GGUAUGUAC CCGAUCUUU UGGUCUGGA AUCUCCAGA AUCACCEU AUCUCCAGU CAAGUUUC AACGUCCAG 12100   CUICUCCUCU UGGUCUACIA AUGUACUUC AGCUUCGAG UUAGAUGACU AAUGUACUUU GAUCUUUG GAUCUUUC AGCUUCGAG 12100   CUAGUAGUUC AGCUUUCGAG UUUCUAGAG UUUUUUU GAUUUUU GAUUUUU AGUUUUU AGUUUUU AGUUUUU AGUUUUU AGUUUUUU AGUUUUUU AGUUUUUU AGUUUUUU AGUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUUU AGUUUUUUUU AGUUUUUUUUU AGUUUUUUUUU AGUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
CUUCCUCGUC UGGGGUAGU AUGUACAMA AUGAAGAC UGUUAUAUCU GACCUGUGC GAAUGUUUGC CUACAGGCCG AUAUGUUUC GACCAGUAAGA 12200   CUUGUGGGUS ACUACUUUC AGCUUUCGAG UGAGCCCAU ACAUUCUUUA GAUUCGUUG GGCGUUUCGC CCGGUAGGCC UAUGGAUACC ACUAUGCAC 12300   CCGGAUGCC UGACUCUU AUAGCACCUA CCUUUCAGGG GAAUAUCGG UUUGUUUGG AUUAGACUG AUUUCCUAG AUUUGUAGA CUACAGCAC AUAUAUGAC 12400   GUUDACUUU GAGCGUGAG UUUCUUCGU CCUUCGGGGA GAAUAUGAG UUUGUUCUG AUUAGACUG AUUUGCUAG GAUUCGACA UAUGACCAUAU 12500   GUUDACUUU GAGCGUGAG UUUCUUCGU CCUUUCAGCUU CUUAUAUA GAGCACAUA UCUUUAGGA CUUAGGCAC UUAGCACUA GGUUUCUUCAA GAUUGCAAU 12500   AUAUUUUCAA GAAUUCCUU CGUGGGUUU CUUAUAUA CUUCAUAUU UUUUAGGA CUUCAGACA UA CUUCUUAUA CUUGAUAUA AGUUCAGA CUUCUUCAA GGUUUCUA CAACUUUU UUUUACAUAU UUUCUUCAUA CUUCUUUUUU UUUUAUUACU CUUAUAUU UUUCUAUAU UUUCUUAUU UUUUAUUACUU AUGUUUCUU AGUUUCUUA AAUUUUUU UUUUCAAUU UUUUUAUUG CUUAUUUU UUUUAUUACUUAU CUUCAUUUUU UUUUAUUUU UUUUAUUU UUUUAUUU UUUUAUUU UUUUAUUUU UUUUUAUUU UUUUAUUUU UUUUAUUUU UUUUAUUUU UUUUAUUUU UUUUAUUUU UUUUAUUUU UUUUUU												
CUMAGUEGGUS ACUACUJUCC AGCUJUCGGG UUGAGUCCU ACUUCUJULA GAUUGGUUG CCGGUUGCG CAUUGAUACU AUUGAUACU AUUGUAUACU AUUGUAUACU AUUGUAUAU AUUGUAUAU AUUGUAUAU AUUGUAUAU AUUGUAUAU AUUGUAUUA AUUGUAUUA AUUGUAUUA AUUGUAUUA AUUGUAUUAU AUUGUUUUAU AUUGUUUUAU A												
CEGGAUGECE UARCUNCUU AUAGEACCUA CUUUCGGEGA AUUUUUUCUC AUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
AGUUGGAGGU GAUUAGAUGA AGUAUCCAAC UUUCUAUGCE GUUGGGUCUA CUUCAAGAGA UCACGUUGU AUAGACACAU GUUCAGCCAAG UUCAGCCAAG UAUUGUUAUA 12500   GUUUACUAUU GUACCGUGAG UUUCUUCUC CCGUGACUAU ACGUGAUUA GAGACAUAG UCGUCUAAUA CGAUUGACCC GAUUCAGCA AGUCCAACA AGUCAACAU AUUUCUCUA AUUUUUCUA GGAUUCUCUC GUUGAUUU ACGUAGAUG UACAUUAUU GCCACGUCAA CGUGGAAUCU AAUUUCUCUA UGUUUUUCUU UGUUUUUUA CUAGAUAUU GGACCAUGAG GUUCUUAC CGUGGAUCUC GAUAAAUACCU AUUUAUUUAA CAAUUCUUUA CUACAUAUU GCUCCUUGGU AUGACAUCA CCUGAAAUAC CAUUUUUU ACUUCAUUUUA CUACAUAUU GACCUCUGUGA CCUGAAAUAC CAUUCUAAU GUUCUUUU UUUUAAUU GCCCACGACAU AUUUUCUA CCUGGAUCUU AUGACAUGA CGUUUCUUAC CUUCGAAAUACCU AUCACAUUA I 13000   GGAAAUAAAA CGAGUUGCA AGUCCAAU GUUCAUAUGU UCUAAUUU UUCUAAUUU GUUCAUUGU AUUGACAUUU AUCAGUUU AUUAAUAUCA CUUAAUUUU UUUAAAUUUU ACUCCUUGE CAUUUCUUAU ACCCCUGUUCU 13100   GGAAAUAAAA CGAGUUGCA AGUCAGUUA GUCAAUUAU UACUGCUAAUU UUUCAAAUAUGU UUUAAAUUUU CUUAAUUAAA CAGUUUGU AUUCAAAUAU UUUCAAAUAU UUUCAAUAAUAU UUUCAAUAAAU CCAUUUUUU AAGUAUAUUU UUUCAAUAAA CGAGUUGUU AUUUAAUAAAA CGAGUUGUU AUUAAUAAUAU UUUUAUUAAA ACAGUUUU UUUAAUAAAA CGAGUUGUU UUUUAAUAAUAAU UUUUAUUAAA ACAGUUUUU UUUUAAAUAAUAU UUUUAUAAAAA CGAGUUUUU UUUUUUUU UUUUUUUUUU												
GUUUACUAUU GUACCGUGAG UUUUUUCGL CCUCAAGUU CUUAGUA CAUUCAUAU AGUUUACUAU CGUUAGUACUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUA AGUUUUUUA AGUUUUUA AGUUUUUUA AGUUUUUUA AGUUUUUUUA AGUUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUU AGUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
AUACUCUAUA UUCUUUCCA GGAAUCCCUU CGUGUACUCU UUCUUUUCA GGGGUUCCA GGUGUUCU AUACUCUAUA UUGUUUUUUA UUGUUUUUUUA GGGGUUUCU GUGUUAUUUU GGGGUUUCU GUGUUAUUUU GGGGUUCUU GUGUUAUUUU GUGUUUUUU GUGUUUUUUU GUGUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
GGGGUUCCA GGUGUUCUA UGUUUUUUA UUUUAUUUA CAACAUGUG GUUUCUACA CGUUUUUUA UUUAAUUUU CUUUAUUUU UUUAUUUU UUUUAUUUU UUUUAUUUU UUUUAUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUUU UUUUUUUUUU UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
UCCAGUCUCU ACAACUGU GUUCAACUGU ACUGAAUAAC CAGUCUACUA CUUCAAUAGU CUCGUUGUC AUAGACAUGA CGUUACUGCU AUCACUUAG 12900   UUACAGAGUU AUCUCU GUUGAAUUU UCUCUACUAG CGUGAUCAU UACUGCUACU ACAGUUGUC AUAGACAUGA CGUUACUGCA CUAAUAACA 13000   GGAAAUAAA CGAGUUGCA GCCCCAUAA GAUCAGUUU UCUACUAG CGUGAUGUU UACUGCUACU ACAGUUGU AUUCCUUUCUCUCUC CCUUCUUUAU ACCCCUGUAC 13100   AUCAGGCCU A AGAUUUCUC UGUUGAAUUU UCUCUACUAG CGUAAUUUU UCAAUAGA UUACCAACUA GGUUGGGUU UUAGACUUU CCUAAUAGA CCUGCUACA 13100   GCACCUUGGA CACUACCCG GAUUGGAGUU UUUGUAUAGUCUU UUCAAUAGA ACCGGGGAGA CAGACACUU UUAAACUUU CCUAAUAGA CCUGCUCACA 13100   GUUCCCCUA AGAUUUCU UGAAUAGA UUUGUUUA UGGUCUUA UUCUAUAGAA ACCGGGGAGA CAGACACUU UUAAAACCU UAGUAAAGA CGUGCUAACC 13300   GUUCCUAUAAU CAACUUCUA UGUGAGUCAA CUGACCGG GUUCUAAUCU UAGUACUU CCCACUU CCCACUUU UUAAAAAAAGA CCUGUUUAACC 13400   GUUCCUAUAA CCUAACCCG GUUCUAAUACU UAGUUACUU CAAUACUU UAGUACUUC CCGUCUUUA AUGGACAUC UACCAUAGGG UCUGUUAAUACUUUCAAU 13500   ACUACUGCCA GUUCCUAUAGA UCCUGGGA AAACCCCAA CUCUUCUUA CGUGUUUC CCGUCUUUA UCCCAUUUCU CACUAUAGG UUCUCUAUU 13700   AGGAAACCC UGUUGUUAA UCCUGGGGA AAACCCCCA CUCUUUUAU CCUUUUUA CCUGUUUCU CAGACAACU UACCAUAUG UUCUCAUAUU 13700   AUGCUGUCAA UUCCUGAGGA AAUCCUUCUAUA CCUUUUUUA CCUGUUCUC CCGUCUCUC UAUCUUUUU UACCAUAUUUU III 13700   CCUCUUUAUACCUA GGGACUUCA GUUCUUCUUUUUU CCUUUUUAU CGUUUUUC CGAA AUGUUUUU UUUUUUUU CCAAUUUU 13700											1	
UUACAGAGUU ANUCUAUCUC UGUUGAAUUU UCUCUACUAG CGUGAUCAUU UACUGCUACU ACAGUUGUG AACUAAUGAC UCAAUACCA CUAAUGACA ACUCUCUAUA GAUCACUUAG UCAAACUUU UACUGCUAU GAGUGAGAUG CGAAUUUGU AUCAAUACA CUAACUUACU AACUCUCUU AU ACCUUCUUU UCUAUUAG UCAAAUUU UCAGAAUACA UUACGAGAUA GAGUAGGUU UUAGAAGUUU GCUAAGACU UACGUCCACA 13200   GGAAAUAAAA CGAGUUGCA GCCCCCAUAA GAUCAGUUAG UCAAACGUAU GAGUGAGAUG CGAAUUUGU AUGACCUU CCUUCUUUAU ACCCCUGUCUU UUCUAUAACA UUACGAGAUA GAGUAGGUU UUAGAAGUUU GCUAAGACCU UACGUCCACA 13200   GCACCUUGGA CACAUACCCG GAUUGGAGAG UUUAGACUU UUCUAUGACA ACCGGGGAGA CACAACACUU AUAAGACUCU CUAGAUACC 13300   GUUCCCCCCA AUGGCGAACU CUAGAAAUUG ACACUGUUAU UGUGUUUCUA CCGGGGGGA CACAACACUU AUAAGAACCG UUCUGUAGAA CGUGUUGAACC 13300   GUUCCGCCCA UUCUAUACU UUAGUUUAUCUU GAGUAUCUU AGAGUAUCU UCCGACGAG GAAAGAACCG UUCUGUAAA CGUAUGAACG 13500   ACUACUGGC CAUGCCCA UGUCAGUCA CUGCACUGUUC UAGUUUCUUA CCGAGUUCU CCACGUGUAUUA UUUCUAAA CCUUAGGGC CAUGACUCA UUGCUCUAU GUUUUCUAA CCUAGGGU CCGUCUAUUA CUGCAAUGCG CCUUCAUUAU UUUCUAACACUUU UAUUUUU AUACCACUU UACGUGUCCA AUGGACCAU UUCUUAUA UUUCUAAC CCUUCUUUAU CUUCUAUAG UUCCUCAUUU UUCUUUUUAU ACCACUUG UUCUCAUUUUUU UACCUCCAG AGCCCCCGA AACCGUAUU UUCUAUUUUUUU UAUCCUCCAGA UUCCUCUGAG AAACCUUUCUUAU 13700   CGUGAACUUC AUUCCUCA UGUCGGGGA AAACCCUCA UCUCUUUUUU AUCCGUUUUCU AAUGCCCGAG AAACCUUUUUUUUUU												
GGAAAUAAA CGAGUUGEAA GCCCCAUAA GAUCAGUUAG UCAGUUAG UCAGUUAG UCAGUUAG GGAGAUUAGUUA UGAGAGUUAG UCAGGACUU GGUUGEAA GCUCUUAUA UGAGGGUGG GUCAAAUUU UCAGAAUAGC GUUAGGAGUU UUAGAAGUUU UUAGUUUU UUAGAAGUUU UUAGUUUUU UUUGUUUU AUAAGUGUUU UUUUUUAU UUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUU UUUUUUUU UUUUUUUU UUUUUUUU UUUUUUUU UUUUUUUUU UUUUUUUUU UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU												
AUCAGGECUA AGAAUUUCUA UGGAGGGUGG GUCAAAAUUU UCAGAAUAGA UUACGAGAUA GAGUAGGUU UUAGAAGUUU GCUAAGACCU UACGUCCAAA 13200 GCACCUUGGA CACAUACCCG GAUUGGAGAG UUUAGUCUA UUCUAUGAGA ACCGGGAGAG ACAGACACUU AUAAGACCC UAGAUAAGUA CGUGCUAACC GUUCCCCCAC AUGGCGAACU CUAGAAAUAG ACACUGUUAC UGGUCUACA CCGGCUGUA UCCUCCAGGA GAAAGAACCG UUCUGUAGAA CGUAUGGAU GUUCCCCCAC AUGGCGAACU CUAGAAAUAG ACACUGUUACU UGGUUCUAC CCGGCUGUA UCCUCCAGG GAAAGAACCG UUCUGUAGAA CGUAUGGAU CCUGUCAACCG UCUCUAUAGA UCCCUACCCG GUUCUAUUCU UAGUUACUUG AAGAGUCUCU CCGACCUCAG UGAUUUCUCA AUGGACCUUG AGUGAAAAA AUCAUGGGC CAUGACUCCA UGUCAGUCAA CUGCACAU CAGUAGUUUC AUAAGGGUA AUGAACUGG AUAUAGGCC UCAGUAAUA UUUUCACAU UCCUGUUCU CAUAUCCUCA GGACUUCAA AUCUUUUUA CCGAUGUUCUU AUGGACAUGG AUGAACAUG AUAUAGGCCU UCAGUAAUA UUUUCACAU AGGAAACCC UGUAGUCAA CUGCAGGAA AAACCCCCCAA UCCUCAUAG CUGCUAUUA CGUGCAAUU AUCCAUAGGC CGUCUUUAU GUUGUCUUAU 13700 AAGGAAACCC UGUAGUCAA CUGCAGGGA AAACCUGCCAUAGA CUCUCAUAGA UACAGUGGCU CAUGACAUG UUCCUCUUAU GUUGUCUUAU 13800 CUAUCCGUCU AGUCCACAC CAGACUCCAA UGUUAAUCUA CUCUCAUAGA UUCAGUGGCU CAUGACCGGA GACCGUGU UGUCUUAUU GUUGUCUUAU 13900 GUGAACUUG AUUGCUCA AUCUCGGGGA AUCAACUGU UCUAUUUCU AUCGAUUAC CCGUCUUCU CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 GAUGAACACC UAUACUUGU UACAAUCUGU UCUAUUUU AUCGAUUAC CCGUUCUUC UAGUAUUAU AUAGGACGAC UCCACCGUGA 14100 UCCCCUUUC UUUAUUUGU UACAAUCUG ACACCCAGUU UCUCAUUUCU AUCGAUAUAC CUGCUCUUC UAGUAUUAU AUAGGACGAC UCCACCGUGG 14100 UGUAACUUC UUUAUUUGU UACAAUCUCU UGACUAUCU CUCAUUUUC ACAUAAGUU CCUCUUUCUU CUUUAUCUU ACUACUUCA AUGACCUUCA 14200 UGGAAUGUC ACUUUAUGU UACAAUCAUC CAAUUGGU UUUCAAUUUU CUUUUAUAGA CCGUGUCUUA CUAACCUUA AUGACCUUCA 14200 UUUAACAUCU AUAGCUCU UAGGCAUAU ACACCCAGU UUUCUAACGU GGACGAUGU CUUCA CUAGGGUUCAA CAUGACCUUC 14400 UUUGACCUUUC UAUAUGCU UAGCACCUGU UACACCCGU CAAUAUUU UUUUUU CUUUUCUU UUUUAUCUU UUUCUAUCGU AUGAGGAUAG CUCCGGUGGG 14500 UUUGCAAUUGU AAUAGUCUU UACCAGAU ACACCCGCU UAAUUGAA AGUACCUUCU UUCAAUCUU UUCUAUCGU AGUGCUUCUA CUAAUGUCU 14600 UUUGACCUUUC AAACAUCUCUU AAACAUCAU CUAACCUCUU UUCUAACGU AGUUUUUU CUUUUUCUUUUC												
GCACCUUGA CACAUACCCG GAUUGAGAAG UUUAGUCUA UUCUAUGAA ACCGGAGAG ACAGACCUU AUAAAACCC UAGAUAAGUA CGUGCUAACCC 13300 GUUCCCCCAC AUGGCGAACU CUAGAAAUAG ACACUGUUAC UGGGUCUACA CCGGCUGUAC UCCUCCAGA GAAAGAACCG UUCUGUAGAA CGUAUGGAUA (13400 CGUCGAACCG UCUCUAUAGA UCCCUACCCG GUUCUAUCU UAGUUACUUG AGAGAUCUU CCGAGCUCAG UGAUUUCUCA AUGACCUUG AGUGUAAAGA (13500 ACUACUGGGC CAUGACUCCA UGUCAGUCAA CUGACCGAU CAGUAUGUUC AUAAGGUAG AUGAAACUG AUAUGCCCU UCUGUAGAA CGUAUGAAA (13500 UCCUGUUCUC CAUAUCCUCA GGACUUCAG AUGCACUGAU CAGUAUGUUC AUAAGGUAGA UGAAACUGG AUAUGGCCU UCUGUAGAUA UUUUCACAAU 13600 UCCUGUUCUC CAUAUCCUCA GGACUUCAG AUUUCUAA CCUUAGGCU CCUUAUGG UUCACAAUC UACCAUAGG CCUUCUUAU UUUUCACAU 13600 UCCUGUUCUC CAUAUCCUCA GGACUUCAG AAUCUUCUAA CCUUAGGGU CGUUCUUA CGUGACAAUC AUCCAUAGU UUCUCUUAU GUUGACUUAU AAGGAAACCC UGUAGUCAG AUCUCUGA AUGUUCUAA CCUUCUAUGG UUCACUAUG CAUGACGAGA GACCCUCAU UUCUAUGGU GUUCUCCACUU 13900 CUUUCCCUUUC UUUAAUUUCUA CAACUCGGA AUCAAUCUU UCUAUUGA CUUACAUAG CAGGACAGUU UUCUAUGGU UGUCUCACUU 13900 GUGAACUUG AGUCCACAAC CAACUCGGG AAUCAAUCUU UCUAUUGU AUGCGAUUAG CACUCUCU CAAUUUUAU AUAGGACGAC UCCACCGUGA 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAU CGACCUGUU UCUAUUUU AUCCGAUUAC CCUUAGGA CCGAGCUGUA CCUAACCCUU AUAUCUCAA 14200 CUCCGAUACUU AUUAGUUUU UCCAAUUGU UUCUAUUUU AUCCGAUUAC CUUUAUGA CCUUACCUUA												
GUUCCCCAC AUGGCGAACU CUAGAAAUAG ACACUGUUAC UGGGUCUACA CCGCCUGUAC UCCCAGGA GAAAGAACCG UUCUGUAGAA CGUAUGGAA CGUAUGAAUAG UUGUCAACU UAGUUACU UAGUUACU UAGUUACUU GAAGAGUCUU CGAACCG UUCUGUAGAA CGUAUGAAU UUGUCAAUU UAGUUACUU AAUAGGUUGU CGAACGG UAGUAGAA CGUAUGAAU UUGUCAAUU UAGUUACUUA AGAGAAUCU CAGUAGUACA UUGUCAACUUCUAAUCUU AAUAGGUUCU CGAACACUG UUCUGUUCUA AUGACCUGA UGUCCUAAUCUUAAU CGUACAAUAUUACUUAAUAGGC UUCAGUAGAU AUUUUCACAAU 13600 UUCUGUUUCU CAUAUCCUCA GGACUUCAG AAUCUUCUAA CCUAGGUUCUA AUAAGGGUAG AUGAAACUG AUGACUAGAG CCUCUUAUU GUUGUUAUU 13700 AAGGAAACCC UGUAUCCUCA GGACUUCAG AAUCUUCUAA CCUAGUUAUA CGUGACAAUC UACCAUAGGC CCUCUUAUU GUUCUCACUU 13800 CUAUCCGUUU AGUCCACAAC CAGACUGCAA UGUAAGGUA CUUCUAUGA UUCUCAUGG UUCACAAUC UACCAUGG GUUCUCUAGU GUUCUCACUU 13900 CGUGAACCC GGUACGAA UACUCGGGG AUGAACUGU UCUCAUAG AUGAGGUGU CGACCCGGG AAACCGUAGU UGUCUCAGUU GUCGCAACUUU 13900 GAUGAGAACC GGGUACGUAG UUGAUAUAU UGAGUCCCA UAUGAGAACA CUACAGUUAC CCGUUUAC CUAGCCCGG GUACGAAGUU GUCCACACUU GAUGAGAACC GGGUACGUAG UUGAUAUAUU UGAGUCCCA UAUGAGAACA CUACAGUUAC CCGUUUGCU CAAUUUAUUAU AUAGGACGAC UCCACCGUGA 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAU CGACUUUU UCUAAUGUU CUACAGUUAC CUGCUCUCU CAAUUUAUUAU AUAGAACGAC UCCACCGUGA 14100 UCGGUAUGUC ACAUUAGUCU UAAUGUCUUA UCGAAUUUC ACAAUAAGUU CCUCUUAAGG ACGUACCU ACUAGUUCAA CAUGACGUAC 14300 UCGGUAUGUC ACAUUAGGCU UAGGCAUUA CCGAUUAUC ACUACGUUAC CUUCUAUCGU AGUUCUAACCUAA CAUGACGUAC 14300 UUUAACCUUCU UUAAUUUGU UACCAACCCCCU AGCUCUCAA CACGAAUAUU CUUCUAUCGU AGUUCUAAC UAAGGUAGU CUCGGUGGG 14500 UUUGAAUGUC ACAUUAGCCU UCAAUGAGU UAUCACGAA CAGACCCCU ACCAAUGUU UUUGUAACCUAA AAGGUUCUU CAACUUAG AUUUAUCCUCU UUCAACCUAA AAGUAGUCU 14600 UUUGAAUGUC CAUUAGGCUUCUU ACCAACCCCCU ACGUCUCAAC ACUUGAAUUCUU UUUCAACCUAA AAUUACCUG GUCCUUACC UUAACAACCUAA 14700 AUUUCGAAUGU CUUCCUAACCCGAU CUAAUGGAU CAGAGGAGGG AACAGUUUU UUCUUAUCGU AGUUCUAUCU UUUAACCUAA AAUUACCUCU 14600 UUCCGUUUCG AACAUUUCUU AAACUCCU CAAUCGGA AACCUCCAA AGUUCCUU CUUCUUAUCGA AUGUUCUAUCU UUAAACAUCU GAAACCUCA 14700 AUUUCGAACGCUUAU AAUAUUCCC AAUAAUCCU AAUACGGU UUCAACCUAA AGUACCUCU AAUAUGACU AAAUUUCUA AAAUAU											1	
CGUCGAACCG UCULAUAGA UCCCUACCCG GUUCUAAUCU UAGUUACUU AGUACUU CGAAGUCUCU CCGACCUAG UGAUUUCUA AUGACCUUG AGUGUAAAGA 13500 ACUACUGGGC CAUGACUCCA UGUCAGUCAA CUGACCGGAU CAGUAGUUUC AUAAGGGUAG AUGAAACUG AUAAGGCCU UCAGUAGAUA UUUUCACAAU 13600 UCCUGUUCUC CAUAUCCUCA GGACUUCAG AAUCUUCUAA CCCUAAGGCU CGUCUAUUA GUGACAAUC UACCAUAGGC CGUCUUUAU GUUGUCUAAU 13700 AAGGAAACCC UGUAGUCUA UCUCGGGGAA AACCCCCAA CUCUCUAUGG UUCAGUGUC AUGACGACG GGCCCCAUG UUCUCUAUGU UUUUCACAAU 13800 CUAUCCGUCU AGUCCACAAC CAGACUGCAG AUGUUUCUAA CCCUAUGG UUCAGUGUC AUGACGCAG GGCCCCAUG UUCUCUAUGU UUUUCACAUU 13900 CGUGAACUU AGUCCACAAC CAGACUGCAG AUGAUCUGU UCUCAUUG UUCAGUGUC AUGACGCAG GGCCCCAUG UUCUCUUAUG UUUCUCAUU GGUGAACUU AGUCCACAAC CAGACUGCAG AUGAACCUG UCUCAUUGU UUCUAUGA UAGCCCGG AAACCGGAG ACACUACUUU 13900 GAUGGAACC GGUACCUUA UUCAAUAU UGAGUCCCA UUCUCUUUUU AUCCGAUUUC CCUGUUCU CUGAAUGUU CUCAUUGU 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAU UCUCAAUUUC AUCCGAUAUG CCGUCUUC CUAGUUUAUA AUAGCAGGAC UCACCCUUA CUCGGAACU AAACCUUACU UAAUGUCUUA UCGACUUU CUCAAUUUC ACCAAUAAUU CGUCCUUACCU CAUUUUAUA AUAGACGGAC CCUACCCCUU ACUACUCACA 14200 CUCGUAAUGU AAACCUUACU UAAUGUCUUA UCGAGUUU CUCAAUUUC ACAAUAAUU CGUCUUAAGGA CGGGUCGGA CCUACCCUUA ACUACUCACA 14300 UCGUAAUGUC ACAUUAGGCU UACCCAUUU UUCCAAUUUC ACAAUAAUU CGUCUUAACG AGGUUCUAAC UAAGGAUAG UCCGGUGGG 14400 CGGUGAAGUCU AUAACCUUA UGCACCUCUU UCCAAUUUC AUUGUAAUUU GUUCUAACG AGGUUCUAAC UAAGGAUAG UCCGUUCGG 14400 UCCGUUUCG AAAGCUUAU UAACCCUGU CACAUCGAU CAGGAGGG AAACAGUUUU CUUCUAUCGA AGGUUCUACCU ACAAAGAGUACU 14600 UUUUAGCUUCC AUGACUCCU UGCUUCUCU ACCACCCCCU ACCACCCCCU ACGAAUAUU CUUCUAUCGA AUGUCUUU UUCACCUAA AUUAUCUCU 14600 UCCGUUUCCG ACGACUUAC CAAUGACCC UUAACUUCU UCCUAACGU AGUUCUAUCU UUCAACCUAA AAUUACUCU 14600 14600 GUUAAGACCUUAC CAAUAUUUA ACCAGGU UUAUCACGA AGUGUUUUU CUUCUAUCGA AUGUUCUACUU GGAACGUAG AUUUCUCU 14900 GUUAAGACCUUAC CAAUAUUUA ACUACGUCUU AAAGAACAGG UGUUUUUG AAUAAUCGU AGUUCUAACUU GAAACUUCU 14600 14600 UUUUAGAAGU UUCCAAACUCCU AAAGAACCCU UUACUAUUU CCGAAUCUCU AAACUUCA AAGUACCUCU UAAUUACCUUCU AAACUUCU AAACUUCU AAACUUCU AAACUUCU AAAGAACCUCA 14700 1460												
ACUACUGGGC CAUGACUCCA UGUCAGUCAA CUGACCGGAU CAGUAGUUUC AUAGGGUAG AUGAAACUGG AUAUAGGCCU UCAGUAGAU UUUUCACAUAU 13600 UCCUGUUCUC CAUAUCCUCA GGACUUCAG AAUCUUCUAA CCCUAGGGCU CGUCUAUUA GUGACAAUC UACCAUAGGC CGUCUUUAU GUUGUCUUAU AAGGAAACCC UGUAGUCUA UCUCGGGGAA AAACCCCCAA CUCUCUUAUG UUCAUGGGU CAGGCCCGA GGCCCCAUG UUCUCUAUG GUUCUCACU 13800 CUAUCCGUCU AGUCCACAAC CAGACUGCAA UGGUAACUGA CCUUCUUAUG UUCAUGGGU CAGCCCGG AAACCGUGA UUCUCUUAUG GUUCUCACU 13800 CGUGAACUG AAUGGAUGA UCUCGGGGA AAACCCCCCA CUCUCUUAUG UUCAUUAU AUCCCGGA AAACCGCGG GAACCGUG UUCUCUUAU GUUCUUCACU GGUAACUG AAUGGAUGA UUCUCGGGG AAUCAACUGU UCUUAUUGU UAUCCGAUAUA AAUCCCCCG GAACCGGG GACGAAUCUG CAGCACCGG GAACCGUAGU GUCUUCACU GUGAACUUC AAUGAUGUA UUCUGAGUCAA UGUAACUGU UCUCAUUUUC UAUCGAUAUA AAUCCCCCGU CAUGUUAUAU AUAGACGAC GCGACCUGA CCACCCCGUA 14000 UCACCCUUUC UUUAAUUUGU UACAAUGAU GAGCCCCC UAUGAAACA CUACAGUUAC CGUCUCUCU CAAUUUUAU AUAGACGAC CCGCCCGUAA 14100 UCCCUUUCU UUUAAUUUGU UACAAUGAU GGAGCUAUC GGAUCAAGU GUCUCAUGU CUUCAAUGU CUCCAUUUCU AUACCUAA CAUGACGUAC CUCCGAAACU AAACCUUACU UAAUGUCUU UCGAGCUAUC GGAUCAAGU GUCUCUACC CUAGCUUCU CUUGUAACGAAC CCGGUCGCU ACUCUCUCAC UUUAGCUUUC UUUAUUUGU UACAAUGAU CGGACUAUC GGAUCAGGU GACCUGUAC CUCCCUCCU UAGUUAAA CAUGACGUAC UCCGUUUCU UUAAUUAGU UACAAUGAU CCGACUGU CUCUCAAUGUA CUCCUUUCUAUCGU AUUCACCUAAC UAGACGUUCU UUUAGCUUUC AAUAUUUCU UUAUGUUA UCGACCUAU CAGGAGGG AAACAGUUUU CUUCUAUCGU AGUUCUAAC AGAGGUUCU UUCAACCUAA CUUCGUUCUG UCCGUUUCG AAUGAUUCU UUCCAUGUCUU CACCUCUU CCUUCGAGA AGUGUUU CUUCUAUCGU AUGUCUAUCU UUCAACCUAG AUUAUUCUC UUUUAGCUUCCA AAGAUCUUUU CCAAUUGAU UACCAAUGUUUU CUUCUAUCGA AUGUUCUACUU GAACCUAGA AUUAUUCUC UACCUUUCG AAUAUUUA CUCCUUCUU AAGAAACAGG UGUUUUGA AUGACCUAGG AAGGUUCUU UUCAACCUAG AAUAUUUCU CAAUAGUUCU UUUACCUUCU AAGACACUCU UUCUAUCGA AGGACCUGU UUCAACCUAG AAUAUUUCU CAAUACUUCU AAGACACUCAA 11400 UCCUGUUUCCA AAGUAUUUU CAAAAACACGU UUAAUCUUG AUGACCUAUG AAUGUUUUG AAUAAUCUCU UUAQUUACCUG AAACCUGC AAACCUCAA 11400 UUUAGAUUUCCA AAGUAUUUU AAAUAUAG AUACAGGUGU UUUAACCUAG AAGGUAGUU UUAACCUAA AGUACUUCU AACCUAGA UUCUAACCUAA GUAGAACCUAA UUCUAAUUUU CAAAUAUUUU												
UCCUGUUCUC CAUAUCCUCA GGACUUCAG AAUCUUCUAA CCCUAGGCU CCGUCUAUUA CGUGACAAUC UACCAUAGCG CCGUCUUUAU GUUGUCUAU 13700 AAGAAACCC UGUAGUCUGA UCUCGGGAA AAACCCCCAA CUCUCUAUGG UUCAGUGUC AUGACGCAGA GGCCCCCAUG UUCUCUAUG GUUCUCCACU 13800 CUAUCCGUCU AGUCCACAC CAGACUGGAA UGGUAAGCUA CCUUCUAUGA UUCUAUGG UCACCCCG GAAACGUGU UGUCAUGAUC GACGAACUUU 13900 CGUGAACUUG AUGCACAC CAGACUGGGA AUCAACUGU UCCUAUUGC AUGACGAUA AAUCCCCUUC CUCGACCCCG GUACGAAAG ACAAUACUGC 14000 GAUGAGAACC GGUACGUAG UUGAUAAUAU UGAGUCCCC UAUGAGAACA CUACAGUUA ACUCCCUUC CUCGACCCCG GUACGAAAG ACAAUACUGC 14000 GAUGAGAACC GGUACGUAG UUGAUAAUAU UGAGUCCCC UAUGAGAACA CUACAGUUG CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCACCCUUUC UUUAAUUUGU UAAAUAUU UAGUCCCCA UAUGAGACA CUACAGUUG CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCGUGAAACUU AAACCUUACU UAUGAGUCUA UCGACUUUC CGAAUGUG ACCUGUCCUC CAAUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCGUGAAAUGUC AAAUGAUCU UAUGAGUCUUA UCGACUUUC CGUACUUGC ACAAUAAGUU GCUCUUCUU CUUAUGGA CCAAGUUCUA ACAUGAUGAA 14300 CGUGUAAUGUC AAUAUCUCU UAUGACCUAU GCACCCCCU AGCUCUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CCUGGCCUAA CUAGGUCCGU UCGUAAUGUC AAUAUCUCCU GUCGUUCUGU CAAUUGGAU UAUCACCAAU UUUGUAAUU CGUGUUUAUACGU AGUUCUUACCU UUUCACCUG AAUUAUCUCU 14600 UCCGGUUUG AAUAUCUCCU GUCGUUCUGU CACAUCGAU AUUCCGAA ACGAAUAUU CGUCUAUCGU AUGUCUAUCU UUUCACCUG AAUUAUCUCU 14600 UUUAGACUGU AUAACUCUCU GUCGUUCUGU CACAUCGAU CAGAGAGGG AAACGUUUU CUUCUAUCGU AGUUCUAUCU UUUCACCUG AAUUAUCUCU 14600 UCCGUUUGGAA AUAUCUCCU GUCGUUCU AAGAACAGG UGUGUUGGA UUGAACAUU GUUCUAUCGU AUGUAUGUU UUCACCUG AAUUAUCUCU 14600 UUUAGACCUUU AAACAUCUCUU AAUAAACAGG UGUAUUGU AUCGACUUU UUCAACCUU AUCGAACUUCU UUCACCUGA AUUUUCUA AACUUGGUUUG AACAUUUU AAUACACCU AUAAAACAGG UGUUUGGA UGGACUUCU AAGGAAUGU CUUAAUCUU GUAACAUUCUU AAUAUUCUCU AAACAUUCU AAUAUGACU UUAACCUUCU AUACCAUUCU AAACAUUCUU AAUAUUCUU AAUAUUCUU AAUAUUUU CCGAAUUUU CCGAAUUUCU AAACAUUCU AAUAUUCU AAUAUUCU AAUAUUCU AAUAUUUCU AAAUAUUGU UUAUUAUAUAUU UUAUUCUAACUUCU AAACUUCU AAUAUUCU AAUAUUUU CCGAAUUUU GAAAUUUU GAAAUUCUU UUCACAUAUGGUCUU CAAACUUCU AAUAUUUU CAAAUUUUU AAUAUUGACU												
AAGGAAACCC UGUAGUCUGA UCUCGGGGAA AAACCCCCAA CUCUCUAUGG UUCAGUGUCC AUGACGCAGA GGCCCCCAU UUCUCUAUG UUCUCCCUU 13800 CUAUCCGUCU AGUCCACAAC CAGACUGCAA UGGUAAGCUA CCUUCUAUGA UUCUAUGG UGACUCCGAG AAACCGUAGU UGUCUAUGAU GACGAACUU 13900 CGUGAACUUG AUGGAUGGA UAACUCGGGG AAUCAACUGU UCCUAUUGU AUCCGAUAUA AAUCCCCUUC CUCAACCGUG GUACGAAGG ACAAUACUGC 14000 GAUGAGAACC GGUACGUAG UUCAUAUAUAU UGAGUCCCCA UAUGAGAACA CUACAGUUAC CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCACCCUUUC UUUAAUUUGU UAAUAUAUA UGAGUCCCCA UAUGAGAACA CUGAAUAUGU CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCGCCUUUC UUUAAUUUGU UAAUAUAU UGAGUCCCCA UAUGAGAACA CUACAGUUAC CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCGUCAAUGUC UUAAUUUGU UAAUGUCUUA UCGACUUCCAUUUC ACAAUAAGUU CCCUCUCCU AAUUAUAUA AUAGGACGAC UCCACCGUGA 14100 UCGUAAUGUC UUAAUUUGU UAAUGUCUUA UCGACUUUC CGGAUCAGGU GACAUGGU CCUCUCUCU AGUUUUAUCCU ACUAUUCAA CAUGACGUAC UCCGAAUCUUA UAAACUCUUA UGACCCUGCU ACCUCUGCAA CACGAAUAUU CGUUCUAACGA GGGUCCGAC CCUGGCCUAA CUGGUCCGU UCGUAAUGUC ACAUUAGGCC UAGCCCCUCU AGCUCUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CCGGGCCUAA CUGGUCCGGU UUUAGACUGU AUAGCUCUUA UGACCUCUCU CCAUUGGAU UAUCACCAAU UUUGUAACUU GGUCCUAACG AGGUCCGAC CGUGCCUAA CUGGUCCGG UUUAGACUGU AUAACUCUCU GUCCUUGU CACAAUCGAU AUACACCAAU UUUGUAACUU GGACGAAGAGG UGUCUUAACCU UUUCACCUUG AAUUAUCUCU 14600 UCCGUUUGC AGUGCUUAC CAAUGAGCC UUAACUCUU UCCUUCGAA AGUAGUCCU AGAAUUGU CUCUAUCCU AUGUACUCU UUCAACCUU GAAACGUACU 14600 UGUUAGACCUU AUACACCUCU UAAGAACCAG UGGUACUUGU AUGGACUAUG UGUGUUGACG UACUAUCGU UUCAACCUU GAAACCUCU 14600 UGUUAGAACG UUACCCCAUCUUAUAUA ACUCGUCUU AAGAACAGG UGGUAUUGU UUGAACAUUGGA UAGACAUUC UAUCACCUUC AAAGUACUU UUCAAACAUUCUU AACCUUCUAACCUU AUAUAUCACCUUCAAACGUACUU AAACAUUCUU AAUAUAUUCUU AAACAUCUUG AUGACCUUC AAACUUCUU AACCAUAUUGGAAUAGUU CUAAACUUC UUAAUCAUUU CAAACUUCU AAAGAACUU UUAAUAUUUCCAA AACUACUU UAAUCUUGAAUAGAG GAUUGUUU GAAAUAUUUG AAAUAAUUCUU UAAUAUAUUCUU AAACAUUCUU AAAGAACCUU UUAAUCUUCUAAAGA AACUACUU UAAUUUUC CGAAAUUCUU AAUAUGGACUU GAAACUUCU UAAUGGAGUACUU UAAUAUUUCUAA AACUACUUUU AAUAUUUUC AAACUUCUU AAUAUUUU CAAAAC												
CUAUCCGUCU AGUCCACAAC CAGACUGCAA UGGUAAGCUA CCUUCUAUGA AUAGAGUGGU CGACUCCGAG AAACCGUAGU UGUCAUGAUC GACGAACUUU 13900 CGUGAACUUG AAUGGAUGGA UAACUCGGGG AAUCAACUGU UCCUAUUUCU AUCCGAUAUA AAUCCCCUUC CUCAACUUUAUAU AUAGGACGAC UCCACGUGG AAUGAGAACC GGGUACGUAG UUGAUAAUAU UGAGUCCCCA UAUGAGAACA CUACAGUUAC CCGUCUCUC CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 GAUGAGAACC GGGUACGUAG UUGAUAAUAU UGAGUCCCCA UAUGAGAACA CUACAGUUAC CCGUCUCUCU CAAUUUAUAU AUAGGACGAC UCCACCGUGA 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAU AGACCCAGUU UCUCAAUUUC ACAAUAAGUU GCCCUUAGGA CCGAGCUGUA CCUAACCCUU AUAUACCUCAC 14200 UCGUAAUGUC ACAUUAGGCC UAGCGCAUAG ACACCCCCCU AGCUCUGCA CACAGUUAC CUCCCUCUCU CAGUUUAUCCU ACUACUCACA CAUGACGUAC UCGUAAUGUC ACAUUAGGCC UAGCGCAUAG ACCACCCCCU AGCUCUGCA CACGAAUAUU CGUUCUAACG AGGUCCGAC CCGUGGCUAA CCUGGUCGGI 14400 UGUAAUGUC ACAUUAGGCC UAGCGCAUGA ACACCCCCCU AGCUCUGCAA CACGAAUAUU CGUUCUAACG AGGUCCUAC UAGAGGAUAG CUCCGGUGGGG 14500 UUUQACCUGU AUAAGCUCU UAUGCCU CCAAUUGGAU UAUCACGAACAAU UUUGUAGAUU GGGACGAGG UGCUCUAACC UAGAGGAUAG CUCCGGUGGGG 14600 UUCCGUUUCCG AGUGCUUACC CAAUGAGCC UUAACUGCU UUUCUAACGAA UUUGUAACCUG AGUUCUAACCUU GUACACCUAG AAUAAUCUCU 14600 UCCGUUUGGA ACAUAUUUA ACUCGUCUCU ACAAUGGAU UUUCUAAGGAAUUU UUUGUAUCU AGUUCUAACCUU GAAACACUGGA 14600 UUUGAACCUUU AAAGAUCCU UUAAUGACCU UUCUUUCUU UCCUUAUCGU AGUUCUAACCUU GGACGUCU GCAAACCUCA 14600 UUGUAGAACG UUACCCCAAUGAGCC UUAACUCUU UUCUUUU CCGAAUUUG UUGACCUUG AGUACUUU GUGACGUCU GCAAACUCCA AAUAAUCUU 14600 UGUUAAAGAUC UUACUGAACACC GAUAAACCU AUAAAUAGA UUGAUCUUU AUGACCUU CUAACCUU AUGUGAGAU UUCAAACAUUCUUU CAAACUUCU 14900 GUUAAACAUCU UUACUGAACAC GAUAAACCU AUAAAUAGA GAUCCCCAAUAUCAA AGUAUUUG AAUAAUCUG UUCAAACUUC UUAAUAGACCUUCU AAUACACUU UAACCCUUAU CAAAGUAAUU UUAUAAAUAGA GAUCCCCAAU GUUUUUU GAAAAUUCGU UAAUAUGUCU UAAUAGUACUU UUAAUACUUCU AAAGGUAGU UUAAACAUCUU AAUAUUCUU AAAGGUAGU UUAAAUAUUCUU AAAGAACCUU UUAAUAUUUU AAAUUUUU AAUUUUU AAAUAUUG AAUAUUCUU AAAGGAACUUC AAACCUCAA AGUACUUCU AAACCUUCU AAAGGUAUUU UAAAUAUUGU UAAUAUUCUU AAUAUUGUAA AACUACUUUU AAUAUUCUU AAAUAUUGA AGGACUUCUA AGGACCUUU UAAUAUAUUUU CAAAAUAUUCUU AAUA												1 3800
CGUGAACUUG AAUGGAUGGA UAACUCGGGG AAUCAACUGU UCCUAUUUCU AUCCGAUAUA AAUCCCCUU CUCAACCCG GUACGAAAGG ACAAUACUGC 14000 GAUGAGAACC GGUACGUAG UUGAUAAAUAU UGAGUCCCA UAUGAGAACA CUACAGUUAC CCGUCUCU CAAUUUAUAU AUAGGACGAC UCCACGUGA 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAU AGACCCAGUU UCUCAAUUUC ACAAUAAGUU GCCCUUAGGA CCGAGCUGU CCUAACCCUU ACUACUCACA 14200 UCCGGAAACU AAACCUUACU UAAUGUCUUA UCGAACUGU CUCAAUUUC ACAAUAAGUU GCCCUUAGGA CCGAGCUGUA CCUAACCCUU ACUACUCACA 14200 UCGGAAACUU AUACCUUACU UAAUGUCUUA UCGAACUUU CUCAAUUUC ACAAUAAGUU GCCCUUAGGA CCGAGCUGUA CCUAACCCUU ACUACUCACA 14200 UCGGAAUGUC ACAUUAGGCC UACCCUCCU AGCUCUCCAA CACGAAUAUU CUUCUAUCGA CAUGAUGUCU ACUAGUUCAA CAUGACGUUA CGAGUCGGAU AUAACUCUU UAGCCUGGU CCAAUUGGAU UAUCACGAAU UUUGUAGUUAC GGACGAGG UGUCUUACC UAGAGGAUAG CUCGGUGGG 14400 UUUGACCUCU AUAUGUCCU GUCGUUCUGU CAACUUGGAU UAUCACGAAU UUUGUAGUU GGGACGAGG UGUCUUACC UAGAGAUAG CUCGGUGGG 14500 UUUGGUUUG AAUAUCUCCU GUCGUUCUGU CAAUUGGAU UAUCACGAA UUUUGUAGUUAGU AGUUCUAUCU UUUCACCUAG AAUAAUCUCU 14600 UCCGUUUCCG AGUGCUUACC CAAUGACCCC UUAACUGUU CUCCUCACAA AGUAGUUCU CUUCUAUCGU AGUUCUAUCU UUUAACCUAG AAUAAUCUCU 14600 UUUGAACAGC UUACCCGAUC GUCAUUCUGU UCCUUCUCU UCCUUCAAC AGUAGUUGU CUGCAUUCU UUCACCUAG AAUUAUCUCU 14900 GUUAAAGAUC UUACUGAACAC GAUAGAACCU AUAGAAUGA GUGGUACUUG UUGACCAUU AUGUGAGUU CUGAACCUCU GCAAACUUCU 14900 GUUAAAGAUC UUCGAAACA GAUAGAACCU AUAGAAUGA GUACCAGUU UUAACCAU AUGUGGAC UAGGACAUCU CUAAAUGCG UUCAACUUCU 14900 GUUAAAGAUC UUCGAAACA GAUAGAACCU AUAGAAUGA UACAGGUGU UCUAACCUU AACCGAAAUCUG UUCAAACUUC UUAAUAGAG UUCAAACUUCU 14900 UAACCCUUAU CAAAGUAAUU UUAUUUCUAA AACUACUUCU AAACUGGA UUCUAACCUUCU AUACUGAACU CUAAUUGAAUAUGAC 15100 UAACCUUAUG AAUAGUUCUU UUAUUUCUAAAAUAG AUACAGCUCCAAU AGGACCUGU CUCAAACUC UUCAAACUUCU UAAUUGAGA CUCCUAAUUCUG AAUAUCUAAUGGUUCU UUAAUUUCUAA AACUACUUCUU AAAUUCCUG AAAGCCCGUUCUAAUGGG CAUUAUUUUC CAAACUCCA 15200 UACCUAAUGA UCCUAAUAG CUCGGAUUAUUUU AACUUCUU AAAUUACAG GGAUUUCUUCUAAAGGGC CUUCUAACCUU CAAACCUCC AISS00												13900
GAUGAGAACC GGGUACGUAG UUGAUAAUAU UGAGUCCCCA UAUGAGAACA CUACAGUUAC CCGUCUCUC CAAUUUAUAU AUAGACGAC UCCACGUGA 14100 UCACCCUUUC UUUAAUUUGU UACAAUGAUC AGACCCAGUU UCUCAAUUUC ACAAUAAGUU GCCCUUAGGA CCGAGCUGUA CCUAACCCUA ACUCCCCGUGA CUCCGAAACCU AAACCUUACU UAAUGUCUUA UCGACCUAUC GGAUCAGGU GACACUGUAC CUCCCUCUC UAGUAUUCCU ACUACUUAA CAUGACGUAC UCGUAAUGUC ACAUUAGGCC UAGGCCUUA UCGACCUCU CGAUCAGGU GACACUGUAC CUCCCCUCUC UAGUAUUCCU ACUAGUUCAA CAUGACGUAC CAAUGUCGAAUAUGC ACAUUAGGCC UAGGCCUUA CCAACCCCCU ACCUCUCUCAA CACGAAUAUU CUUCUAUCGA GGGUCCUAAC ACUGGUCCUA CAUGAGGAUG CGAGUCGGAU AUAACUCUA UGACCCUGCU CCAAUUGGAU UAUCACGAAU UUUGUAGAUU GGGACGAGG GGUCUCUAAC AAGAGGAUG CCUCGGUGGGG 14500 UUUAGCUUG AUAUUCUCU GUCGUUCUGU CCAAUGGAU CAGAGGAGG AAACAGUUUU CUUCUAUCGU AGUUCUAUCU UUUAACCUAA AAUAACUCU UUUAGCUUUCC ACUGUUCUGU CCAAUGACCC UUAAUUCUCU UUCUGACCUAA AUAAGGUUGU CUUCAAUCAG AAUUAUCUCU GCAAUCUGU CUAACCUAG AAUUAUCUCU UCCGUUUGCG AGUGCUUACC CAAUGACCC UUAACUCUCU UCCUUCCAGA AAGUAGUUCU CUUCUAUCGA AUGUCUAUCU UUAACCUAG AAUUAUCUCU GUUAGAAGAC UUACCCCGAUC GAAGAACCCU AAAGAACAGG UGGUACUUGU AUGACCUAG AAUGUUGUU CGGAAUGUGU CGGACGUCU GCAAACCUCA 14600 UGUUAGAAGA CUUCCUGAACA GAUGAACCU AUAGAACAGG UGUAUUUU CUUCAACCAUG AGUGUGUUC GGAACGUCU GCAAACUUCU 14900 GUUAAAGAUC UUCCGAACAC GAUAGAACCU AUAGAAUAG AUACAGGUGU UCUAACCAUU CAUAUGGAC UACUAUCGA AGUUGUCU GCAAACUUCU 14900 GUUAAAGAUC UUCCAAACA GAUAGAACCU AUAGAAUAG AUACAGGUGU UCUAACCAUU CAUAUGGAC UUCAAACUUC UUCAAACUUCU 14900 GUUAAGAUC UUCCAAACA GAUAGAACCU AUAGAAUUGA UACAAUGUGU UCUAACCUU AAGGACAUUC AUACUGAACUUCU UAAUAUGAC 15100 UACCUUAUG AAUAUUCUUA AACUACUUCUUA AACUCCCG UCCAAUUCGA AAGCCCCAUU GAUAAUUGU UUAAUAUUGU UACUAUAUUGC 15100 UACUAUAUGAU UCCAAACUUCUUA AACUACUUCUU AAAUUACAG GAUUGCACGU UUCAACCUCU UUAAUAUGU UAAUUUUCUAAACCUACUUCU UACUUAUGGUUCU UUAAUUUCUAA AACUACUUCUU AAACUUCU UAAUUGUGACUAUUUGC												14000
CUCCGANACU ANACCUUACU UAAUGUCUUA UCGAGCUAUC CGGAUCAGGU GACACUGUAC CUCCUCCUC UAGUAUUCCU ACUAGUUCAA CAUGACGUAC 14300   UCGUAAUGUC ACAUUAGGCC UAGCCCAUAG ACCACCUCU AGUCUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CGUGCCUAA CCUGGUCCGU 14400   CGAGUCGAU AUAGACUCU UGACCCUGC UCAUUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CGUGCCUAA CCUGGUCCGU 14400   UCGUUAUCU AUAGACUCU UGACCCUGCU CCAUUGGAA UAUCACCAAU UUUGUACAUU GGACGAAGA UGUUCUAAC UAGAGAUAG UCCCUGCGG AUAGACUGGU CUCUAAUUUCU UUUCACCUAA UAGAGAUAG UCCCUGGGG 14500 14600   UUUUAGACUGU AUAACCUGU UGCUUCUGU CAAUGAU UACCAAU UUUGUACAUU GGACGACGA UGUUCUAAUUU UUUCACCUAA AAUAGAUGU UCCACUUGA AUUAUUCUU UCCUUCUACCU AGAAUCGGA UAGAAUCGU UCCACUUGA AUUAUUCUCU UUUCACCUAA AUUAUCUCU UUCACCUAA AUUAUUCUCU UUCUACCUAA AUUAUUCUCU UUCUACCUAA AUUAUUCUCU UACUUUCU UCCUUCAACUU CUUCUACCUU ACAUACGAA AUUAUUUU UCCUUUUU UCCUUUUGAAUUUU UUCUUUU UUCUUUU UCCUUUUU ACAUAUUU ACUUUU AAUAUUUU UCGAAUUUUU UUCUUUU UUUUACUUUU UUCUUAUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUU AUAUUUUU UUUUU AUAUUUUU UUAAUUUUU UUAAUUUUU UUAUUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU AAUAUAUUUU AAUAUUGAAUUGU UUAACCUUUU UAAUUUUUU UAAAUUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU AAUAUUGU UUAAUUUUU AAUAUUUUU AAUAUUUUU AAUAUUUU AAUAUUUUU AAUAUUUU UAAUUUUU UAAUUUUU UUAUUUUU UUUUUU												
CUCCGANACU ANACCUUACU UAAUGUCUUA UCGAGCUAUC CGGAUCAGGU GACACUGUAC CUCCUCCUC UAGUAUUCCU ACUAGUUCAA CAUGACGUAC 14300   UCGUAAUGUC ACAUUAGGCC UAGCCCAUAG ACCACCUCU AGUCUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CGUGCCUAA CCUGGUCCGU 14400   CGAGUCGAU AUAGACUCU UGACCCUGC UCAUUGCAA CACGAAUAUU CGUUCUAACG AGGUCCGAC CGUGCCUAA CCUGGUCCGU 14400   UCGUUAUCU AUAGACUCU UGACCCUGCU CCAUUGGAA UAUCACCAAU UUUGUACAUU GGACGAAGA UGUUCUAAC UAGAGAUAG UCCCUGCGG AUAGACUGGU CUCUAAUUUCU UUUCACCUAA UAGAGAUAG UCCCUGGGG 14500 14600   UUUUAGACUGU AUAACCUGU UGCUUCUGU CAAUGAU UACCAAU UUUGUACAUU GGACGACGA UGUUCUAAUUU UUUCACCUAA AAUAGAUGU UCCACUUGA AUUAUUCUU UCCUUCUACCU AGAAUCGGA UAGAAUCGU UCCACUUGA AUUAUUCUCU UUUCACCUAA AUUAUCUCU UUCACCUAA AUUAUUCUCU UUCUACCUAA AUUAUUCUCU UUCUACCUAA AUUAUUCUCU UACUUUCU UCCUUCAACUU CUUCUACCUU ACAUACGAA AUUAUUUU UCCUUUUU UCCUUUUGAAUUUU UUCUUUU UUCUUUU UCCUUUUU ACAUAUUU ACUUUU AAUAUUUU UCGAAUUUUU UUCUUUU UUUUACUUUU UUCUUAUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUUU AUAUUUUU UUUU AUAUUUUU UUUUU AUAUUUUU UUAAUUUUU UUAAUUUUU UUAUUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU AAUAUAUUUU AAUAUUGAAUUGU UUAACCUUUU UAAUUUUUU UAAAUUUUU UUAAUUUUU UUAAUUUUU UUAAUUUUU AAUAUUGU UUAAUUUUU AAUAUUUUU AAUAUUUUU AAUAUUUU AAUAUUUUU AAUAUUUU UAAUUUUU UAAUUUUU UUAUUUUU UUUUUU		UCACCCUUUC	UUUAAUUUGU	UACAAUGAUC	AGACCCAGUU	UCUCAAUUUC	ACAAUAAGUU	GCCCUUAGGA	CCGAGCUGUA	CCUAACCCUU	ACUACUCACA	14200
CGAGUCGAU AUAGACUCUA UGACCCUGCU CCAAUUGGAU UAUCACGAAU UUUGUAGAUU GGGACGAAGG UGUCUCUAAA UAGAGAUAG CUCCCUGGGG 14500 UUUAGACUGU AAUAUCUCCU GUCGUUCUGU CACAAUCGAU CAGAGGAGGG AAACAGUUUU CUUCUAUCGU AGUUCUAUCU UUUCACCUAG AAUUAUCUCU 1 4600 UCCGUUUUCG AGUGCUUACC CAAUGAGCCC UUAACUCUU UCCUUCGAGA AGUAGUCCU ACGAAUCUGG AAUGUUCUAUCU UUUCACCUAG AAUUAUUCU ACUUGGUUUG AACAUAUUUA ACUCGUCUCU AAAGAACAGG UGGUACUUGU AUCGACUAUG GUGUUUGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU A 4800 UGUUAGAACG UUACCCCAUC UUAUUGACU AAAGAACAGG UGGUACUUGU AUCGACUAUG GUGUUUGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU 4900 GUUAAAGAUC UUACUCAAACA GAUAGAACUC AUGAAAUAG AUACAGUGU UCUAACCAUU GACCAUUC AUAGGACA UAGGACACUC UUCAAACUUC UUCAACUUCCU 14900 UAACCCUUAU CAAAGUAAUA GUAGGAACU UUAAGAAUAG AUACAGGUGU UCUAACCAUU GACCAAUUG GACCAAUUC UUCAACUUC UUCAACUUCCU UUCAACUUC UUCAACUUCU AAACUUCUUCUA AACUACUUCU UAUUUCUAA AACUACUUCU UAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAUUUCUAA AACUACUUCU UAUUUCUAA AACUACUUCU UAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAAUUUCUAA AACUACUUCU UAAUUUCUA AACUUCU UAUUUUCUAACUUCU UAUUUUCUAACUUCU UAUUUUCUAACUUCUAUAUCAG GAUUCUCCCG UUCUACUUCUUCUAUAUUU UAUUUUCUAACUUCUU UAUUUUCUAACUUCUU UAUUUUCUAACUUCUU UAUUUCUA AACUCUUCUACUUCUUCUAUUCUUCUUCUAUUCUUCUAUAUGAG GAUUUUUGCUAACUUCUUCUUCUUUCUUCUUCUUCUAACUUCUUCUUC												14300
UUUAGACUGU AAUAUCUCCU GUGGUUCUGU CACAAUCGAU CAGAGGAGGG AAACAGUUUU CUUCUAUCGU AGUUCUAUCU UUUCACCUAG AAUAUAUCUCU 14600 UCCGUUUUCG AGUGCUUACC CAAUGAGCCC UUAACUCUU UUCUUUGAGA AGUAGUCCU ACGAAUCUGG AAUGUUGUG UGGACGUCU GCAAACCGAA 14700 ACUUGGUUUG AACAUAUUUA ACUCGUCUCU AAAGAACAGG UGGUACUUGU AUCGACUAUG UGUGUUGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU A 14800 UGUUAGAACG UUACCCCAUC UUAUUGACUC AGUCUAUUUU CCGAAUUUG AUGACCAUCU AUACGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU 14900 GUUAAAGAUC UUCGAACAC GAUAGAACCU AUAGAAUAG AUACAGUGU UCUAACCAU GACCAGUA GGACUGUC UUCAACUUC UUCAACUUC UUAUGACUC AUACUAUUG AUACACUUC AAACACUCAU GACCAGUA GGACUGUC UUCAACUUC UUAAUGAGACG UUCGAACUUC UUCAACUUCU UUAAUGAGUAUU UAAGAGUUCUUG AAUAACAUCUG UCCAAACUCC UUCAACUUC UUCAACAUUC UUCAACAUUC UUAAUGAUCUU UAAUGUCUUG AAUAACUUCU AAAGAACUG UUCUAACAUUCU UAAUGUCUUG AAUAUAUUGU UAAUGUGUCUG AAUAAUUCU UAAUGUCUUG AAUAAUUCU UAAUGUCUUG AAUAAUUCU UAAUGUCUU UAAUGUCUUG AAUAACUUCU UAAUGUCUUCU AAAGAUCUG UUCAACAUUCU UAAUGUCUUCUA AACUACUUCU AAAGUUCUUCU AAAGUUCUU UAUGUGCUGG CACUAAUUGCC UUACCUAAGG AGUGGUUUCU UUAUUUUCUAA AACUACUUCU AAAUAUCCCG UUCGUACACU UUCUAACGU UUCUAACGUUCUUCU UAUGUGCUGG CACUAACUCAC 15200 UACCUAUAGG UUCCUAUAUG CUCGGAUUUCUUA AUAUAUCAG GGAUUGCACG UUUUUGCUAACGU UUCGAGCCG CAUGGACCUU CAAACCCCA 15300		UCGUAAUGUC	ACAUUAGGCC	UAGCGCAUAG	ACCACCCCCU	AGCUCUGCAA	CACGAAUAUU	CGUUCUAACG	AGGGUCCGAC	CCGUGCCUAA	CCUGGUCCGU	14400
UUUAGACUGU AAUAUCUCCU GUGGUUCUGU CACAAUCGAU CAGAGGAGGG AAACAGUUUU CUUCUAUCGU AGUUCUAUCU UUUCACCUAG AAUAUAUCUCU 14600 UCCGUUUUCG AGUGCUUACC CAAUGAGCCC UUAACUCUU UUCUUUGAGA AGUAGUCCU ACGAAUCUGG AAUGUUGUG UGGACGUCU GCAAACCGAA 14700 ACUUGGUUUG AACAUAUUUA ACUCGUCUCU AAAGAACAGG UGGUACUUGU AUCGACUAUG UGUGUUGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU A 14800 UGUUAGAACG UUACCCCAUC UUAUUGACUC AGUCUAUUUU CCGAAUUUG AUGACCAUCU AUACGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU 14900 GUUAAAGAUC UUCGAACAC GAUAGAACCU AUAGAAUAG AUACAGUGU UCUAACCAU GACCAGUA GGACUGUC UUCAACUUC UUCAACUUC UUAUGACUC AUACUAUUG AUACACUUC AAACACUCAU GACCAGUA GGACUGUC UUCAACUUC UUAAUGAGACG UUCGAACUUC UUCAACUUCU UUAAUGAGUAUU UAAGAGUUCUUG AAUAACAUCUG UCCAAACUCC UUCAACUUC UUCAACAUUC UUCAACAUUC UUAAUGAUCUU UAAUGUCUUG AAUAACUUCU AAAGAACUG UUCUAACAUUCU UAAUGUCUUG AAUAUAUUGU UAAUGUGUCUG AAUAAUUCU UAAUGUCUUG AAUAAUUCU UAAUGUCUUG AAUAAUUCU UAAUGUCUU UAAUGUCUUG AAUAACUUCU UAAUGUCUUCU AAAGAUCUG UUCAACAUUCU UAAUGUCUUCUA AACUACUUCU AAAGUUCUUCU AAAGUUCUU UAUGUGCUGG CACUAAUUGCC UUACCUAAGG AGUGGUUUCU UUAUUUUCUAA AACUACUUCU AAAUAUCCCG UUCGUACACU UUCUAACGU UUCUAACGUUCUUCU UAUGUGCUGG CACUAACUCAC 15200 UACCUAUAGG UUCCUAUAUG CUCGGAUUUCUUA AUAUAUCAG GGAUUGCACG UUUUUGCUAACGU UUCGAGCCG CAUGGACCUU CAAACCCCA 15300		CGAGUCGGAU	AUAGACUCUA	UGACCCUGCU	CCAAUUGGAU	UAUCACGAAU	UUUGUAGAUU	GGGACGAAGG	UGUCUCUACA	UAGAGGAUAG	CUCCEUGEGE	14500
UCCGUUUCCG AGUGCUUACC CAAUGAGCCC UUAACUCUCU UCCUUCGAGA AGUAGUCCCU ACGAAUCUGG AAUGGUAGUU CGUGACGUCU GCAAACCGAA 14700 ACUUGGUUUG AACAUAUUUA ACUCGUCUCU AAAGAACAGG UGGUACUUGU AUCGACUAUG UGUGUUGACG UACUAUCGAA AGUUGUCCCA AAACUUCCU 14800 UGUUAGAACC UUACCCGAUC UUAAUGACUC AGUCUAUUUU CCGAAUUUG UUGACCAUUC AUACUGGAC UACGACACUC UCUAAGUCCC GUUCAACUUCU 14900 GUUAAACAUC UUCUGAACAC GAUAGAACCU AUAGAAUAG AUACAGUGU UCUAACCAUU GACCAGUA GGACUGGUC UUCAACUUC GUUCGAACU UAACCCUUAU CAAAGUAAUA GUAGGCACU UUAGUUCUUG GACUACCAUU GACCAGUC AAAGUACUGU UUCAACCUUC GUUCGAACU UAACCCUUAU CAAAGUAAUA GUAGGCACU UUAGUUCUUG GACUACCAUU GACCAGUG GCCAAUUCUG UUCAACUUC UUCAACCUUC GUUCGAACU AUACCUUAU CAAAGUAAUU UUAUUUCUAA AACUACUU <u>CU AAAUUCCCCG UCAUCUUCA AAGCCCCGGU CCUUUAUUUU UUAUUGUGCUGG GACUAACUUCU</u> UAAUCUAAGGA CUCGGUUUCU UUAUUUCUAA AACUACUUCU AAAUUCCCG UCGUCUGACGU CUCAACUUCU UAUGUGCUGG GACUAACUUCU UACUAUGA UCCACUAUAG CUCGGUUUCUU UAUUUCUAA											1	14600
ACUUGGUUUG AACAUAUUUA ACUCGUCUCU AAAGAACAG UGGUACUUGU AUGACUAUG UGUGUUGACG UACUAUGAA AGUUGUCCA AAACUUCCUA UGUUAGAAGC UUACCCGAUC UUAUUGACUC AGUCUAUUUU CCGAAUUUG UUGACCAUUC AUACUGGACA UAGGACACUC UCUAAGUCCG UUCAACUUCU UGUAAAGAUC UUCUGAACAC GAUAGAACUC AUAGAAUAG AUACAGGUGU UCUAACAUU GACCAGUA GGACUGGUC UUCAACUUC GUUCUAAU UAACCCUUAU CAAAGUAAUA GUAGGCACU UUAGUUCUUA GACUCCCA UAGUUUUUUG AAAUAUCUG UCCAAACUCC UUCUAAUUUG GUUCUAAUAUGA UACAUUCU AUACCCUUAU CAAAGUAAUA GUAGGCACU UUAUUCUAA AACUACUCCCA UAGUUUUUUG AAAUAUGUG UCCAAACUCC UUCUAAUUUCUAAUAUUGU UAAUGUCUUG GACUCCCAU AGUUUUUUG AAAUAUGUG UCCAAACUCC UUCUAAUUAUUCUA AUAUCUAAGG AGUGGUUUCU UUAUUUCUAA AACUACUUCU AAAUCCCCG UCAGUUUUAUG AACUAUUCUG UCAAACUCC UUCUACUAUUGUGCUGG GACUAGUUCU UACUAUGA UCCACUUAUAG CUCGGUUUCUU UUAUUUCUAA AACUACUUCU UAUUGUCCGG CAUGGACCUU CAGAACCCCG 15300												14700
UGUUAGAAGC UUACCCGAUC UUAUUGACUC AGUCUAUUUU CCGAAUUUU AUGACCAUU AUACUGGACA UAGACACUC UCUAAGUCCG UUCAACUUCU 14900 GUUAAAGAUC UUCUGAACAC GAUAGAACU AUAGAAAUAG AUACAGUGU UCUAACAUU GACCCAGUA GGACUGGU UUCAACUUC GUUCUGAAGU 15000 UAACCCUUAU CAAAGUAAUA GUAGGGCACU UUAGUCCUUG GACUCCCAAU AGUGUUUUG AAAUAAUCUG UCCAAACUC UUAAACUUC GUUCUGAAGU 15100 AUAUCUAAGG AGUGUUUCU UUAUUUCUAA AACUACUU <u>U AAAAUCCCG UCAGUUCUAC AAGCCCGGU CGUUUUACU UAUGUGCUGG CACUAACUUC</u> 15200 UACCUAAGG AUCACUAUAG CUCGGUAUAC UGUCGAGGU UAUUAAUCAG GGAUAGCACG UCUUGCUAC AUCAGGCCG CAUGAACCUG CAGAACCUG 15300												14800
GUUAAAGAUC UUCUGAACAC GAUAGAACUA AUAGAAAUAG AUACAGGUGU UCUAACAUU GACCCAGUAA GGGACUGGUC UUCAAACUUC GUUCUGAAGU 15000   UAACCCUUAU CAAAGUAAUA GUAGGGCACU UUAGUCCUUG GACUCCCAAU AGUGUUUUG AAAUAAUCUG UCCAAACUUC UUUAUUAUAUAUGA 15100   AUAUCUAAG AGUGUUUCU UUAUUUCUAA AACUACUUCU AAAAUCCCG UCAGUUCUAC AAGCCCCGGU CGUUUUACU UUAUGGCUGG CACUAACUAC 15200   UACCUAAGG AGUGUUUCU UUAUUUCUAA AACUACUUCU AAAAUCCCG UCAGUUCUAC AAGCCCCGGU CGUUUUACU UUAUGGCUGG CACUAACUAC 15200   UACCUAAGG AUCCACUAUAG CUCGGUAUUA QUUCUUAA AACUACUGU UAUUAAUCAG GGAUAGCACG UCUUGCUACC UUCGAGCCCG CAUGAACCUG A 15300		UGUUAGAAGC	UUACCCGAUC	UUAUUGACUC	AGUCUAUUUU	CCGAAUUUGA	UUGACCAUUC	AUACUGGACA	UAGGACACUC	UCUAAGUCCG	UUCAACUUCU	14900
UAACCCUUAU CAAAGUAAUA GUAGGGCACU UUAGUCCUUG GACUCCCAAU AGUGUUUUG AAAUAAUCUG UCCAAACUCC UAUAAUAUGU AUCAUAUUGC 15100 AUAUCUAAGG AGUGUUUCU UUAUUUCUAA AACUACUU <u>CU AAAAUCCCCG UCAGUUCUAC AAGCCCCGGU CGUUUUACU UAUGUGCUGG CACUAACUAC</u> 15200 UACCUAGGA UCCACUAUAG CUCGGUAUAA UGUCGAGGU UAUUAAUCAG GGAUAGCACG UCUUGCUAGC UUCGAGGCGC CAUGGACCUU CAGAACCUGA 15300												15000
UACCUAGUA UCCACUAUAG CUCGGUAUAC UGUCGAGCAU UAUUAAUCAG GGAUAGCACG UCUUGCUAGC UUCGAGGCGC CAUGGACCUU CAGAACCUGA 15300												15100
		AUAUCUAAGG	AGUGGUUUCU	UUAUUUCUAA	AACUACUU <u>CU</u>	AAAAUCCCCG	UCAGUUCUAC	AAGCCCCGGU	CCGUUUUACU	UAUGUGCUGG	CACUAACUAC	15200
ACAGGUAUAC UGUUAUC <u>AUU<sup>K</sup>CUUUUU</u> GAAU GUUCUUCUGU UCUUUUAAAU UUUCCUAUGU AUAGAGAAUU UGAGAACAGA CCA 15383	I										CAGAACCUGA	15300
		ACAGGUAUAC	UGUUAUCAUU	CUUUUUGAAU	GUUCUUCUGU	UCUUUUAAAU	UUUCCUAUGU	AUAGAGAAUU	UGAGAACAGA	CCA		15383

Fig. 2. The RNA sequence of the Sendai virus genome (Z strain) from nucleotide position 4,781 to 15,383, the 5' end of the genome. R1 and R2 are the repeating consensus sequences. Op-4, op-5 and op-6 denote the large open reading frames.

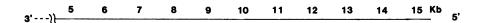


Fig. 3. Distribution of translation termination codons (vertical bars) in the plus strand RNA of the genome region presented in Fig. 2. An arrow indicates the direction of translation.

namely a cDNA fragment correponding to the 3' end portion of op-6 (from 8,540 to 9,719), which was prepared from clone 3S-222 by digestion with HindIII, followed by labeling <u>in vitro</u> by nick-translation. This probe hybridized with the largest virus specific poly(A) RNA as shown in Fig. 4. From these observations, we concluded that op-5 and op-6 encode the HN and L proteins, respectively.

No open reading frame was detected within the sequence of the 5' terminal 54 nucleotides following -GAA after the last gene, and the 3' half of this sequence was found to be U-rich and the 5' half to be A-rich. Thus, this region may be considered as the 5' leader region as reported for vesicular stomatitis virus (VSV). It is interesting to note that the most 5' terminal 12 nucleotides of this region are complementary (including one wobbling base pair) to the most 3' terminal 12 nucleotides of the genome,

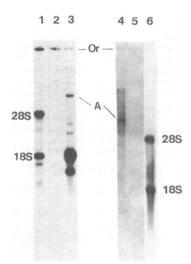


Fig.4. Left: Agarose gel electrophoresis of <sup>32</sup>P-labeled mRNAs from Sendai virus infected BHK-21 cells (lane 3) and those from uninfected cells (lane 2). Right: Hybridization of <sup>32</sup>P-labeled cDNA corresponding to the 3' end portion of OP-6 (nucleotide position 8,540 to 9,721) with mRNAs from infected cells (lane 4) and uninfected cells (lane 5). 32Plabeled ribosomal RNA from uninfected cells served as size markers (lanes 1 and 6). "A" indicates the largest mRNA. "Or" means the origin.

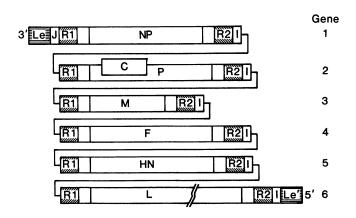


Fig. 5. A schematic illustration of the entire Sendai virus genome RNA. Le and Le' indicate the 3' leader region and the putative 5' leader region, respectively. I is the sequence, GAA or GGG, and J the sequence, AAAA.

which is in good agreement with a previous observation by Re et al.(24). This observation suggests the possibility that both ends of the genome RNA construct a stable secondary structure, giving the genome RNA a panhandle structure.

Combining the present results with the previous ones (3,4), we present

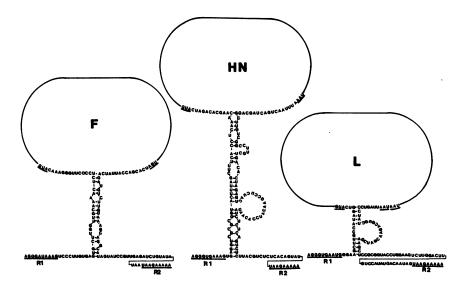


Fig. 6. Proposed secondary structures of the mRNAs for the F, HN and L proteins. The initiation and termination codons are underlined. R1 and R2 are the consensus sequences.

F protein 10 20 30 40 50 60 70 80 MTAYIQRSQC ISTSLLVVLT TLVSCQIPRD RLSNIGVIVD EGKSLKIAGS HESRYIVLSL VPGVDFENGC GTAQVIQYKS 80 LUNRLLIPLE DALDLOEALI TVINDITIONA GAPOSEFFGA VIGTIALGVA TSAQITAGIA LAEAREAKED IALIKESMIK 160 THKSIELLON AVGEQILALK TLOOFVNDEI KPAISELGCE TAALRLGIKL TOHYSELLTA FGSNFGTIGE KSLTLOALSS 240 LYSANTIDEIM TTIRTGOSNI SOVIYTEOIK GTVIDVOLER YMVTLSVKIP ILSEVPGVLI HKASSISYNI DGEEWYVTVP 320 SHILSRASFL GGADITDCVE SRLTYICPRD PAQLIPDSQQ KCILGDTTRC PVTKVVDSLI PKFAFVNGGV VANCIASTCT 400 CGTGRRPISQ DRSKGVVFLT HDNCGLIGVN GVELYANRRG HDATWGVQNL TVGPAIAIRP VDISLNLADA TNFLODSKAE 480 LEKARKILSE VGRWYNSRET VITIIVVMVV ILVVIIVIII VLYRLRRSML MGNPDDRIPR DTYTLEPKIR HMYTNGGPDA 560 MAEKR (565) HN protein 10 20 30 40 50 60 70 80 MDGDRGKRDS YWSTSPSGST TKPASGWERS SKADTWLLIL SFTOWALSIA TVIICIIISA ROGYSMKEYS MTVEALINMSE 80 REVKESLTSL IRQEVIARAV NIQSSVQTGI PVLLNKNSRD VIQMIDKSCS RQELTOHCES TIAVHHADGI APLEPHSPWR 160 CPVGEPYLSS DPEISLLPGP SLLSGSTTIS GCVRLPSLSI GEAIYAYSSN LITQGCADIG KSYQVLQLGY ISLNSDMPPD 240 LNPVVSHTYD INDNRKSCSV VATGTRGYQL CSMPTVDERT DYSSDGIEDL VLDVLDLKGR TKSHRYRNSE VDLDHPFSAL 320 YPSVGNGIAT EGSLIFLGYG GLTTPLQGDT KCRTQGCQQV SQDTCNEALK ITWLGGKQVV SVIIQVNDYL SERPKIRVTT 400 IPITONYLGA EGRLLKLGDR VYIYTRSSGW HSOLOIGVLD VSHPLTINNT PHEALSRPGN KECNWYNKCP KECISGVYTD 480 AYPLSPDAAN VATVTLYANT SRVNPTIMYS NTTNIINMLR IKDVQLEAAY TTTSCITHFG KGYCFHIIEI NQKSLNTLQP 560 MLFKTSIPKL CKAES (575) L protein 10 20 40 30 50 60 70 80 MDGQESSONP SDILYPECHL NSPIVRGKIA OLHVLLDVNO PYRLKDDSII NITKHKIRNG GLSPROIKIR SLGKALORTI 80 KDLDRYTFEP YPTYSQELLR LDIPEICDKI RSVFAVSDRL TRELSSGFOD LWLNIFKOLG NIEGREGYDP LODIGTIPEI 160 TOKYSRNRWY RPFLTWFSIK YDMRWMOKTR PGGPLDTSNS HNLLECKSYT LVTYGDLVMI LNKLTLTGYI LTPELVLMYC 240 DVVEGRWNMS AAGHLDKKSI GITSKGEELW ELVDSLFSSL GEEIYNVIAL LEPLSLALIQ LNDPVIPLRG AFMRHVLTEL 320 QTVLTSRDVY TDAEADTIVE SLLAIFHGTS IDEKAEIFSF FRTFGHPSLE AVTAADKVRA HMYAOKAIKL KTLYECHAVF 400 CTILINGYRE RHGGOWPPCD FPDHVCLELR NAOGSNTAIS YECAVDNYTS FIGFKFRKFI EPOLDEDLTI YMKDKALSPR 480 KEAWDSVYPD SNLYYKAPES EETRRLIEVE INDENENPEE INNYVESGOW LKDEEFNISY SLKEKEIKOE GRLFAKMTYK 560 MRAVQVLAET LLAKGIGELF RENGMVKGEI DLLKRLTTLS VSGVPRTDSV YNNSKSSEKR NEGMENKNSG GYWDEKKRSR 640 HEFKATDSST DGYETLSCFL TTDLKKYCLN WRFESTALFG ORCNEIFGFK TFFNWMHPVL ERCTIYVGDP YCPVADRMHR 720 QLQDHADSGI FIHNPRGGIE GYCOKLWTLI SISAIHLAAV RVGVRVSAMV OGDNOAIAVT SRVPVAOTYK OKKNHVYEEI 800 TKYFGALRHV MFDVGHELKL NETIISSKMF VYSKRIYYDG KILPQCLKAL TKCVFWSETL VDENRSACSN ISTSIAKAIE 880 NGYSPILGYC IALYKTCQQV CISLGMTINP TISPTVRDQY FKGKNWLRCA VLIPANVGGF NYMSTSRCFV RNIGDPAVAA 960 LADLKRFIRA DLLDKQVLYR VMNQEPGDSS FLDWASDPYS CNLPHSQSIT TIIKNITARS VLQESPNPLL SGLFTETSGE 1040 EDLNLASFLM DRKVILPRVA HEILGNSLTG VREAIAGMLD TTKSLVRASV RKGGLSYGIL RRLVNYDLLQ YETLTRTLRK 1120 PVKDNIEYEY MCSVELAVGL ROKMWIHLTY GRPIHGLETP DPLELLRGIF IEGSEVCKLC RSEGADPIYT WFYLPDNIDL 1200 DTLTNGCPAI RIPYFGSATD ERSEAQLGYV RNLSKPAKAA IRIAMVYTWA YGTDEISWME AALIAQTRAN LSLENLKLLT 1280 PVSTSTNLSH RLKDTATQMK FSSATLVRAS RFITISNDNM ALKEAGESKD TNLVYQQIML TGLSLFEFNM RYKKGSLGKP 1360 LILHLHLNNG CCIMESPQEA NIPPRSTLDL EITQENNKLI YDPDPLKDVD LELFSKVRDV VHTVDMTYWS DDEVIRATSI CTAMTIADTM SQLDRDNLKE MIALVNDDDV NSLITEFMVI DVPLFCSTFG GILVNQFAYS LYGLNIRGRE EIWGHVVRIL 1520 KDTSHAVLKV LSNALSHPKI FKRFWNAGVV EPVYGPNLSN ODKILLALSV CEYSVDLFMH DWOGGVPLEI FICDNDPDVA 1600 DMRRSSFLAR HLAYLCSLAE ISRDGPRLES MNSLERLESL KSYLELTFLD DPVLRYSOLT GLVIKVFPST LTYIRKSSIK 1680 VLRTRGIGVP EVLEDWDPEA DNALLDGIAA EIQQNIPLGH QTRAPFWGLR VSKSQVLRLR GYKEITRGEI GRSGVGLTLP 1760 FDGRYLSHQL RLFGINSTSC LKALELTYLL SPLVDKDKDR LYLGEGAGAM LSCYDATLGP CINYYNSGVY SCDVNGQREL 1840

NIYPAEVALV GKKLNNVTSL GQRVKVLFNG NPGSTWIGND ECEALIWNEL QNSSIGLVHC DMEGGDHKDD QVVLHEHYSV

RNLRVITKTL LDRFEDIIHS ITYRFLTKEI KILMKILGAV KMFGARONEY TTVIDDGSLG DIEPYDSS

IRIAYLVGDR DVVLISKIAP RLGTDWTRQL SLYLRYWDEV NLIVLKTSNP ASTEMYLLSR HPKSDIIEDS KTVLASLLPL 2000 SKEDSIKIEK WILIEKAKAH EWVTRELREG SSSSGMLRPY HQALQTFGFE PNLYKLSRDF LSTMNIADTH NCMIAFNRVL 2080 KDTIFEWARI TESDKRLKLT GKYDLYPVRD SGKLKTISRR LVLSWISLSM STRLVTGSFP DQKFEARLQL GIVSLSSREI 2160

1920

(2228)

here the primary structure of the entire Sendai virus genome, which is schematically illustrated in Fig. 5. As a whole, 99.17% of the Sendai virus genome is transcribed into mRNA and 93.63% is translated into proteins, indicating that the structure of the genome is utilized quite efficiently.

It is noteworthy that an open reading frame corresponding to 249 amino acids was detected in the genome sense strand within the L gene region (from nucleotide position 9,588 to 8,842), which is longer than that of the Sendai virus C protein, and the only one long open reading frame capable of coding for more than 150 amino acids in the genome strand. However, this frame is not flanked by R1 and R2, nor could a single stranded cDNA probe complementary to this open reading frame detect any subgenomic transcript from the infected cells, either poly (A) plus or poly (A) minus (data not shown). From these observations, it seems unlikely that this open reading frame is transcribed, although further studies should be carried out before a definite conclusion is drawn.

### Proposed structures of mRNAs for the F, HN and L proteins

From the nucleotide sequence of the genome RNA, the nucleotide sequences of mRNAs for the F, HN and L proteins were deduced. On detailed examination of these sequences, it was found that in every mRNA, a part of the 5' noncoding sequence was complementary to that of the 3' noncoding sequence, suggesting that these ends might form a double-stranded structure, which gives the mRNA a panhandle structure (Fig. 6). These secondary structures seems to be fairly stable on the basis of the free energy levels (30).

### Characteristics of the F, HN and L gene products

The amino acid sequences of the F, HN and L gene products are shown in Fig. 7. The deduced amino acid sequence of the F protein indicates that the F protein is highly hydrophobic overall. As we reported previously (4), a signal peptide of 24 to 27 amino acids was detected in its N terminus, and the cleavage site for the F1 and F2 proteins was assumed to be the arginine residue at position 116. The most hydrophobic region of this protein is located near its C terminus (from amino acid position 500 to 523), which is followed by a hydrophilic region of 42 amino acids. The F protein of Sendai virus has been reported to penetrate the viral envelope, leaving an at least 3K portion exposed to the inside of the envelope (31), which is similar to

Fig. 7. The predicted amino acid sequences of the F, HN and L proteins. The putative N-linked carbohydrate attachment sites are boxed. The underlining indicates the most hydrophobic regions.

# **Nucleic Acids Research**

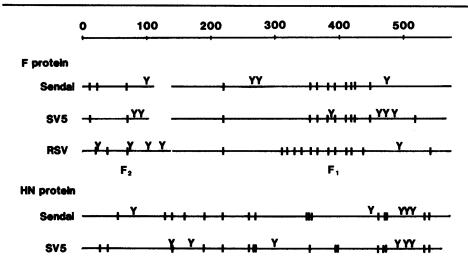


Fig. 8. Locations of the cysteine residues (vertical bars) and the putative N-linked carbohydrate attachment sites (Y) within the F proteins of Sendai virus, simian virus 5 (SV5) and respiratory syncytial virus (RSV), and the HN proteins of Sendai virus and SV5. To align the N termini of the F1 subunits, small gaps are left between the F2 and F1 subunits of Sendai virus and SV5.

in the case of hemagglutinin of influenza virus (32). Thus, it is probable that the hydrophobic region near the C terminus may anchor the F protein within the viral envelope and the very C terminal hydrophilic region may be located at the inner surface of the viral envelope, interacting with the viral M protein. As shown in Fig.8, the F protein has 12 cysteine residues, of which 2 are found in the signal peptide, 1 in the F2 portion and 9 in the F1 portion. Eight of the 9 cysteine residues in the F1 portion are clustered in its middle part, from position 338 to 424. Since the results of secondary structure analysis (33) indicated many reverse turns within the cysteine-rich portion of the F1 protein, this cluster may stabilize the tertiary structure of the F protein by forming intramolecular disulphide bonds. We could detect four putative N-linked carbohydrate attachment sites (Asn-X-Ser or Asn-X-Thr) (34) in the F protein, i.e. from position 104 to 106, 245 to 247, 259 to 261, and 449 to 451, respectively, which is in good agreement with the results reported by Kohama et al. (35) showing that the F protein of Sendai virus has four N-linked carbohydrate chains.

The most hydrophobic region of the HN protein is, in contrast to in the F protein, located near the N terminus, i.e. from position 36 to 58, which is preceded by a hydrophilic domain of 35 amino acids. This structure

resembles that of influenza virus neuraminidase (36,37), in which the hydrophobic region near the N terminus serves as an signal for membrane translocation and as an anchor in the membrane. Thus, it is plausible that this hydrophobic region near the N terminus of the HN protein may act as a signal and an anchor, and the preceding hydrophilic region of 35 amino acids is exposed to the inside of the viral envelope, since the HN protein was reported to penetrate the viral envelope and the molecular weight of the portion exposed to the inside of the envelope was estimated to be at least 2K (31). Within the deduced amino acid sequence of the HN protein, we found five N-linked carbohydrate attachment sites, i.e. 77 to 79, 448 to 450, 499 to 501, 504 to 506, and 511 to 513, which is in accordance with previous suggestion that the HN protein of Sendai virus has at least four N-linked carbohydrate chains (35).

Analyses of the deduced amino acid sequence of the L gene product provided us with little special information as to the structure of the L protein or its functional domains, except that Ser-Asp-Asp found from position 1,429 to 1,431 or Leu-Asp-Asp from 1,649 to 1,651 might be the active site of RNA synthesis as in the case of reverse transcriptases of retroviruses and RNA polymerases of picornaviruses (38). It is interesting to note that there is a possibility that another L gene product might be present which starts from the sixth AUG codon in mRNA, corresponding to the methionine at position 249 of the L protein or nucleotide position 9,299 to 9,301 with a calculated molecular weight of 224,005, since this codon is preceded by A 3 nucleotides upstream while the first five AUG codons are preceded by G or U, and according to Kozak (39), A might be preferentially recognized by ribosomes.

# Comparison of the Z and Harris strains as to the F and HN genes

Recently, Blumberg et al. (8) reported the complete nucleotide sequence of the F gene of the Harris strain of Sendai virus as well as the sequence of the tripeptide of the N terminus of the F2 protein purified from virions of the same strain. On comparison of the results presented in this paper with theirs neither insertion nor deletion could be detected, and 98.63% of the nucleotides and 97.88% of the amino acids were found to be conserved between these two strains. The observation that the N terminus of the F2 protein of the Harris strain was glutamic acid at amino acid position 26 is in good agreement with our previous prediction as to the cleavage site of the signal peptide (4), thus we concluded that the peptide from 1 to 25 is the signal peptide and that from 26 to 116 is the F2 protein.

# **Nucleic Acids Research**

During preparation of this manuscript, the nucleotide sequences of the HNN genes of the Harris strain (40) and Z strain (41) of Sendai virus were also presented. Comparing our present results for the Z strain with those for the Harris strain, we found that the latter is longer than the former by one amino acid residue, namely, the genome of the latter strain has an additional -AAA- between nucleotide positions 8,288 and 8,289 of that of the former, which led to an insertion of a serine residue. This insertion seems to make no significant difference as to the structure of the HNN protein between these strains. Except for this insertion, most of the nucleotide sequence as well as the amino acid sequence was conserved between these two strains, giving 98.52% and 97.57% homology, respectively. On comparison of our results and those presented by Miura et al. (41) on the HN protein of the Z strain, four substitutions of amino acid residues were found while neither insertion nor deletion was detected.

#### Comparison of Sendai virus and SV5 as to the F and HN proteins

The nucleotide sequences of the F and HN genes of SV5, another paramyxovirus, were also reported recently (12,13). When the amino acid sequence of the F protein of the Sendai virus Z strain described above was aligned with that of SV5 to give minimal gaps for comparison, 133 amino acids of the F proteins were found to coincide with each other, but the overall homology was estimated to be only 23.5%. Interestingly, however, certain portions of the proteins show more than 50% homology, which were found from amino acid position 116 to 135 and from 458 to 477 of the F protein of Sendai virus (Fig. 9). This suggests the importance of these sequences for the function of the protein, since the N terminal portion of the Fl protein seemed to act as a functional domain during membrane fusion (42). It is noteworthy that the distribution of cysteine residues within the F proteins was well conserved between these two viruses. Both the F protein of Sendai virus and that of SV5 have 12 cysteine residues, out of which 10 could be aligned at the same positions, 1 in the F2 portion and 9 in the F1 portion, as shown in Fig. 8. The eight cysteine residues

clustered in the middle part of the Fl protein of Sendai virus are all conserved at the corresponding positions of the F protein of SV5. In spite of these similarities, however, there is no indication that the carbohydrate attachment sites are distributed similarly in the F proteins of the two viruses.

When the amino acid sequences of the HN proteins of these viruses were aligned, 138 amino acids coincided and the overall homology was about 24%.

F prot	ein		HN protein	HN protein				
Sendai	116 RFFGAVIGTIALGV	135 ATSAQI	Sendai	409 GAEGRLLK *****	428 CLGDRVYIYTRSS			
SV5	RFAGVVIGLAALGV 102	ATAAQV 121	SV5	GAEGRLYM 388	IYGDSVYYYQRSN 407			
	458 IRPVDISLNLADATI * * *** *** IDPLDISQNLAAVN	* * KSLSDA		* * ** CSATNRCP	482 KECISGVYTDAY * *** ** GFCLTGVYADAW			
	444	463		448	467			

Fig. 9. Highly conserved regions of the F and HN proteins of Sendai virus and SV5. Numbers indicate the amino acid positions. Asterisks indicate the identity of amino acid residues.

As in the case of the F proteins, however, highly conserved portions were found from amino acid position 409 to 428 and from amino 463 to 483 (Fig.9), which may be included in the active sites for hemagglutinin and neuraminidase. Similarities could also be found between these two viruses in the locations of the cysteine residues in the HN proteins and those of carbohydrate attachment sites, three of which were found in the C terminal regions of the proteins.

#### DISCUSSION

At the beginning of our work concerning the determination of the nucleotide sequence of the Sendai virus genome RNA, we obtained relatively long cDNAs, of about 3,000 to 4,000 nucleotides in length, starting from its 3' end, which could be sequenced satisfactorily (3). However, attempts to elongate these cDNAs by the primer extension method generally only yielded short cDNA clones of about 600 to 1,000 nucleotides in length (4), which greatly hampered the determination of the entire nucleotide sequence of the Sendai virus genome RNA. Thus, we decided to adopt a new cDNA cloning strategy, which involved starting cDNA synthesis from multiple sites in the genome RNA in combination with the cloning method of Okayama and Berg (16). It was nesessary for this purpose to cut the genome RNA into a few fragments and to add a poly(A)-tail to the resulting fragments. We found that this could be achieved by using an excess amount of poly A polymerase (P-L Biochemicals, lot no. 206-7) during the polyadenylylation reaction of the genome RNA, since we had found that a trace amount of ribonuclease activity contaminated the poly A polymerase preparation. Thus, we succeeded in establishing a set of cDNA copies that completely covered the Sendai virus

genome RNA, and could determine the whole sequence of the genome RNA of 15,383 nucleotides, which revealed that the gene structure of the Sendai virus is 3'-NP-P+C-M-F-HN-L-5'.

One of the characteristic features of the genome is that each gene is flanked by consensus sequences at both ends, that is, R1 at the 3' end and R2 at the 5' end. Since R1 shows minor variations from gene to gene, i.e. UCCCAGUUUC for the NP gene (3), UCCCACUUUC for the P+C (3), M (4) and HN genes, UCCCUAUUUC for the F gene (4), and UCCCACUUAC for the L gene, its common structure was deduced to be UCCC-A/U-C/G/A-UU-U/A-C or UCCCNNUUNC. On the other hand, R2 was found to be AUUCUUUUU for all genes. It is highly possible that R1 is the recognition sequence for viral RNA polymerase, minor differences in which may play a role in the control of the expression of each gene, while R2 is a polyadenylation signal. Consensus sequences similar to R1 and R2 were also reported for non-segmented negative stranded RNA viruses, i.e. VSV (43,44), respiratory syncytial virus (RSV) (45) and measles virus (46), suggesting their common importance in the transcription and/or replication process of these viruses. It is interesting to note that the sequence, GAA or GGG, which was found between two adjacent genes (or between R2 and R1) and thought to be an intergenic sequence (3,4), was detected after R2 of the L gene. This strongly indicates that R2 together with this trinucleotide may constitute a signal sequence for the termination of transcription as well as for polyadenylation. The finding that the 12 nucleotides of the very 5' end of the genome are complementary to the 12 nucleotides of the very 3' end of the genome is very important because it supports the prediction that the genome would form a very stable panhandle structure, which will provide the signal sequences for recognition by viral RNA polymerase and for association between the RNA and nucleocapsid proteins (47).

In VSV, it has been reported that about 50 nucleotides of the very 3' end of both the genome and antigenome RNA are transcribed to small RNAs designated as plus and minus leader RNAs, respectively (2,48). In Sendai virus, however, plus leader RNA but not minus leader RNA was detected in infected cells (2). Accordingly, it is of interest to investigate whether the 3' terminal 54 nucleotides of the Sendai virus antigenome, which is complementary to the 5' terminal 54 nucleotides of the genome RNA, can be transcribed to produce minus leader RNA, and to determine the function of minus leader RNA.

It is noteworthy that a stable secondary structure could be formed

within the non-coding portions of all the mRNAs for the F, HN and L proteins, giving each mRNA a panhandle structure, and similar structures are also possible for the construction of mRNAs for the NP, P+C and M genes (not shown). This type of secondary structure has been proposed for the mRNA of gene 10 of human rota virus (49), although the panhandle structure involves coding sequences in this case. The model presented in the present paper is of special interest, since the panhandle is constructed from only non-coding sequences, leaving the initiation as well as the termination codon within the single-stranded loop structure. According to this model, this secondary structure of the mRNA may be very important for its translation, because ribosomes may select the first AUG in the loop as the initiation codon, and bind directly to it with the aid of the initiation factors. In this regard, it is interesting to note that two forms of the secondary structure are possible for the construction of the mRNA of the P+C gene, which has two open reading frames that overlap (3). In one form, the initiation codon as well as the termination codon for both the P and C proteins are present within the loop structure, whereas in the other form, both the initation and termination codons for the P protein are buried in the stem structure, while those for the C protein remain in the loop structure. Details of these structures will be published elsewhere.

Recently, the nucleotide sequence of the F gene of respiratory syncytial virus (RSV) was reported (45,50). As expected from the fact that this virus is classified as a pneumovirus different from paramyxoviruses (51), there is little homology between the deduced amino acid sequence of this gene and that of the Sendai virus F gene. It is noteworthy, however, that the F protein of RSV also has a cluster of cysteine residues in the middle portion of the F1 protein (Fig. 8) as the F proteins of Sendai virus and SV5 do, indicating that this structure may be very important in the determination of the tertiary structure of biologically active F proteins.

The L protein of paramyxoviruses is expected to exhibit multifunctional activities as to viral transcription and replication, including the initiation, elongation, termination, polyadenylation, methylation and capping reactions, as in the case of the L protein of VSV. However, we could not find any significant homology in the amino acid sequence between the L gene product of Sendai virus and that of VSV (52), and it is too early to infer the functional domains of L proteins, and it is necessary to have further information on the structures of the L proteins of other paramyxovirus. Therefore, we have started analyzing the L gene of bovine parainfluenza type 3 virus.

#### ACKNOWLEDGEMENT

We wish to express our thanks to Dr. A. Nomoto for his helpful discussions. This work was supported by grants from the Ministry of Education, Science and Culture and the Ministry of Health and Welfare of Japan.

\*To whom correspondence should be addressed

#### REFERENCES

- 1. Lamb, R.A. and Choppin, P.W. (1978) Virology 84, 469-478.
- Leppert, M., Rittenhouse, L. Perraut, J., Summers, D.F. and Kolakofsky, D. (1979) Cell 18, 735-747.
- Shioda, T., Hidaka, Y., Kanda, T., Shibuta, H., Nomoto, A. and Iwasaki, K. (1983) Nucleic Acids Res. 11, 7317-7333.
- Hidaka,Y., Kanda,T., Iwasaki,K., Nomoto,A., Shioda,T. and Shibuta,H. (1984) Nucleic Acids Res. 12, 7965-7972.
- 5. Giorgi, C., Blumberg, B.M. and Kolakofsky, D. (1983) Cell 35, 829-936.
- Blumberg, B.M., Giorge, C., Rose, K. and Kolakofsky, D. (1984) J.gen.Virol. 65, 769-779.
- Blumberg,B.M., Rose,K., Simona,M.G., Roux,L., Giorgi,C. and Kolakofsky,D. (1984) J.Virol. 52, 656-663.
- Blumberg,B.M., Gorgi,C., Rose,K. and Kolakofsky,D. (1985) J.gen.Virol. 66, 317-331.
- 9. Morgan, E.M., Re, G.G. and Kingsbury, D.W. (1984) Virolgy 135, 279-287.
- Dowling, P.C., Giorgi, C., Roux, L., Dethlesen, L.A., Galantowitz, M.E., Blumberg, B.M. and Kolakofsky, D. (1983) Proc.Natl.Acad.Sci. USA. 80, 5213-5216.
- 11. Glazier, K., Raghow, R. and Kingsbury, D.W. (1977) J.Virol. 21, 863-871.
- Paterson,R.G., Harris,T.J.R. and Lamb,R.A. (1984) Proc.Natl.Acad.Sci. USA. 81, 6706-6710.
- 13. Hiebert, S.W., Paterson, R.G. and Lamb, R.A. (1985) J.Virol. 54, 1-6.
- 14. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochmistry 18, 5294-5299.
- Inokuchi,Y. Hirashima,A. and Watanabe,I. (1982) J.Mol.Biol. 158, 711-730.
- 16. Okayama, H. and Berg, P. (1982) Mol.Cel.Biol. 2, 161-170.
- 17. Boliver, F. and Backman, K. (1979) Methods Enzymol. 68, 245-267.
- 18. Messing, J. (1983) Methods Enzymol. 101, 20-78.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc.Natl.Acad.Sci. USA. 74, 5463-5467.
- Grunstein, M. and Hogness, D.S. (1975) Proc.Natl.Acad.Sci. USA. 72, 3961-3965.
- 21. Wahl,G.M., Stern,M. and Stark,G.R. (1979) Proc.Natl.Acad.Sci. USA. 76, 3683-3687.
- 22. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- Rigby, P.W.J., Deckman, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol. 113, 237-259.
- 24. Re,G.G., Gupta,K.C. and Kingsbury,D.W. (1983) Virology 130, 390-396.

- Kolakofsky, D., Boy de la Tour, E. and Delius, H. (1974) 25. J.Virol. 13, 261-268. Shibuta, H., Kanda, T., Adachi, A. and Yogo, Y. (1979) 26. Microbiol. Immunol. 23, 617-628. Nakamura, K., Hommma, M. and Compans, R.W. (1982) Virology 27. 119, 474-487. 28. Hsu, M.-C. and Choppin, P.W. (1984) Proc.Natl.Acad.Sci.USA. 81, 7732-7736. 29. Lamb, R.A., Mahy, B.W.J. and Choppin, P.W. (1976) Virology 69, 116-131. Tinoco Jr.I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., 30. Chorothers, D.M. and Gralla, J. (1973) Nature New Biol. 246, 40-41. Lyle, D.S. (1979) Proc.Natl.Acad.Sci.USA. 76, 5621-5625. 31. Gething, M.J., Bye, J., Skehel, J. and Waterfield, M. (1980) 32. Nature 287, 301-306. 33. Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymology 47, 45-148. 34. Neuberger, A., Gottschalk, A., Marshall, R.D. and Spiro, R.G. (1972) in The Glycoproteins: Their composition, structure and function, Gottschalk, A. Ed., pp. 450-490, Elsevier, Amsterdam. 35. Kohama, T., Shimizu, K. and Ishida, N. (1978) Virology 90, 226-234. 36. Blok, J., Air, G.M., Laver, W.G., Ward, C.W., Lilley, G.G., Woods, E.F., Roxburgh, C.M. and Inglis, A.S. (1982) Virology 119, 109-121. 37. Bos,T.J., Davis,A.R. and Nayak,D.P. (1984) Proc.Natl.Acad.Sci.USA. 81, 2327-2331. 38. Kamer, G. and Agros, P. (1984) Nucleic Acids Res. 12, 7269-7282. 39. Kozak, M. (1984) Nature 308, 241-246. Blumberg, B.M., Giorgi, C., Roux, L., Ramaswamy, R., Dowling, P., 40. Chollet, A. and Kolakofsky, D. (1985) Cell 41, 269-278. 41. Miura, N., Nakatani, Y., Ishimura, M., Uchida. T. and Okada, Y. (1985) FEBS lett. 188, 112-116. Richardson, C.D., Sceid, A. and Choppin, P.W. (1980) Virology 105, 42. 205-222. 43. Rowlands, D.J. (1979) Proc.Natl.Acad.Sci. USA. 76, 4793-4797. Rose, J.K. (1980) Cell 19, 415-421. 44. Collins, P.L., Huang, Y.T. Wertz, G.W. (1984) Proc.Natl.Acad.Sci. 45. 81, 7683-7687. 46. Bellini, W.J., Englund, G., Rozenblatt, S., Arnheiter, H. and Richardson, C.D. (1985) J.Virol. 53, 908-919. 47. Blumberg, B.M., Giorgi, C. and Kolakofsky, D. (1983) Cell 32, 559-567. 48. Colonno, R.J. and Banerjee, A.K. (1976) Cell 8, 197-204. 49. Okada,Y., Richardson,M.A., Ikegami,N., Nomoto,A. and Furuichi,Y. (1984) J.Virol. 51, 856-859. Elango, N., Satake, M., Coligan, J.E., Norrby, E., Camargo, E. 50. and Venkatesen, S. (1985) Nucleic Acids Res. 13, 1559-1574.
  - 51. Matthews, R.E.F. (1982) Intervirology 17, 1-200.
  - 52. Schubert, M., Harmison, G.G. and Meier, E. (1984) J. Virol. 51, 505-514.